BioLaw Journal
Rivista di BioDiritto

UNIVERSITÀ DI TRENTO

1 24

Special Issue

a cura di M. Tomasi, L. Busatta, M. Fasan, C. Nardocci, S. Penasa, S. Sulmicelli

Special Issue || Vulnerabilità e Intelligenza Artificiale

The online Journal about law and life sciences

BioLaw Journal – Rivista di BioDiritto

Editor in chief: Carlo Casonato

Steering Committee: Roberto Bin, Antonio D'Aloia, Alessandro Pajno

Scientific Committee:
Roberto Andorno, Vittorio Angiolini, Charles H. Baron, Alberto Bondolfi, Paolo Benciolini, Patrizia Borsellino, Roger Brownsword, Massimiano Bucchi, Stefano Canestrari, Cinzia Caporale, Maria Chiara Carrozza, Paolo Carrozza (†), Lorenzo Chieffi, Ricardo Chueca Rodríguez, Roberto Cingolani, Roberto Giovanni Conti, Roberto Dias, Frédérique Dreifuss-Netter, Gilda Ferrando, Silvio Garattini, Francesca Giardina, Stefano Guizzi, Stéphanie Hennette-Vauchez, Juan Alberto Lecaros, Sheila McLean, Laura Palazzani, Marco Pandolfi, Barbara Pezzini, Cinzia Piciocchi, Alessandra Pioggia, Anna Maria Poggi, Carlo Alberto Redi, Fernando Rey Martinez, Stefano Rodotà (†), Carlos Maria Romeo Casabona (†), Amedeo Santosuosso, Stefano Semplici, Paula Siverino Bavio, Mariachiara Tallacchini, Chiara Tripodina, Gianni Tognoni, Paolo Veronesi, Umberto Veronesi (†), Paolo Zatti.

Associate Editors: Lucia Busatta and Marta Tomasi

Editorial Boards:
Trento: Giorgia Bincoletto, Lucia Busatta, Marta Fasan, Paolo Guarda, Antonio Iannuzzi, Ilja Richard Pavone, Simone Penasa, Mariassunta Piccinni, Ludovica Poli, Elisabetta Pulice, Carla Maria Reale, Elena Scalcon, Marta Tomasi.
Ferrara: Paolo Veronesi, Giuseppina Barcellona, Fabio Ferrari, Migle Laukyte, Benedetta Liberali, Nicola Lucchi, Irene Pellizzone, Silvia Zullo.
Parma: Stefano Agosta, Giancarlo Anello, Maria Chiara Errigo, Giulia Formici, Valentina Gastaldo, Valeria Marzocco, Erika Ivalù Pampalone, Giovanna Razzano, Lucia Scaffardi, Veronica Valenti.
Napoli: Lorenzo Chieffi, Gianvito Brindisi, Claudia Casella, Gianpiero Coletta, Emilia D'Antuono, Luca Di Majo, Luigi Ferraro, Maria Pia Iadicicco, Carlo Iannello, Raffaele Manfrellotti, Ferdinando Menga, Franca Meola, Andrea Patroni Griffi, Virginia Zambrano.

E-mail: biodiritto@gmail.org
Website: https://teseo.unitn.it/biolaw

Peer review system: All academic articles that are submitted to *BioLaw Journal – Rivista di BioDiritto* are subject to a double blind peer review. Referees remain anonymous for the author during the review procedure and the author's name is removed from the manuscript under review.

December 2024
ISSN 2284-4503
© Copyright 2023

UNIVERSITY OF TRENTO - Italy

Università degli Studi di Trento
Via Calepina, 14 – 38122 Trento
Registrazione presso il Tribunale di Trento n. 6 dell'11/04/2014

In collaborazione con

BioDiritto
UNIVERSITÀ DI PARMA
UCB University Center for Bioethics

Front cover: Graphic project based on "Tomba del tuffatore", Paestum, 5th century b.C., on permission nr. 15/2014 by Ministero dei Beni e delle Attività Culturali e del Turismo – Soprintendenza per i Beni Archeologici di SA, AV, BN e CE.
Cover design: Marta Tomasi

Contents

BioLaw Journal – Rivista di BioDiritto
Special issue n. 1/2024

Table of contents

Downloaded from www.biodiritto.org.
ISSN 2284-4503

CC BY NC ND

Contents

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Contents

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Editorial

Vulnerabilità e Intelligenza Artificiale

Marta Tomasi, Lucia Busatta, Marta Fasan, Costanza Nardocci, Simone Penasa, Sergio Sulmicelli

"Vulnerabilità" è un concetto che reca numerosi significati, complesso e al tempo stesso evocativo, applicabile a diversi contesti, per il perseguimento di obiettivi differenti[1]. Nella maggior parte delle ricostruzioni che si sono interessate dell'argomento, si affiancano (almeno) due macro-significati del termine: l'uno, più generale e immanente, rimanda all'ontologia stessa della persona, evocando una caratteristica universale della condizione umana; l'altro, più specifico e variabile, si riferisce a situazioni, contesti e momenti che espongono la persona a forme di fragilità, anche di gruppo[2]. In questo secondo significato, la vulnerabilità è il prodotto delle società nelle quali viviamo, del modo in cui esse sono costruite e regolate. La vulnerabilità non è tratto identificativo che qualifica il singolo soggetto portatore di una caratteristica o collocato in un certo contesto situazionale, ma è una caratteristica indotta, prodotta da un ambiente nel quale la persona è immersa, al punto che, in alcune elaborazioni, si suggerisce che l'aggettivo "vulnerabile" possa essere sostituito dall'alternativo "vulnerabilizzato", maggiormente rispondente alla realtà di una condizione eterodeterminata[3].

Un termine che viene preferito quando si vuole enfatizzare l'origine strutturale e contingente della vulnerabilità, rispetto alla sua natura intrinseca o universale.

Fra i fattori che oggi possono incidere sulla vita delle persone, le tecnologie – e fra tutte sicuramente l'Intelligenza Artificiale – giocano un ruolo determinante.

I nuovi sistemi intelligenti stanno alimentando processi trasformativi e pervasivi, capaci di incidere profondamente su ogni aspetto della vita contemporanea. Dalla medicina all'educazione, dal lavoro ai trasporti, passando per la comunicazione e le relazioni sociali, l'IA non si limita a innovare i processi esistenti, ma ne sta ridefinendo i paradigmi, imponendo una riflessione complessiva su come viviamo, pensiamo e ci rapportiamo al mondo. Le sue capacità di apprendere, rielaborare e agire in modi che simulano, ampliano o sfidano le capacità umane sollevano interrogativi fondamentali sulla nostra stessa identità come esseri umani e, conseguentemente, sulle situazioni di vulnerabilità che ciascuno può trovarsi a vivere[4].

Le tecnologie intelligenti possono dispiegare una forza ambivalente entrando in una dinamica bidirezionale e articolata con il concetto di vulnerabilità. Da un lato, anche in ragione di specificità di natura tecnica (es. "black box", mancanza di trasparenza, *bias* algoritmici e discriminazioni) e della molteplicità dei contesti di utilizzo, l'I.A. può creare vulnerabilità nuove, inducendo meccanismi di esclusione sociale, o acuire vulnerabilità esistenti per quei soggetti e gruppi sociali storicamente in posizione di subordinazione e marginalizzazione. D'altro canto, però, laddove adeguatamente costruito, imple-

[1] M. Dunn, I. Clare, A. Holland, *To empower or to protect? Constructing the 'vulnerable adult'*, in *English law and public policy. Legal Studies*, 28, 2008, 234-254.

[2] Una prospettiva è quella di M.A. Fineman, *The Vulnerable Subject: Anchoring Equality in the Human Condition*, in *Yale Journal of Law and Feminism*, 20, 1, 2008, 8-10.

[3] Si v., per esempio, B. Casalini, *Politics, justice and the vulnerable subject: the contribution of feminist thought*, in *Revista Gênero & Direito*, 5, 2016, 15-29.

[4] C. Casonato, *Intelligenza artificiale e diritto costituzionale. Prime considerazioni*, in *Diritto pubblico comparato ed europeo*, Fascicolo speciale, 2019, 131-144.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

mentato e regolamentato, l'impiego di tecnologie di I.A. può anche rivelarsi uno strumento capace di rinforzare la posizione della persona nella società, sottraendola a processi esclusivi.

È questa una delle ragioni che giustificano l'attenzione crescente che il costituzionalismo contemporaneo riserva alle questioni della regolamentazione della tecnologia e dell'I.A.. Rappresentando oggi una forma di potere, infatti, l'Intelligenza Artificiale non può che sollecitare l'attenzione di una disciplina che trova la propria ragione di essere nell'obiettivo, duplice, di porre argini ai poteri che determinano condizioni di vulnerabilità e di promuovere strumenti che siano potenzialmente in grado di contribuire alla piena realizzazione della persona umana.

A partire da queste riflessioni, *BioLaw Journal – Rivista di BioDiritto* ha proposto, all'inizio di quest'anno, una *call for papers* volta ad analizzare, anche in chiave interdisciplinare, le complesse relazioni che si sono descritte.

I contributi selezionati sono stati raccolti in tre gruppi tematici.

Nel primo (*Dell'esistenza e delle forme della vulnerabilità*) sono confluiti i saggi caratterizzati da un respiro più ampio, per la maggior parte di natura giusfilosofica che, muovendo da riflessioni intorno all'esistenza della vulnerabilità (Corradi), all'analisi concettuale delle sue molteplici dimensioni (il senso universale e quello particolare, l'approccio categorizzante e quello situazionale, di cui tratta Dadà), passano per indagini circa la necessarietà di un discorso intorno all'etica dell'IA e al grado di effettività dello stesso, giungendo a posizioni conclusive vicine, ma in parte distoniche, che riferiscono da un lato del bisogno di integrare ragionamenti sociali di ampia portata, al fine di garantire uno sviluppo tecnologico inclusivo (Carnevale) e dall'altro di una marcata e forse insuperabile incompatibilità tra la struttura ontologica presupposta dalla vulnerabilità, da una parte, e dai sistemi di intelligenza artificiale, dall'altra (Corradi).

Poggiando su questi presupposti teorici del dibattito, altri autori indagano, in analoga prospettiva, i contenuti di uno dei principali atti normativi intervenuti sul tema: il Regolamento dell'Unione Europea sull'Intelligenza Artificiale (Regolamento (UE) 2024/1689). Il vaglio delle più significative ricorrenze del concetto di vulnerabilità nel testo e l'analisi delle sue principali variazioni (Dadà e Galli Novelli) fanno emergere, da un lato, l'approccio parziale della normativa europea, che sembra focalizzarsi su alcune forme particolari di vulnerabilità, tralasciandone il senso universale, che rimane relegato all'implicito (Dadà). Al tempo stesso, però, emerge anche l'esigenza forte di non rinunciare a trovare un aggancio e una collocazione del concetto anche all'interno dell'AI Act, che potrebbe realizzarsi attraverso un'interpretazione ampia del riferimento a una "specific social situation" contenuto nell'art. 5(b) (Galli - Novelli).

Un terzo contributo (Nardocci), a chiusura di questo gruppo, mette a confronto i contenuti dell'AI Act, nato come tentativo di regolamentazione di un prodotto, con l'obiettivo primario di migliorare il funzionamento del mercato interno, con un altro atto, la *Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law*, recentemente adottata nel contesto europeo in seno al Consiglio d'Europa, un organismo internazionale che da sempre mette la tutela dei diritti fondamentali al centro del suo agire.

Un secondo gruppo di scritti (*Dei luoghi della vulnerabilità*) trova radice comune nel fatto di volgere l'attenzione a contesti particolari, a settori paradigmatici all'interno dei quali la vulnerabilità si manifesta con evidenza estrema ed è destinata a entrare in relazione sempre più

Downloaded from www.biodiritto.org.
ISSN 2284-4503

stretta con le tecnologie dell'I.A. Fra tutti questi, ovviamente, il mondo della rete e lo spazio digitale offrono interessanti spunti di riflessione che mettono in luce almeno due forme di vulnerabilità nuove e spesso inconsapevoli: da un lato, la vulnerabilità individuale e sociale di tutti coloro che vivono una parte sempre più consistente delle proprie esistenze in uno spazio virtuale, gestito da poteri economici e sociali le cui forme di controllo e limitazione devono ancora essere affinate (le *social platforms* di Di Majo). Dall'altro, sono ormai emersi con chiarezza i profili di forme di vulnerabilità tecnologica, determinate dal diffondersi delle minacce informatiche, al contenimento delle quali l'IA sembra poter attivamente e positivamente contribuire, purché il diritto si apra a un approccio integrato che coinvolga tecnologia, normativa e cooperazione tra gli stakeholder (l'analisi giusinformatica di Brighi).

Il collegamento stretto fra virtuale e digitale emerge, infine, in un ulteriore spazio, che è fisico, ma tecnologicamente intriso, quello delle smart cities, un luogo nel quale i fondamenti dell'identità, dell'autonomia, della libertà e della responsabilità sono sottoposti a tensione, imponendo di riconsiderare attentamente il senso stesso dell'essere cittadini (Vignola).

Altrettanto significativo, seppure più specifico, pare il contesto medico-sanitario che è stato attraversato in pieno dalla rivoluzione tecnologica: qui, l'impatto tecnologico, oltre a presentare profili specifici connessi alla vulnerabilità digitale (si pensi alle difficili sfide del bilanciamento tutela della salute e diritto alla riservatezza di dati sempre più "sensibili"), ha reso manifesto e più pressante il problema delle discriminazioni e delle diseguaglianze. Tale questione, muovendo dalla fase dei *trial* clinici, passando per l'accesso alle cure, fino all'erogazione delle stesse, sembra non essere mai stata risolta, e forse nemmeno sistemicamente affrontata.

I contributi qui raccolti mettono in luce il ruolo "disvelatore" delle nuove tecnologie che, a causa del loro funzionamento, basato su uno storico di dati raccolti, rischiano di perpetrare iniquità consolidate (Cicu - Colangelo - Saba), soprattutto con riferimento a determinati gruppi sociali (le minoranze etniche, nel contributo di Lando e le donne, in quello di Piva e di Previti). Sempre in questa sezione, si è deciso di dare spazio a un contesto particolare, quello agricolo, in quanto rappresentativo di alcune delle sfide "globali" che caratterizzeranno, se già non caratterizzano, le riflessioni giuridiche e costituzionalistiche dei prossimi anni: la povertà alimentare e i rischi ambientali, infatti, integrano forme di vulnerabilità sempre più diffuse che prolificano anche in ragione delle loro intersezioni con altre forme di svantaggio sociale (De Tullio).

La varietà delle situazioni che il concetto di vulnerabilità evoca si percepisce, infine, in pienezza guardando ai contributi raccolti nella terza e ultima sezione, dedicata ai gruppi, o meglio alle categorie, di persone che presentano caratteristiche specifiche di fragilità (*Dei volti umani della vulnerabilità*). Qui trovano collocazione quasi spontanea le analisi dedicate ai rischi della discriminazione algoritmica che, in ragione dei meccanismi di funzionamento stesso della tecnologia, basata su set di dati organizzati, rischia di rinforzare categorizzazioni più o meno palesi e leggibili o di crearne di nuove. In questo senso, ai contributi dedicati, in termini differenti, alle questioni di genere, che interessano uomini e donne (Pulito, Viggiani, Gambatesa), ma anche le persone trans (Di Giovanni), si affiancano, nelle pagine che seguono, stimolanti riflessioni che indagano meccanismi di esclusione espliciti, come quelli che si determinano alla frontiera,

Downloaded from www.biodiritto.org.
ISSN 2284-4503

dove i controlli sono sempre più di frequente basati su strumenti di riconoscimento biometrico (Nobile), o impliciti, che determinano situazioni di marginalità (o vulnerabilità) che possono essere date, per esempio, dall'appartenenza a un determinato gruppo linguistico (Morganti - Zuaro) o, addirittura, a un gruppo di consumatori identificato sulla base di criteri di classificazione discriminatori (Curcuruto - Inturri).

Seguono, infine, stimolanti riflessioni intorno alle complesse sfide dei processi di costruzione di identità personale che interessano – seppure in termini diversi – i minori di età (Maccabiani, Pileggi e Vadalà), le persone con disabilità (Pagnanelli) e le persone trans (ancora Di Giovanni e Ciccarella - Fortuna - Lambiase - Mogetti), sfide che impegnano i poteri pubblici e privati al pari delle esigenze di fornire loro adeguata protezione e tutela.

L'insieme composito di questi studi dedicati alla vulnerabilità fa emergere, con evidenza, il valore irrinunciabile della diversità, al quale le società fondate sui valori del costituzionalismo democratico non possono rinunciare in nome dell'esigenza di rendere computabile e classificabile la realtà.

Senza alcuna pretesa di completezza, la presente Special Issue, rivela il valore fecondo della nozione di vulnerabilità. Sebbene la maggior parte dei contributi sia dedicato a un profilo specifico di fragilità, l'accento che ciascuno pone sulla parzialità del proprio sguardo, sull'esigenza di aprirsi a un approccio intersezionale, che sia capace di coordinare più vulnerabilità e di sciogliere la rigidità di alcune categorie, fa pensare che, per avvicinarsi a un tentativo di comprensione del tema, sia indispensabile guardare alla complessiva condizione umana e mettere a sistema più fattori. Un obiettivo che, in fin dei conti, le strutture dell'IA potrebbero aiutare a perseguire.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Special issue

Comprendere la vulnerabilità. Pluralismo ontologico e sistemi di intelligenza artificiale nel diritto

Silvia Corradi*

GRASPING VULNERABILITY. ONTOLOGICAL PLURALISM AND ARTIFICIAL INTELLIGENCE SYSTEMS IN THE LAW

ABSTRACT: The paper aims at investigating the relationship between vulnerability and artificial intelligence systems in law, following a philosophical-legal reading. The investigation therefore questions, firstly, the existence of vulnerability, in an attempt to clarify its definitions and ways of grasping it. Secondly, the paper proposes the example of pain, dwelling on its ontological and (onto-)epistemological profiles, nominating liberalised naturalism as the appropriate realist framework in identifying it. Thirdly, by investigating informational structural realism, the partial incompatibility between the ontological structure presupposed by vulnerability, on the one hand, and artificial intelligence systems, on the other, is highlighted.

KEY WORDS: vulnerability; ontological pluralism; law; reality; artificial intelligence systems.

ABSTRACT: Il contributo si propone di indagare il rapporto tra vulnerabilità e sistemi di intelligenza artificiale nel diritto, seguendo una lettura filosofico-giuridica. L'indagine si interroga quindi, *in primis*, sull'esistenza della vulnerabilità, nel tentativo di chiarirne le definizioni e modalità di comprensione. In secondo luogo, lo scritto propone l'esempio del dolore, soffermandosi sui relativi profili ontologici ed (onto-)epistemologici, candidando il naturalismo liberalizzato ad appropriata cornice realista nell'individuazione di esso. In terzo luogo, indagando il realismo strutturale informazionale, viene posta in luce la parziale incompatibilità tra la struttura ontologica presupposta dalla vulnerabilità, da una parte, e dai sistemi di intelligenza artificiale, dall'altra.

PAROLE CHIAVE: Vulnerabilità; pluralismo ontologico; diritto; realtà; sistemi di intelligenza artificiale.

* *Assegnista di ricerca, Università di Palermo. Mail: silvia.corradi@unipa.it.* *Contributo sottoposto a doppio refereraggio anonimo.*

Downloaded from www.biodiritto.org.
ISSN 2284-4503

1. Introduzione

Il contributo si propone di indagare il rapporto tra vulnerabilità e sistemi di intelligenza artificiale: nello specifico, si interroga sulla possibilità di comprensione, da parte di un sistema di intelligenza artificiale, della situazione di vulnerabilità in cui versa un soggetto. A ciò consegue domandarsi circa l'esistenza della vulnerabilità, e dunque, sia di come questa possa essere definita, sia di come essa possa essere rilevata. L'indagine, che prediligerà una lettura filosofico-giuridica, intende porre in luce la diversità – e, quindi, in questo caso, della parziale incompatibilità – tra la struttura ontologica presupposta dalla vulnerabilità, da una parte, e dai sistemi di intelligenza artificiale, dall'altra.

L'occasione per questa riflessione scaturisce dall'approvazione istituzionale europea del Regolamento inerente allo sviluppo di sistemi di intelligenza artificiale (anche denominato "Legge sull'I.A."), il quale, dopo una sua prima formulazione nelle vesti di "Proposta di Regolamento del Parlamento Europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale) e modifica alcuni atti legislativi dell'Unione" del 21 aprile 2021, giunge nell'anno corrente ad una sua versione definitiva[1]. Dopo aver chiarito, nei considerando del Regolamento, i rischi sottesi all'utilizzo di sistemi di intelligenza artificiale per i casi di vulnerabilità[2], l'art. 5 del Regolamento al co. 1, lettera b) pone il divieto di commercializzazione di sistemi siffatti che sfruttino le vulnerabilità di determinati soggetti[3].

La formulazione normativa, secondo la lettura che si intende proporre, pone qualche perplessità: essa sembra, infatti, accogliere una concezione "categoriale" di vulnerabilità (su cui si dirà qualcosa *infra* par. 2), tralasciando invece la concezione "ontologica", che non è riconducibile a singoli e precisi criteri (come invece il Regolamento lascerebbe intendere, giacché indica criteri quali l'età o l'afferenza ad un gruppo di persone). Si scopre così che la vulnerabilità presuppone una realtà particolare,

[1] Il Regolamento europeo in esame è denominato "Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio, del 13 giugno 2024 che stabilisce regole armonizzate sull'intelligenza artificiale e modifica i regolamenti (CE) n. 300/2008, (UE) n. 167/2013, (UE) n. 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 (regolamento sull'intelligenza artificiale)".

[2] Così si legge al punto 29 del Regolamento: «[I] sistemi di IA possono [inoltre] sfruttare [in altro modo] le vulnerabilità di una persona o di uno specifico gruppo di persone dovute all'età, a disabilità ai sensi della direttiva (UE) 2019/882 del Parlamento europeo e del Consiglio o a una specifica situazione sociale o economica che potrebbe rendere tali persone più vulnerabili allo sfruttamento, come le persone che vivono in condizioni di povertà estrema e le minoranze etniche o religiose»; al punto 48 del Regolamento viene ricordata la delicata posizione dei minori: «è importante sottolineare il fatto che i minori godono di diritti specifici sanciti dall'articolo 24 della Carta e dalla Convenzione delle Nazioni Unite sui diritti dell'infanzia e dell'adolescenza, ulteriormente sviluppati nell'osservazione generale n. 25 della Convenzione delle Nazioni Unite dell'infanzia e dell'adolescenza per quanto riguarda l'ambiente digitale, che prevedono la necessità di tenere conto delle loro vulnerabilità e di fornire la protezione e l'assistenza necessarie al loro benessere». Per un inquadramento generale della Proposta di Regolamento, G. SARTOR, *L'intelligenza artificiale e il diritto*, Torino, 2022, 91-94; C. CASONATO, B. MARCHETTI, *Prime osservazioni sulla Proposta di Regolamento dell'Unione Europea in materia di intelligenza artificiale*, in *BioLaw Journal – Rivista di Biodiritto*, 3, 2021, 418 ss.

[3] L'art. 5 co. 1 lettera b) del Regolamento europeo enuncia il divieto di «immissione sul mercato, la messa in servizio o l'uso di un sistema di IA che sfrutta le vulnerabilità di una persona fisica o di uno specifico gruppo di persone, dovute all'età, alla disabilità o a una specifica situazione sociale o economica, con l'obiettivo o l'effetto di distorcere materialmente il comportamento di tale persona o di una persona che appartiene a tale gruppo in un modo che provochi o possa ragionevolmente provocare a tale persona o a un'altra persona un danno significativo».

Downloaded from www.biodiritto.org.
ISSN 2284-4503

strettamente legata all'esistenza umana e pertanto individuabile a partire da quella (la questione sarà trattata in *infra* par. 3). Nel tentativo di approfondire le suggestioni che provengono a tal riguardo dalla filosofia del diritto, verrà considerato, in via esemplificativa, lo statuto ontologico del dolore (par. 3.1.); a seguito della trattazione della questione ontologica seguirà una domanda (onto-)epistemologica, che si interroga sulle modalità tramite cui la vulnerabilità possa essere compresa (par. 3.2.). Infine, si cercherà di capire quale sia la realtà presupposta dai sistemi di intelligenza artificiale, approfondimento che verrà condotto a partire dal realismo strutturale informazionale (par. 4).

2. La vulnerabilità: due concezioni e un concetto

Una definizione normativa del concetto di "vulnerabilità" non è presente né all'interno dell'ordinamento italiano né tantomeno nel testo della CEDU[4]: tuttavia, la giurisprudenza fa uso di esso, imponendo così, in determinati casi, di vagliare quando un soggetto versi in siffatto stato. Un esempio nel contesto nazionale è rappresentato dalla sussistenza del reato *ex* art. 643 c.p., che punisce la circonvenzione di incapace. In questo caso, ai fini della configurabilità dell'illecito penale, è richiesto, *inter alia*, che sia accertato «l'abuso dello stato di vulnerabilità che si verifica quando l'agente, consapevole di detto stato, ne sfrutti la debolezza per raggiungere il suo fine, ossia quello di procurare a sé o ad altri un profitto»[5]. Lo stato di vulnerabilità viene così presupposto dal terzo requisito[6] per la sussistenza del reato in questione, ma resta sprovvisto di una formulazione (giurisprudenziale o codicistica). Richiamando un orientamento che pare consolidato, la Cassazione penale, nella sentenza citata, ha chiarito che il delitto *de quo* «non postula che la vittima versi in stato di incapacità di intendere e di volere, essendo sufficiente […] un'alterazione dello stato psichico che […] risulti idoneo a porla in uno stato di minorata capacità intellettiva, volitiva o affettiva, che ne affievolisca le capacità critiche»[7]. La Corte parla anche di "fragilità" al fine di identificare siffatta situazione, la quale «deve avere

[4] E. Diciotti, *La vulnerabilità nelle sentenze della Corte europea dei diritti dell'uomo*, in *Ars interpretandi*, 2, 2018, 13; T. Casadei, *Diritti umani in contesto: forme della vulnerabilità e "diritto diseguale"*, in *Ragion pratica*, 2, 2008, 291.

[5] Cass. penale, sez. II, sent. 31 maggio 2024, n. 30551, punto 1.2. in diritto: nel caso di specie il ricorrente lamentava il fatto che la "condizione di debolezza psichica" della donna di ottantasei anni sarebbe iniziata nel giugno 2020, un periodo susseguente a quello all'interno del quale l'anziana signora avrebbe effettuato dei versamenti a beneficio del ricorrente, che era solita frequentare tramite il social network Facebook. Il requisito dell'abuso dello stato di vulnerabilità per la configurazione dell'illecito penale di circonvenzione di persona incapace trova conferma in altri luoghi nella giurisprudenza nazionale: *ex multis*, Cass. penale, sez. II, sent. 13 marzo 2024, n. 13557; Cass. penale, sez. II, 1 marzo 2019, n. 19834; Trib. Milano, sez. spec. impresa, sent. 4 aprile 2022, n. 2913.

[6] I requisiti enunciati sono i seguenti: «a) l'instaurazione di un rapporto squilibrato fra vittima e agente, in cui quest'ultimo abbia la possibilità di manipolare la volontà della vittima, che, in ragione di specifiche situazioni concrete (minore età, infermità o deficienza psichica), sia incapace di opporre alcuna resistenza per l'assenza o la diminuzione della capacità critica; b) l'induzione a compiere un atto che importi per il soggetto passivo o per altri qualsiasi effetto giuridico dannoso; c) l'abuso dello stato di vulnerabilità che si verifica quando l'agente, consapevole di detto stato, ne sfrutti la debolezza per raggiungere il suo fine, ossia quello di procurare a sé o ad altri un profitto; d) l'oggettiva riconoscibilità della minorata capacità, in modo che chiunque possa abusarne per raggiungere i suoi fini illeciti» (Cass. penale, sez. II, sent. 31 maggio 2024, n. 30551, punto 1.2. in diritto).

[7] Loc. ult. cit.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

natura oggettiva»[8]. Tuttavia, le precisazioni della Suprema Corte non sembrano potersi ricondurre specificamente allo stato di vulnerabilità ma alle condizioni, complessivamente considerate, necessarie per la configurabilità del reato: il concetto di vulnerabilità mantiene così contorni sfumati per quanto concerne la sua precisa individuazione.

In effetti, "vulnerabilità" è un concetto versatile[9], che non ha ricevuto formulazione univoca nemmeno da parte della giurisprudenza della Corte Europea dei Diritti Umani[10]: ciò spiega perché sia sovente evidenziato che quella delle «"persone deboli e vulnerabili" non sia né una categoria giuridica, né – più in generale – una categoria ben definita»[11]. Ciononostante, negli ultimi anni l'utilizzo quantitativo del concetto di "vulnerabilità" da parte della Corte di Strasburgo è cresciuto sensibilmente (mentre nell'anno 2000 solo sette sentenze contenevano questo termine, nell'anno 2013 se ne registravano settanta)[12] e anche sul piano nazionale la protezione ai soggetti vulnerabili viene assicurata attraverso gli articoli 2 e 3 co. 2 della Costituzione, in base al "principio di solidarietà"[13]. Ciò testimonia il crescente interesse per la vulnerabilità, al punto che è stato possibile individuare, in ambito politico e giuridico, due principali accezioni[14].

Secondo una prima accezione, vulnerabilità è intesa come «una caratteristica di gruppi all'interno della società e della comunità politica»[15]; in una seconda accezione, invece, essa designa «una condizione universale dell'essere umano, suscettibile di manifestazioni diverse per tipologia e intensità, a

[8] Loc. ult. cit.

[9] Viene considerato un concetto «sia descrittivo che normativo e funziona come uno strumento euristico con il quale analizzare e correggere l'agire istituzionale» (L. Re, *Politica e istituzioni al tempo del cambiamento climatico. Il paradigma della vulnerabilità come proposta di trasformazione*, in *Materiali per una storia della cultura giuridica*, 1, 2020, 296); Chenal, invece, lo considera un concetto «funzionale e strumentale funzionale e strumentale a garantire la massima effettività dei diritti fondamentali» (R. Chenal, *La definizione della nozione di vulnerabilità e la tutela dei diritti fondamentali*, in *Ars interpretandi*, 2, 2018, 51). La letteratura nazionale ed internazionale inerente alla vulnerabilità è vasta e comprende diverse ramificazioni più specifiche (per esempio, la vulnerabilità con riferimento alla disabilità o alla situazione di migrante). In questo scritto ci limiteremo ad analizzare i più recenti sviluppi di tale tematica, sotto una prospettiva filosofico-giuridica, prediligendo la letteratura nazionale.

[10] R. Chenal, *op. cit.*, 35. L'Autore precisa, inoltre, che tale formulazione non si possa legittimamente accogliere, sia che si tratti di provenienza convenzionale, sia giurisprudenziale: «Se[, come si spera di essere riusciti a provare,] la Corte europea, in virtù dell'argomentazione per principi che caratterizza il sistema di protezione della Convenzione, ritiene, in sostanza, che non si possa pervenire in astratto a una definizione unitaria, coerente (e tassativa) della nozione di vulnerabilità sul piano legislativo, non si potrà che giungere alla stessa conclusione se si rivolge l'attenzione alla stessa Corte», R. Chenal, *op. cit.*, 49.

[11] F. Poggi, *Il caso Cappato: la Corte costituzionale nelle strettoie tra uccidere e lasciar morire*, in *BioLaw Journal – Rivista di BioDiritto*, 1, 2020, 85.

[12] E. Diciotti, *op. cit.*, 13-14.

[13] P.F. Bresciani, *La protezione dei deboli e vulnerabili come giustificazione costituzionale del reato*, in *Quaderni costituzionali*, 1, 2020, 118-119.

[14] E. Pariotti, *Vulnerabilità ontologica e linguaggio dei diritti*, in *Ars interpretandi*, 2, 2019, 155; M.G. Bernardini, *Vulnerabilità e disabilità a Strasburgo: il vulnerable groups approach in pratica*, in *Ars interpretandi*, 2, 2018, 80.

[15] E. Pariotti, *op. cit.*, 155.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

seconda dei contesti»[16]. La prima accezione di vulnerabilità (comprensiva tanto di definizioni formali[17] quanto di definizioni sostanziali[18]) è stata criticata per la sua inadeguatezza alla formulazione normativa: viene, ad esempio, evidenziato che la predisposizione di un elenco tassativo di caratteristiche rischierebbe di trascurare una molteplicità di situazioni in cui la vulnerabilità si manifesta in modalità differenti rispetto alle peculiarità legislativamente prescelte[19]. Ciò è da ricondursi al carattere «notevolmente indeterminato»[20] del concetto, che possiede «diverse gradazioni interne»[21]. Si distinguono, ad esempio, tre differenti significati[22]: 1) in senso stretto indica l'essere inclini ad un una fragilità; 2) in senso lato denota l'essere *particolarmente* inclini ad una certa fragilità, e si riferisce quindi a persone che hanno una maggiore probabilità di incorrere in una situazione di vulnerabilità; 3) in senso latissimo si riferisce ad una «generica situazione di svantaggio nei confronti degli altri»[23].

Per questo motivo si propende, nella giurisprudenza di Strasburgo[24], per l'accoglimento della seconda accezione di vulnerabilità, chiamata "vulnerabilità ontologica", ovvero una «condizione universale che accomuna tutti gli esseri umani in quanto "esposti alla ferita"»[25]. La speculazione inaugurale delle riflessioni inerenti alla odierna vulnerabilità ontologica fonda le proprie radici nel Novecento, in parte causata dalle esperienze belliche mondiali, in parte spinta dall'interesse dei movimenti femministi

[16] E. Pariotti, *op. cit.*, 155-156.

[17] Così Chenal (R. Chenal, *op. cit.*, 39): «Definire in termini formali consiste, perlomeno secondo un'ottica convenzionale, nel ritenere vulnerabile un soggetto sul mero presupposto che l'ordinamento lo qualifichi come tale in assenza di qualunque riferimento a criteri di natura sostanziale. È l'ipotesi, ad esempio, della definizione per categorie tassative. Si potrebbe stabilire in astratto che la vulnerabilità è propria di alcune (e solo quelle) categorie di soggetti, quali ad esempio le donne, gli immigrati, i disabili, i minori, o di gruppi, quali le minoranze etniche o religiose».

[18] La definizione sostanziale consiste nell'«identificazione degli elementi essenziali o proprietà che un soggetto o gruppo di persone devono possedere per poter essere considerate vulnerabili. Si potrebbe ipotizzare di considerare come tali i soggetti che si trovano in una situazione di dipendenza, quali ad esempio coloro che sono privati della libertà personale o affidati alla custodia, potestà, vigilanza, controllo, cura o assistenza di altre persone» (R. Chenal, *op. cit.*, 42).

[19] R. Chenal, *op. cit.*, 42.

[20] E. Diciotti, *op. cit.*, 16.

[21] R. Chenal, *op. cit.*, 46.

[22] E. Diciotti, *op. cit.*, 14-17.

[23] Loc. ult. cit. Altre gradazioni sono proposte da P.F. Bresciani, *op. cit.*, 111-112.

[24] Le pronunce della Corte EDU non sono, tuttavia, prive di eccezioni: si considera, ad esempio, giurisprudenza consolidata il principio per il quale «nel caso in cui si tratti di effettuare una restrizione dei diritti fondamentali in capo a gruppi particolarmente vulnerabili (intendendosi come tali quelli che hanno sofferto una storica discriminazione), il margine di apprezzamento degli Stati è ristretto» (M.G. Bernardini, *op. cit.*, 90). In questo modo, quindi, la vulnerabilità è ancora considerata una "concezione essenzialista", per la valutazione della quale è presa in considerazione la circostanza per cui un certo gruppo sia stato storicamente sottoposto ad una particolare discriminazione.

[25] L. Re, *Vulnerabilità e cura nell'orizzonte dello Stato costituzionale di diritto*, cit., 183. L'Autrice riprende l'espressione di Adriana Cavarero in *Inclinazioni. Critica della rettitudine* del 2013. Adriana Cavarero, sostenitrice della vulnerabilità ontologica (insieme a Judith Butler), ha introdotto in Italia agli inizi degli anni Duemila studi legati alle tesi di Autori e Autrici del Novecento considerate pioniere delle riflessioni odierne sulla vulnerabilità, come Simone Weil, Hannah Arendt, Emmanuel Lévinas e Paul Ricoeur.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

degli anni Settanta[26]. A partire da questo periodo storico è stata percepita l'esigenza di discostarsi dall'astrattezza della speculazione teorica e di volgere invece l'attenzione alla concretezza dell'esperienza: in questo modo, è stata proposta «una ontologia fondata sulla relazione e, conseguentemente, una nuova concezione dell'autonomia»[27]. Il paradigma della vulnerabilità (e l'etica della cura connessa a tale concetto) «ha dovuto mettere radicalmente in discussione la struttura del pensiero morale e politico di matrice liberale, e la sua assunzione di autonomia da parte degli esseri umani individuali. *Humans are not fully independent*»[28].

Questa seconda accezione di vulnerabilità rifiuta, pertanto, il "mito dell'autonomia"[29] tipico dell'individualismo moderno e del pensiero liberale, che ne risulta profondamente ridimensionato: ciò ha condotto alcune Autrici a parlare di "autonomia relazionale"[30] proprio ad indicare non la mera possibilità, ma la necessità, stante i presupposti ontologici menzionati, di inserire l'autonomia della persona in un contesto interrelato con diverse soggettività[31], in cui autonomia ed eguaglianza non siano considerati come termini antitetici ma vicendevolmente connessi[32].

---

[26] L. Re, *Politica e istituzioni al tempo del cambiamento climatico. Il paradigma della vulnerabilità come proposta di trasformazione*, cit., 295; T. Casadei, O. Giolo, S. Pozzolo, L. Re, *Introduzione. Dalla istituzionalizzazione della critica di genere alla costruzione di una società inclusiva: questioni e sfide per la filosofia del diritto*, in *Rivista di filosofia del diritto*, 2, 2022, 289-290.

[27] L. Re, *Politica e istituzioni al tempo del cambiamento climatico. Il paradigma della vulnerabilità come proposta di trasformazione*, cit., 295. L'autonomia è oggi solitamente considerata connessa con la libertà (positiva), almeno secondo la lettura che ne offre Andronico, riprendendo il pensiero di Isaiah Berlin. L'Autore ricorda il berliniano distinguo tra "libertà negativa", descritta come un senso di indipendenza, cioè, letteralmente, non dipendenza (da altro), di assenza di vincoli esterni; "libertà positiva" è invece intesa come autonomia, come possibilità di fare ciò che si vuole. Cfr. A. Andronico, *Libertà. La legge come misura*, in A. Andronico, T. Greco, F. Macioce (a cura di), *Dimensioni del diritto*, Torino, 2019, 115-116.

[28] G. Zanetti, *L'etica della cultura e i diritti*, in *Ragion pratica*, 2, 2004, 252, corsivo dell'A.; *ibidem*, 528.

[29] L'espressione è riconducibile al testo di M.A. Fineman dal titolo *The Autonomy Myth. A theory of dependency*, del 2004. «Si tratta di un mito che interpreta l'autonomia in termini molto ristretti, collegandola all'autosufficienza economica e "a un senso di separazione dagli altri all'interno della società"», L. Re, *Vulnerabilità e cura nell'orizzonte dello Stato costituzionale di diritto*, cit., 186.

[30] E. Pariotti, *op. cit*., 160.

[31] Si noti che questa concezione di autonomia, sotto un profilo storico, non risulta una novità: i Greci e i Romani intendevano in modo similare la libertà. Essa, infatti, presupponeva una situazione di appartenenza ad una famiglia, ad una stirpe o ad un gruppo, pertanto «solo nell'"essere con gli altri" si poteva davvero "essere se stessi". E dunque essere liberi». A. Andronico, *Libertà. La legge come misura*, in A. Andronico, T. Greco, F. Macioce (a cura di), *Dimensioni del diritto*, Torino, 2019, 118.

[32] Scrive l'Autrice: «La mia tesi è che né l'eguaglianza né l'autonomia possono essere comprese separandole l'una dall'altra, mentre nella società sembra che l'una sia enfatizzata o privilegiata a scapito dell'altra» (M.A. Fineman, *op. cit.*, 158; similmente anche Casadei, «si tratta[, piuttosto,] di muoversi entro un nuovo equilibrio che riconosce l'ineliminabile tensione, ma al contempo anche la possibile convivenza, tra eguaglianza e diversità, tra universalità e contesti, tra una artificialità necessaria e una realtà che non può essere cancellata o occultata», T. Casadei, *Diritti umani in contesto: forme della vulnerabilità e "diritto diseguale"*, cit., 310. Si impone, a tal riguardo, un'ulteriore precisazione. Fineman ha cura di specificare che «[q]uesta non vuole essere un'affermazione in favore dell'uguaglianza dei risultati. Non ignoro, né nego che vi siano differenze nelle abilità o nell'iniziativa individuale, né respingo del tutto l'idea che gli individui siano responsabili di se stessi e delle proprie condizioni di vita» M.A. Fineman, *op. cit*., 39.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Ecco che quindi, facendo sunto di tali indicazioni, in un recente contributo Baldassare Pastore ha proposto, come tratti essenziali della vulnerabilità, i seguenti:

1. La vulnerabilità «si lega alla dimensione della corporeità, che ne costituisce la radice intrascendibile»[33]. Rifiutando il "mito dell'autonomia" «tale dato, però, ha bisogno di essere percepito e riconosciuto, riportando il soggetto alla consapevolezza del suo costitutivo legame con gli altri»[34];
2. La vulnerabilità è un concetto relazionale, composto tanto di una dimensione 'ontologica' quanto di una dimensione 'situazionale'[35];
3. La vulnerabilità è un concetto strettamente connesso al riconoscimento dei diritti umani, poiché essi si preoccupano di garantire quelle possibilità sociali ed economiche per prevenire o gestire una situazione di vulnerabilità[36];
4. La vulnerabilità presuppone una ricostruzione fattuale che, tenendo insieme le dimensioni 'ontologica' e 'situazionale', permetta di discernere, di caso in caso, l'*ubi consistam* dell'essere vulnerabili, in una attiva valutazione del caso di specie che avviene in presenza di giudizi di valore[37].

È possibile quindi concludere che, in base a quanto emerso, la vulnerabilità designa una «una categoria euristica che apre ad una questione di senso, rinviando alla comprensione delle 'cose umane'. È espressione della finitezza e della fragilità proprie degli esseri umani, ma è vissuta nella concretezza esistenziale e risulta influenzata da molteplici specifici fattori»[38].

3. Realtà e vulnerabilità: l'esempio del dolore

L'analisi condotta nel paragrafo precedente ha cercato di fornire un resoconto di alcuni dei tratti essenziali riconducibili al paradigma della vulnerabilità da una prospettiva filosofico-giuridica. L'attenzione sarà ora volta alla prima e seconda caratteristiche elencate, nel tentativo di comprendere quali siano i canali epistemici tramite i quali poter comprendere quando un soggetto versi in una situazione di vulnerabilità e che tipo di realtà essi rilevino. Qual è, dunque, la posizione filosofica sullo sfondo della comprensione dello stato di vulnerabilità?

[33] B. Pastore, *Vulnerabilità, diritto, ragionamento giuridico*, in *Teoria e storia del diritto privato*, n. speciale, 2022, 2; in termini simili anche L. Corso, *op. cit.*, 57; S. Zullo, *Lo spazio sociale della vulnerabilità tra "pretese di giustizia" e "pretese di diritto". Alcune considerazioni critiche*, in *Politica del diritto*, 3, 2016, 475-476.

[34] B. Pastore, *op. cit.*, 6.

[35] B. Pastore, *op. cit.*, 3-5, 11; sulla dimensione 'situazionale' anche R. Chenal, *op. cit.*, 38 e 50, allorquando precisa che «[l]a Corte effettua un bilanciamento tra i diritti che sono in gioco, in base alle specificità del caso, delle circostanze fattuali, degli argomenti sollevati dalle parti. È sufficiente che uno di questi elementi muti perché la Corte pervenga a una conclusione differente»; conferma la nozione relazionale anche M.G. Bernardini, *Vulnerabilità e disabilità a Strasburgo: il vulnerable groups approach in pratica*, in *Ars interpretandi*, 2, 2018, 84; L. Corso, *op. cit.*, 57; L. Re, *Vulnerabilità e cura nell'orizzonte dello Stato costituzionale di diritto*, cit., 184 scrive: «La vulnerabilità che sperimentiamo dipende dunque dalle reti di protezione di cui possiamo o non possiamo avvalerci».

[36] B. Pastore, *op. cit.*, 6; vengono così confermate le suggestioni di Fineman e Nussbaum, in particolare quando Pastore scrive, nello stesso luogo, che «[v]i sono, infatti, beni essenziali per ogni essere umano che è non possibile manomettere, violare, calpestare, senza compiere un torto. In questo senso, i diritti umani costituiscono criteri di giustizia, legati all'aspettativa della eliminazione della sofferenza socialmente prodotta».

[37] B. Pastore, *op. cit.*, 16

[38] B. Pastore, *op. cit.*, 4.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Come è stato posto in luce, molteplici aspetti sono da tenersi in considerazione per individuare una situazione siffatta: ad esempio, il legame della vulnerabilità con la corporeità, la quale diviene un peculiare accesso epistemico, unicamente umano, che permette una certa relazione tra essere umano e mondo[39]. Oppure, la seconda caratteristica ricordata, relativa al carattere ontologico e situazionale della vulnerabilità, richiede una sorta di immedesimazione del soggetto giudicante nei confronti del soggetto giudicando. Come viene infatti notato, la vulnerabilità ha un duplice aspetto: il primo, "oggettivo", per cui il danno che affligge il soggetto vulnerabile è manifesto; il secondo, "soggettivo"[40], per cui è opportuno chiedersi quali siano le difficoltà che la persona vive in certo momento e come siano da questa percepite. Si consideri, ad esempio, il dolore (che è stato considerato criterio di vulnerabilità dalla Corte Costituzionale italiana[41]), che può assumere diverse manifestazioni nella vita di ciascuno. Come identificarlo in sede giudiziale? E, preliminarmente, che tipo di realtà è coinvolta? La prima domanda concerne il carattere epistemologico[42] della vulnerabilità; la seconda riguarda l'aspetto ontologico[43]. Cercheremo nei sottoparagrafi che seguono di indagare tali questioni.

[39] Hart avrebbe considerato tale caratteristica come rientrante nelle «"ovvie verità" della natura umana che dobbiamo dare per scontate». Tra queste, secondo la lettura di Villa, rientra proprio «il fatto che essi [scil. gli esseri umani] sono *vulnerabili*», includendo così anche la vulnerabilità ontologica; Villa invece le chiamerebbe «presupposizioni su "come è fatto il mondo"» (V. VILLA, *Disaccordi interpretativi profondi. Saggio di metagiurisprudenza ricostruttiva*, Torino, 2017, 200, corsivo dell'A.).

[40] Il distinguo è di L. CORSO, *op. cit.*, 62-63; l'Autrice precisa che tale aspetto soggettivo ha condotto la Corte di Strasburgo ad attribuire rilevanza non soltanto all'"essere vulnerabili" ma anche al "sentirsi vulnerabili".

[41] L. CORSO, *op. cit*., 68.

[42] Nella spiegazione che segue viene mantenuta a fini di chiarezza un distinguo tra epistemologia ed ontologia: tuttavia, le due aree di sapere sono strettamente connesse l'una all'altra, come emergerà *infra*. Il modo tramite cui una certa 'entità' può essere conosciuta ne determina anche l'esistenza.

[43] Si precisa che utilizzeremo il termine "ontologia" e il relativo aggettivo senza una marcata distinzione semantica dal termine "metafisica" ed il suo aggettivo. La scelta è giustificata da due circostanze. In primo luogo, pur designando oggi questioni diverse, il distinguo concettuale è di conio piuttosto recente. Secondo quanto riporta l'enciclopedia Treccani, infatti, "ontologia" viene utilizzato per la prima volta all'inizio del XVII secolo dai filosofi Jacob Lorhard e Rodolfo Goclènio e successivamente divulgato in particolare da Christian Wolff (https://www.treccani.it/enciclopedia/ontologia/, ultima consultazione 22/06/2024). Sembra che le principali linee definitorie tra "ontologia" e "metafisica" inizino a delinearsi nel corso del Novecento, specie grazie ai contributi di Quine, Carnap, Russell e Strawson (cfr. A. VARZI, *Ontologia e metafisica*, in F. D'AGOSTINI, N. VASSALLO (a cura di), *Storia della Filosofia Analitica*, Torino, 2002, 81-117). Le definizioni di "ontologia" e "metafisica" sono oggi molteplici: seguendo le indicazioni di Varzi, si ricorda che «la metafisica – secondo una definizione diffusa alla quale ci atterremo – si occupa fondamentalmente della natura ultima di tutto ciò che esiste, attiene alla metafisica anche il compito preliminare di stabilire che cosa esiste, o quantomeno di fissare dei criteri per stabilire che cosa sia ragionevole includere in un accurato inventario del mondo. La messa a punto di tali criteri definisce, appunto, la questione ontologica» (A. VARZI, *op. cit*., 82). Pertanto, «[b]y 'ontology' here, I mean the inquiry into *what there is*, and by 'metaphysics' the inquiry into *what it is*» (F. FRANDA, *On Whether It Is and What It Is*, in *Acta Analytica*, 2023, 4, corsivo dell'A.; si ringrazia l'Autore per la condivisione di questo contributo). In secondo luogo, la domanda ontologica e la domanda metafisica, pur analiticamente distinguibili, sembrano in taluni casi coimplicanti. Scrive Franda: «[I]n some cases *we have to* characterize our entities in a metaphysical sense in order to be able to discuss them from an ontological point of view. And here I mean 'we have to' in a stronger sense than the methodological one» (F. FRANDA, *op. cit*., 8, corsivo dell'A.). Si configurano così delle eccezioni alla c.d. 'priority thesis' che può essere definita, molto succintamente, come quella posizione filosofica che ammette la possibilità dell'esistenza di 'qualcosa' senza doversi impegnare a stabilire che cosa quel 'qualcosa' sia esattamente; in questo modo, si sostiene che la domanda ontologica sia prioritaria rispetto a quella metafisica (cfr. la

Downloaded from www.biodiritto.org.
ISSN 2284-4503

3.1. Che tipo di realtà: il problema ontologico

Il problema in esame rientra all'interno del "realismo ontologico", una categoria filosofica che designa una molteplicità di questioni sottese alle seguenti domande: «ci si può chiedere se una determinata cosa esista veramente oppure, concedendo che esista, ci si può chiedere se essa esista indipendentemente dalle menti che la pensano»[44]. La prima domanda concerne, ad esempio, l'esistenza degli atomi[45]; la seconda domanda, invece, riguarda, a titolo esemplificativo, l'esistenza dei colori[46]. Orbene, presupponendo un accordo sull'esistenza del dolore (come il diritto presuppone: alcune tipologie di danno morale sono infatti classificate come *pretium doloris*), l'analisi si concentrerà sulla seconda domanda; il dolore esiste indipendentemente dalle menti che lo pensano? La questione non suscita interesse soltanto da un punto di vista teoretico, ma assume una concreta rilevanza pratica: se l'esistenza del dolore restasse relegata nella mente di chi lo prova, si cadrebbe in una forma di solipsismo, non intellegibile dal giudicante, il quale non potrebbe discernere lo stato di vulnerabilità. Allo stesso tempo, tuttavia, il dolore non può considerarsi al pari di un'"entità astratta", cioè quelle «entità che per definizione non hanno una collocazione spaziotemporale, come gli universali, i numeri, gli insiemi, le specie e i significati»[47]. Né tantomeno parrebbe assimilabile agli "oggetti sociali" menzionati da Ferraris, come, ad esempio, un debito, oppure agli "oggetti naturali", come una montagna o un fiume[48]. Il dolore è piuttosto una "capacità"[49] profondamente incarnata. Esso «sortisce un effetto di reale. Noi percepiamo la realtà soprattutto a partire dalla resistenza, che provoca dolore»[50] al punto che esso viene descritto come un «affidabile criterio di verità»[51]; esso possiede una qualche forma specifica di concretizzazione in un certo spazio e tempo, tuttavia è comprensibile in maniera universale.

lettura congiunta di A. VARZI, *On Doing Ontology Without Metaphysics*, *passim* e F. FRANDA, *op. cit.*, *passim*). Un esempio di questa co-implicazione (e dunque di eccezione rispetto alla priority thesis) è fornito dai mondi possibili, per i quali «the ontological commitment takes place thanks to the attribution of a metaphysical property: philosophers posit the existence of possible worlds because, in their view, they are what makes modal sentences true or false». (F. FRANDA, *op. cit.*, 11). Al fine di evitare qualsiasi rischio di riduzionismo (storico o concettuale), "ontologia" viene utilizzato come termine inclusivo e dunque comprensivo della domanda metafisica.

[44] M. DE CARO, *Realtà*, Torino, 2020, 17.

[45] M. DE CARO, *op. cit.*, 17; lo stesso potrebbe dirsi per l'esistenza dell'elettrone: la questione è stata oggetto di dibattito nell'ambito della filosofia della scienza. È infatti possibile misurare proprietà che è ragionevole attribuirsi all'elettrone, ma esso non si incontra mai come "cosa". «Perché [non] dovremmo allora essere autorizzati a parlare di esso come qualcosa di realmente esistente senza evidenza percettiva della sua esistenza?», E. AGAZZI, *L'oggettività scientifica e i suoi contesti*, Milano, 2018, 444-445.

[46] Non è questo il luogo per approfondire tale inciso, ma si noti che l'esempio dei colori è stato anche utilizzato al fine di spiegare l'esistenza di valori: cfr. A. VARZI, *I colori del bene*, cit., 53 ss.

[47] M. DE CARO, *op. cit.*, 19.

[48] M. FERRARIS, *Manifesto del nuovo realismo*, Bari, 2012, 71 ss.

[49] B.C. HAN, *La società senza dolore. Perché abbiamo bandito la sofferenza dalle nostre vite*, Torino, 2021, 50.

[50] B.C. HAN, *op. cit.*, 41.

[51] B.C. HAN, *op. cit.*, 39. Han lo ritiene anche criterio di felicità: nel suo scritto infatti critica la società "palliativa" contemporanea che, nella pretesa di eliminare il dolore, incoraggia una "società della sopravvivenza" perdendo di vista la vita buona (*ibidem*, 19-23). L'Autore esorta, a contrario, ad accogliere, interrogarsi (sul) e comprendere il dolore, nella convinzione che «[l]a profonda felicità resta inaccessibile a chi non è aperto al dolore» (*ibidem*, 17).

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Un possibile statuto ontologico per il dolore è rinvenibile nella teoria del soggettivismo sofisticato proposta da Porciello[52]. Secondo l'Autore, lo stato d'animo di meraviglia che sorprende l'essere umano di fronte al sublime della natura non è disponibile al soggetto, il quale non può scegliere se meravigliarsi o meno: a ciò, tuttavia, non consegue che la bellezza della natura possa darsi al di fuori dello sguardo dell'essere umano che osserva, ad esempio, il tramonto sul mare. Pertanto, essa non potrà predicarsi "mind-independent", giacché accade implicando una forma di dipendenza con il soggetto osservante. Ma tale forma di dipendenza, che l'Autore chiama "relazione", è comprensibile ed ipotizzabile da chiunque, e pertanto, pur acquisendo concretezza nel momento in cui il soggetto si trova di fronte al mare ed osserva il tramonto, può essere universalmente compresa. Questo è il motivo per cui l'Autore parla di una relazione "oggettiva"[53], poiché la capacità di stupore non è dispensabile, cioè non è disponibile alla modifica da parte del soggetto, che non può scegliere se meravigliarsi o meno. Ecco che quindi, sulla scorta di tale lettura, si scopre che lo stupore e il dolore sembrano possedere la stessa struttura esistenziale: la loro esistenza implica una forma di dipendenza con l'esistenza dell'essere umano. Calando queste riflessioni nella questione in esame, è plausibile quindi sostenere che senza l'essere umano alcune forme di dolore (come, ad esempio, lo stato d'animo di discriminazione) non esisterebbero: ed è quindi ragionevole concludere che la stessa vulnerabilità – se intesa come esperienza in qualche modo dolorosa – non potrebbe trovare luogo.

3.2. Come conoscere quel tipo di realtà: il problema (onto-)epistemologico

Se si conviene sul suddetto tipo di esistenza del dolore, è opportuno comprendere come esso possa essere conosciuto da un soggetto altro rispetto a colui/colei che prova dolore. La questione verrà indagata attraverso il "conflitto dei realismi"[54] che vede contrapposto il realismo ordinario al realismo scientifico, due versioni del realismo ontologico. Il realismo ordinario, di provenienza aristotelica, «attribuisce realtà esclusivamente alle cose di cui possiamo avere esperienza»[55]: per i realisti ordinari

[52] In particolare, riprendendo gli studi di Arne Naess, A. Porciello, *Filosofia dell'ambiente. Ontologia, etica, diritto*, Roma, 2022, 90 ss.; Id., *Una giustificazione metaetica del valore intrinseco della natura: il soggettivismo sofisticato (una variante)*, cit., 219-247.

[53] Si ricorda che anche Enrico Opocher, interrogatosi circa lo statuto ontologico dei valori, individua un preliminare livello "soggettivo" ed una conseguente valenza "oggettiva". Prediligendo una terza posizione tra all'immanentismo dei valori (secondo cui i valori sono storicamente giustificati) ed il giusnaturalismo dei valori (per il quale i valori sono fondati su ideologie), Opocher conduce la propria indagine a partire dall'esistenza umana. In tal modo è possibile comprendere che cosa sia il valore. Sia l'esistenza umana che il valore posseggono, infatti, la medesima struttura: l'esistenza, mediante un processo che Opocher chiama di "oggettivazione", tende a sottrarsi dal nulla prodotto dal vuoto di senso per acquisirne uno. Anche in tal caso, dal soggettivismo dell'esistenza si giunge all'oggettivismo del valore. Per una spiegazione più esaustiva, di cui qui sono riportati i soli passaggi essenziali, si v. M. Manzin, *op. cit.*, 97-101.

[54] M. De Caro, *op. cit.*, 33; S. Bonicalzi, *Naturalismo liberalizzato e altri realismi. La proposta teorica di Mario De Caro*, in *Iride*, 3, 2021, 712-716.

[55] M. De Caro, *op. cit.*, 18. Il realismo ordinario spiegato dall'Autore implica due forme di esperienza: l'esperienza diretta implica introspezione o sensi; l'esperienza indiretta ricorre invece a «strumenti che estendono i sensi, come microscopi e telescopi» (loc. ult. cit.). Non assume rilevanza per il distinguo che traccia l'Autore (che si ringrazia per il chiarimento), ma si ricorda che l'esperienza indiretta implicata dal realismo ordinario, poiché condotta tramite un qualche forma di strumento tecno-scientifico, implica confini molto labili con il realismo scientifico (sul punto, cfr. F. Russo, *Techno-Scientific Practices. An Informational Approach*, Londra, 2022, *pas-*

Downloaded from www.biodiritto.org.
ISSN 2284-4503

sono, quindi, percezione e senso comune a «determinare l'ambito ontologico»[56]. Viceversa, il perimetro ontologico per il realismo scientifico è definito dalla "scienza naturale"[57]: questa versione del realismo ontologico – galileiana[58] e, in origine, platonica – attribuisce alla fisica il ruolo di «scienza fondamentale, perché tutte le altre scienze sono ad essa riducibili»[59]. A ciò consegue che il realista ordinario sarà antirealista nei confronti di «entità inosservabili postulate dalla scienza naturale, come gli elettroni, le radiazioni e i buchi neri»[60]; essi non potranno rientrare nel catalogo ontologico del mondo proprio perché i sensi non rilevano l'esistenza di tali entità. D'altro canto, il realista scientifico sarà antirealista con riferimento a tutto ciò che «non possa, in linea di principio, essere indagato con metodi e concetti propri delle scienze naturali»[61].

Orbene, le posizioni del realismo ordinario e del realista scientifico, singolarmente considerate, sono criticabili giacché ambedue pretendono di possedere il "monopolio dell'ontologia"[62] che si traduce nella convinzione che «noi disponiamo di un'unica chiave di accesso epistemico alla realtà»[63]. A partire dalla nascita della scienza moderna sino al culmine del naturalismo radicale di Quine[64], il realismo scientifico ha goduto di particolare successo; di esso si rinvengono tracce anche nel contesto giuridico, ad esempio nella "matematizzazione della logica" inaugurata da Leibniz[65] e nei processi di codificazione del Novecento che, presupponendo un ordine a "narrazione continua", hanno tradotto il principio di causalità nell'impostazione sistematica del codice[66]. È significativo notare sin d'ora, per i

sim).

[56] M. De Caro, *op. cit.*, 24.

[57] M. De Caro, *op. cit.*, 39.

[58] Scrive De Caro: «[L]'affermazione del suo modo [di Galilei] di concepire la scienza e il mondo naturale segnò il trionfo di una concezione di matrice platonica, che riguardava alla fisica in prospettiva realistica e matematizzata e riteneva che il mondo fosse composto esclusivamente da proprietà geometriche». Un elaborato *ad hoc* sarebbe necessario per indagare nel dettaglio la posizione galileiana: si basti qui ricordare la sintesi di P. Musso, *Techne e conoscenza nella modernità*, in M. Ferrari (a cura di), *Logos e Techne*, Milano-Udine, 2017, 53. L'Autore facendo sunto delle varie formulazioni del metodo galileiano, elenca i seguenti quattro principî: «1) Non cercare l'essenza delle cose, ma limitarsi a studiare alcune proprietà. 2) Non solo generica osservazione, ma esperimento. 3) Uso della matematica. 4) Nessun principio di autorità».

[59] M. De Caro, *op. cit.*, 19.

[60] M. De Caro, *op. cit.*, 35.

[61] M. De Caro, *op. cit.*, 43. Semplificando la questione, il conflitto è, storicamente ed in via esemplificativa, evidente nelle posizioni di Edmund Husserl e Wilfrid Sellars. Mentre il primo «è un realista ordinario e un antirealista rispetto alla scienza, Sellars adotta la prospettiva opposta: è cioè realista rispetto alla visione scientifica e antirealista rispetto alla visione ordinaria del mondo» (*ibidem*, 40).

[62] L'espressione è di F. Eustacchi, *M. De Caro, Realtà*, in *Bollettino della società filosofica italiana*, 1, 2022, 96.

[63] M. De Caro, *op. cit.*, 69.

[64] M. De Caro, *op. cit.*, 41; S. Bonicalzi, *op. cit.*, 713-714, che sintetizza così le tre tesi chiave del naturalismo radicale di Quine: «La prima [tesi], ontologica, comune anche al realismo scientifico non radicale, afferma che sono reali solo le entità e le proprietà spiegabili dalle scienze naturali o a esse riconducibili. La seconda, epistemologica, stabilisce che solo la conoscenza scientifica, o le forme di conoscenza a essa riconducibili, è una forma valida di conoscenza. La terza, metafilosofica, afferma che la filosofia si sviluppa in continuità con la scienza per contenuti, metodi e scopi».

[65] F. Puppo, *Diritto e retorica*, Torino, 2023, 15-17.

[66] M. Manzin, *op. cit.*, 31-32. Si precisa che l'idea sottesa al principio di causalità all'epoca era quella per cui ad *un* effetto era riconducibile *una* causa; fallacia oggi smascherata anche nel contesto della filosofia della scienza

Downloaded from www.biodiritto.org.
ISSN 2284-4503

propositi di questo scritto, che il realismo scientifico, in una sua versione particolare, quella del realismo strutturale – che verrà chiarito *infra* –, funge oggi da teoria filosofica di sfondo per i sistemi di intelligenza artificiale, almeno secondo l'approccio informazionale supportato da Luciano Floridi[67].

Il realismo scientifico, tuttavia, si scontra con quello che De Caro chiama "problema della collocazione": vi sono cioè «delle entità e delle proprietà (proprietà secondarie, libero arbitrio, coscienza, valori e così via) che tanta importanza hanno per la visione ordinaria del mondo ma che, almeno apparentemente, non sembrano trattabili dalle scienze naturali»[68]. Tali entità e proprietà non sono dunque esaustivamente "collocabili" né all'interno di una spiegazione guidata dal senso comune e dalla percezione né all'interno di una teoria (neuro)scientifica[69]: per questo motivo viene proposto il "naturalismo liberalizzato", «una terza forma di realismo caratterizzata da pluralismo ontologico, epistemologico e causale»[70]. Allo stesso tempo, tuttavia, il naturalismo liberalizzato sostiene che «si possono anche accettare come reali entità che sono implicite nelle altre pratiche epistemiche solide e coronate da successo (come senso comune e le scienze umane e sociali) nella misura in cui tali entità non sono incompatibili con la concezione del mondo propria delle scienze della natura»[71]. In altre parole, «[s]i tratta di una posizione che prende atto dell'infinita varietà della realtà, nonché della sua eccedenza rispetto ai nostri schemi conoscitivi»[72].

Il proposito sotteso al naturalismo liberalizzato, ovvero l'intento di includere molteplici e diversificati accessi epistemici nel nostro catalogo epistemologico (e, di conseguenza, ontologico), sembra trovar riscontro, con i dovuti distinguo e pur in mancanza di un richiamo esplicito, in alcune voci della filosofia giuridica recente. Ci riferiamo qui alle proposte, diverse, di Jori e Villa, che sono però accumunate dal richiamo al senso comune.

(si v. F. Russo, *L'esposizione all'amianto causa il mesotelioma? Domande scientifiche e analisi filosofiche*, cit., 219).

[67] L. Floridi, *A defence of informational structural realism*, in *Synthese*, 161, 2008, 219-253, che leggiamo alla luce di F. Russo, *Techno-Scientific Practices. An Informational Approach*, cit., 220 ss.

[68] M. De Caro, *op. cit.*, 59.

[69] Il terzo capitolo del testo di De Caro considera come caso studio il libero arbitrio: spiega l'Autore che né le ricerche neuroscientifiche (inaugurate da Benjamin Libet negli anni Settanta e sviluppate da Chun Siong Soon e colleghi) né tantomeno la "sfida epifenomenistica" (secondo la quale gli stati coscienti non avrebbero potere causale nelle azioni del soggetto) sono in grado di spiegare esaustivamente se il libero arbitrio esista. Allo stesso tempo, i risultati ottenuti da questi studi non possono essere ignorati dalla speculazione filosofica sul tema. Proprio per questo motivo il naturalismo liberalizzato si mostra una chiave di lettura adeguata della realtà, poiché in grado di coniugare scienza e filosofia: viene così restituita un'antropologia che permette di «pensare gli esseri umani, allo stesso tempo, come agenti liberi e come agenti naturali. Nel primo senso, apparteniamo alla sfera normativa dello spazio delle ragioni; nel secondo senso, alla sfera della legalità naturale» (M. De Caro, *op. cit.*, 114).

[70] S. Bonicalzi, *op. cit*., 712. Pur essendo una forma di realismo più 'inclusiva' del realismo scientifico, è opportuno notare che essa non sfocia nel soprannaturalismo, poiché «liberal naturalists are committed to accept the constitutive claim of naturalism, according to which no entity or explanation should be accepted whose existence or truth would contradict the laws of nature, insofar as we know them» M. De Caro, A. Voltolini, *Is Liberal Naturalism Possible?*, in M. De Caro, D. Macarthur (a cura di), *Naturalism and Normativity*, Cambridge, 2010, 75. Sul perché il pluralismo causale non sia una mera opzione, cfr. F. Russo, *L'esposizione all'amianto causa il mesotelioma? Domande scientifiche e analisi filosofiche*, cit., 219 ss.

[71] M. De Caro, *op. cit.*, 69.

[72] F. Eustacchi, *op. cit.*, 97.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Secondo Jori è il senso comune, infatti, ad individuare il diritto vigente, permettendo così di distinguere un «matto che faceva partire i treni»[73] dal capostazione. In altre parole, ogni «formidabile apparato di dottrina non disporrà di nessun criterio per *individuare* quale sia il diritto vigente e quindi per scegliere tra due o più diritti che si pongono come rivali»[74]. Solo a seguito di tale individuazione il «diritto può essere determinato e descritto [solo] con strumenti tecno-giuridici»[75], da parte della giurisprudenza e metagiurisprudenza (per questo, nel testo di Jori, *descrittiva*). Il senso comune sembrerebbe assumere un ruolo ancor più pregnante nelle opere di Villa, il quale riconduce alla giurisprudenza e alla metagiurisprudenza non un mero ruolo descrittivo, bensì "ricostruttivo"[76]: il senso comune non resta, per Villa, relegato all'individuazione del diritto vigente ma penetra inevitabilmente lo strato "tecno-giuridico" della giurisprudenza e della metagiurisprudenza, le quali manifestano un apporto costruttivo nell'interpretazione della disposizione legislativa, nel primo caso, e nell'analisi della giurisprudenza, nel secondo. Ciò è reso possibile grazie agli schemi concettuali, che si pongono ad uno stadio antecedente tanto del giuridico quanto del metagiuridico, e dunque anche alla base dell'individuazione del diritto vigente. Tali schemi concettuali sono trascendentali, ovvero presupposizioni che riguardando «"come siamo fatti noi in quanto abitanti del mondo"»[77], che intrattengono legami con le assunzioni empiriche ma non si esauriscono in esse. Inoltre – per ciò che pertiene specificamente al confronto con il naturalismo liberalizzato –, gli schemi concettuali si compongono non soltanto di concetti di senso comune ma anche di concetti scientifici[78]: queste due tipologie non sono da intendersi come categorie incomunicabili, ma interagenti tra di loro nelle pratiche sociali.

[73] M. Jori, *Del diritto inesistente. Saggio di metagiurisprudenza descrittiva*, Pisa, 2010, 11. L'aneddoto di Jori, in esordio al suo testo, è ambientato alla stazione ferroviaria di Pavia: l'Autore racconta di un signore (il "matto") che era solito alzare e agitare le braccia a seguito di ogni fischio del capostazione, volto a segnalare al treno di poter partire.

[74] M. Jori, *op. cit.*, 52, corsivo dell'A. Il senso comune permette dunque di «sapere cosa è il diritto senza sapere niente del diritto» (*ibidem*, 24). Secondo Jori inoltre «il senso comune collega indubbiamente l'esistenza del diritto all'esistenza di una pratica condivisa, a un qualche tipo di *accettazione* collettiva» (*ibidem*, 25, corsivo dell'A.), che comprende "indizi" del diritto vigente. Ad esempio, «[l]a presenza di insegne, uniformi e distintivi non è mai menzionata dalla teoria del diritto, ma è caratteristica del modo in cui il senso comune individua quell'elemento basilare e antichissimo del diritto che sono le (persone dotate di) autorità» (*ibidem*, 33). Sul ruolo della metagiurisprudenza e della scienza giuridica cfr. *ex multis*, N. Bobbio, *Essere e dover essere nella scienza giuridica*, in *Rivista di filosofia*, 58, 1967, 239-240; R. Guastini, *Bobbio sulla scienza giuridica. Introduzione alla lettura*, in *Saggi sulla scienza giuridica di Norberto Bobbio*, Torino, 11.

[75] M. Jori, *op. cit*., 74.

[76] V. Villa, *op. cit.*, *passim*.

[77] V. Villa, *op. cit*., 202; cfr. anche Id., *Costruttivismo e teorie del diritto*, cit., 23-34.

[78] V. Villa, *Costruttivismo e teorie del diritto*, cit., 14-15. I concetti di senso comune «esprimono complessivamente la visione del mondo sulla quale i membri laici di una determinata comunità sociale, o, meglio ancora, di più comunità tra loro affini (le comunità che condividono forme di vita di tipo "occidentale", o comunque ad esse assimilabili, costituiscono, grosso modo, il campo di riferimento di queste osservazioni), fanno affidamento nelle loro attività quotidiane» (*ibidem*, 23); i concetti scientifici, invece, esprimono il «contenuto di tutte quelle credenze, di tipo sostanziale [inerenti al contenuto delle teorie elaborate in un certo contesto] o semantico [cioè inerenti al significato delle nozioni impiegate, ad es. "elettrone" o "democrazia"], che sono presupposte (implicitamente o esplicitamente), in modo assolutamente non problematico, dai membri di una determinata comunità scientifica nel corso delle loro svariate attività di carattere teorico e/o empirico» (*ibidem*, 15). In quest'ultima categoria rientrano i concetti "tecnici", tanto del diritto (ad esempio, "democrazia") quanto della scienza (appunto, "elettrone").

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Per ciò che interessa i propositi di questo elaborato, sembra più che plausibile sostenere che il naturalismo liberalizzato possa candidarsi ad adeguata posizione realista per comprendere quando un soggetto versi in uno stato di vulnerabilità. Come è stato ricordato, oltre a qualificare la vulnerabilità tramite il criterio del danno "oggettivo"[79], un criterio "soggettivo" è complementarmente necessario e parrebbe propedeutico a rilevare proprio quelle caratteristiche che Pastore ricorda nel suo studio[80], come ad esempio la corporeità o le dimensioni 'ontologica' e 'situazionale' della vulnerabilità. Permetterebbe così, per restare nell'esempio proposto, l'immedesimazione e la comprensione del dolore a partire (anche) dalla percezione (che necessita della struttura corporea umana) e del senso comune, ossia assunzioni basilari circa il "come siamo fatti noi in quanto abitanti del mondo"[81]: a partire dalla limitatezza dell'essere umano[82], solo quest'ultimo è in grado di capire se, come e quanto una vicenda personale possa essere dolorosa, fino al punto di farla propria e di viverla in prima persona. Questa capacità presenterebbe, invero, risvolti concreti ben oltre il mero ambito della vulnerabilità: si pensi, ad esempio, all'applicazione di circostanze attenuanti comuni in sede processuale penale. L'art 62, co. 1, punto 2) c.p. consente al giudice di attenuare la pena prevista qualora la persona imputata nel processo abbia agito in stato di ira, a fronte di un altrui fatto ingiusto. È evidente, quindi, che al fine di vagliare l'applicabilità di tale attenuante, percezione e senso comune, implicando immedesimazione, sembrano requisiti necessari proprio per comprendere le circostanze che compongono il *dictat* legislativo, cioè lo stato di ira ed il fatto ingiusto.

Al fine di qualificare come vulnerabile un soggetto, non pare quindi sufficiente collocarlo all'interno di una precisa categoria, come confermano le difficoltà nella predisposizione di elenchi chiusi sul piano legislativo; sarà invece opportuno operare altresì una valutazione sulla base di senso comune e percezione quali criteri di giudizio che, accomunando gli esseri umani, sono comprensibili in maniera universale[83].

---

[79] L. CORSO, *op. cit.*, 62-63, che scrive che «[l]a vulnerabilità ha di certo un aspetto oggettivo, per così dire quantificabile, direttamente proporzionale alla misurabilità del rischio o della sofferenza che il vulnus produce. Sotto questo profilo, come pure afferma la Cedu, anche dati statistici possono venire in soccorso».

[80] Su cui si v. *supra*, in particolare primo e secondo punto nell'elenco.

[81] Secondo Villa trattasi di mere "credenze" e non di "fatti", giacché questi ultimi implicano elaborazione mediante il linguaggio (cfr. V. VILLA, *Disaccordi interpretativi profondi. Saggio di metagiurisprudenza ricostruttiva*, cit., 171). Tuttavia, accogliendo le suggestioni di De Caro, anche i "fatti" di cui parla Villa, reggendosi sugli schemi concettuali, non sono costruiti in modo arbitrario e pertanto implicheranno sempre, a nostro modo di vedere, il fondamento ad una qualche realtà.

[82] Non è questo il luogo per svolgere tale approfondimento ma si segnala che la valenza del limite è indagata con molteplici e diverse sensibilità: *ex multis*, in ambito filosofico-giuridico, V. VILLA, *Costruttivismo e teorie del diritto*, cit., 116 ss. nella spiegazione dei diversi "vincoli" del suo costruttivismo post-positivistico; M. MANZIN, *Reasonableness of Limits, Reasonableness as Limit (in Legal Interpretation)*, cit., 147 ss.; con riferimento all'antropologia aristotelica, M. HEIDEGGER, *Concetti fondamentali della filosofia aristotelica* (= *Grundbegriffe der aristotelischen Philosophie*, ed. it. a cura di G. GURISATTI), 2017, Milano, 66, 75; in ambito di etica ambientale, L. VARELA, *Tecnologia ed ecologia. Dall'etica alla metafisica, dalla negazione del limite alla negazione dell'uomo*, in *Pensamiento*, 71, 2015, 1456 ss.; in ambito tecno-scientifico ne viene fatta menzione in F. RUSSO, *Techno-Scientific Practices. An Informational Approach*, cit., 169-170.

[83] Su temi consimili si è occupato di recente A. LO GIUDICE, *Il dramma del giudizio*, Milano, 2023, *passim*, cui si rimanda per approfondimenti sul tema del giudizio.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

4. Realismo e sistemi di intelligenza artificiale

Accogliere il naturalismo liberalizzato proposto da De Caro permette, come si è sin qui visto, di dar conto di elementi percettivi, come il dolore, che entrano a far parte della comprensione dello stato di vulnerabilità. Allo stesso tempo, consente di evidenziare alcuni limiti dei sistemi di intelligenza artificiale[84] nel comprendere situazioni siffatte, poiché tali sistemi, in virtù della loro natura informazionale, sono sprovvisti proprio di quelle capacità percettive necessarie a rilevare una situazione di vulnerabilità facente capo ad un dato soggetto in un caso concreto. Ad oggi, infatti, i sistemi di intelligenza artificiale disponibili afferiscono a quella che si è soliti chiamare "I.A. debole": tali sistemi sono strumenti utili per una singola funzione circoscritta, ma non posseggono "stati cognitivi" come è invece richiesto per un'intelligenza artificiale 'forte'[85]. Solo quest'ultima sarebbe (forse?) in grado di pensare e comportarsi includendo nel ragionamento senso comune e percezione, necessari per intendere tanto la dimensione 'ontologica' della vulnerabilità quanto quella 'situazionale'.

Ciò è da ricondursi al tipo di realismo che sottace ai sistemi di intelligenza artificiale: seguendo la teoria di Floridi, trattasi del realismo scientifico di tipo strutturale nella sua versione informazionale, di cui riassumeremo brevemente le caratteristiche essenziali[86]. Il realismo strutturale, genericamente inteso, si basa sull'assunto per cui «le nostre migliori teorie fisiche non descrivono la natura intrinseca dei fenomeni inosservabili a cui fanno riferimento, bensì la loro struttura, ossia le relazioni che

[84] Si riprende la definizione di "sistema di intelligenza artificiale" proposta dal "High-Level Expert Group on Artificial Intelligence" (denominato anche "AI HLEG"), nominato dalla Commissione Europea nel giugno 2018, incaricato di predisporre "linee guida per un'intelligenza artificiale affidabile": «Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions» Independent High-Level Expert Group on Artificial Intelligence, *Ethics Guidelines for Trustworthy AI*, 8 aprile 2019 (https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai, ultima consultazione 04/07/2024).

[85] Il distinguo tra "I.A. debole" e "I.A. forte" è originariamente riconducibile a Searle, per il quale «according to strong AI the computer is not merely a tool in the study of the mind; rather, the appropriately programmed computer really *is* a mind, in the sense that computers given the right programs can be literally said to *understand* and have other cognitive states»; viceversa, «[a]ccording to weak AI, the principal value of the computer in the study of the mind is that it gives us a very powerful tool» (J.R. SEARLE, *Minds, Brains and Programmes*, in *The behavioral and brain sciences*, 3, 1980, 417, corsivi dell'A.). Similmente, Sartor definisce "intelligenza specifica artificiale" «tutte le applicazioni di IA oggi disponibili: si tratta di sistemi capaci di ottenere risultati utili in attività che richiedono intelligenza, con prestazioni, in alcuni casi, di livello umano o anche sovrumano. Per esempio, nel riconoscimento di immagini o di volti, l'IA ha già raggiunto prestazioni paragonabili a quelle di un umano esperto; nel gioco degli scacchi, è invece capace di prestazioni sovrumane, superiori a quelle dei migliori giocatori»; "intelligenza generale artificiale" invece «dovrebbe possedere la maggior parte delle abilità cognitive umane, a livello umano, o anche a un livello sovrumano» (G. SARTOR, *op. cit.*, 18).

[86] Un'analisi adeguata del pensiero del realismo strutturale informazionale di Floridi richiederebbe di chiarire altri elementi connessi a questa posizione realista, come ad esempio i presupposti costruzionistici, il metodo dei livelli di astrazione e la filosofia dell'informazione, i quali non possono trovare luogo in questa sede. Si basti notare che il metodo dei livelli di astrazione è orientato ad un fine specifico, quello della creazione di modelli della realtà. Torneremo in chiusura del paragrafo su questo punto.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

questi fenomeni stabiliscono»[87]. Il realismo strutturale (d'ora innanzi anche "RS") si ramifica in RS epistemico ed RS ontico (quest'ultimo si distingue al suo interno in eliminativista e non-eliminativista)[88]: il RS epistemico considera la 'cosa in sé' inaccessibile, ciononostante ritiene possibile conoscere le relazioni che si instaurano tra differenti 'cose in sé' (i c.d. 'noumeni', a voler riprendere l'ispirazione kantiana del RS[89]). Per il realista strutturale epistemico è pertanto possibile cogliere e conoscere i soli aspetti strutturali/relazionali dei noumeni. Per il RS epistemico, dunque, l'"impegno ontologico" ("ontological commitment"; dunque la 'forza' di tale posizione realista) è piuttosto debole. Si riconosce, infatti, la possibilità che qualcos'altro esista oltre alle strutture (cioè la 'cosa in sé', il noumeno): pertanto, l'"oggetto" identificato grazie a tali strutture disponibili non potrebbe mai conoscersi interamente, né le strutture individuate possono elevarsi a rango di "oggetti".

Il problema sembra apparentemente risolto dal realista strutturale ontico eliminativista, il quale «asserisce[, più radicalmente,] che non esistono oggetti inosservabili ma soltanto caratteristiche strutturali»[90]: in questo modo, lo scarto tra epistemologia ed ontologia sembra ricongiunto, giacché l'unità minima della realtà sarebbe composta da sole strutture, conoscibili grazie all'indagine scientifica[91]. Accogliere il RS ontico eliminativista, tuttavia, conduce a delle criticità, non facilmente superabili[92]: per questo motivo, Floridi respinge la posizione dell'eliminativista e propende invece per l'accoglimento del RS ontico non-eliminativista, che ritiene compatibile con il RS epistemico. Il RS ontico non-eliminativista, infatti, assume non soltanto che sia possibile conoscere le strutture dei fenomeni sotto indagine (al pari del realista strutturale epistemico) ma altresì che si possa attribuire unità a tali

[87] M. De Caro, *op. cit.*, 57; cfr. anche R. Ridi, *La piramide dell'informazione e il realismo strutturale*, in *AIB studi*, 61, 2021, 237.

[88] L. Floridi, *op. cit*., 220-223.

[89] Il realismo strutturale epistemico viene, infatti, anche chiamato «"bifurcated structuralism" because of its obvious (and typically Kantian) dualism» L. Floridi, *op. cit*., 222 (l'espressione è di A.E. Heath, *Contribution to the symposium "Materialism in the Light of Scientific Thought"*, in *Proceedings of the Aristotelian Society*, Supplement, 8, 1928, 130-142).

[90] M. De Caro, *op. cit.*, 57.

[91] L. Floridi, *op. cit*., 222. Così riassume Morganti il distinguo tra RS epistemico e RS ontico (eliminativista): «[w]hile the former version states that we can only have a justified realist attitude towards structures but there is, or might be, something more, the latter argues that structures are all there is out there. According to ESR, that is, the intrinsic nature of the physical world exhibiting the relations expressed by the mathematical structures remains hidden. OSR, on the other hand, suggests that there is nothing more in the world than those structures we get to (partially) know through our scientific inquiry» (M. Morganti, *On the preferability of epistemic structural realism*, in *Synthese*, 142, 2004, 82). Si noti che secondo Morganti è possibile conoscere in parte tali strutture; così non sembrerebbe secondo Floridi, per il quale «ontic monism and structural knowledge guarantee that reality is fully knowable in principle» (L. Floridi, *op. cit.*, 223).

[92] Si pongono, *in primis*, dubbi sulla collocazione di tale posizione all'interno del RS stesso: quest'ultimo, infatti, si basa sull'assunto per cui sia possibile conoscere la sola struttura dei fenomeni, ammette che vi sia altro rispetto alla sola struttura. In secondo luogo, postulare che esistano solo strutture e che esse siano conoscibili condurrebbe a supportare una versione 'forte' di realismo scientifico, che invece Floridi intende evitare (e per questo accoglie una posizione costruzionista). Infine, se non fosse possibile identificare "oggetti strutturali" – come Floridi propone accogliendo la posizione RS ontica non – eliminativista – si cadrebbe in una *regressio ad infinitum*: «Relations (structures) require relata (structured/able objects), which therefore cannot be further identified as relations (structures) without running into some vicious circularity or infinite regress» (L. Floridi, *op. cit*., 234).

Downloaded from www.biodiritto.org.
ISSN 2284-4503

strutture, e dunque 'impegnarle' ontologicamente. Per questo motivo, si accetta che le strutture conoscibili compongano un "oggetto strutturale". La particolarità dell'approccio informazionale di Floridi consiste nel fatto che tali aspetti relazionali, conoscibili e ontologicamente 'impegnati', sono informazioni. Come viene spiegato, «a un determinato livello di astrazione, *tutti gli oggetti nell'universo sono strutture di dati*»[93]: l'informazione è dunque composta dalle relazioni disponibili in un certo momento, e acquisisce rango di oggetto strutturale.

Tuttavia, per ciò che riguarda la possibilità di dare conto della vulnerabilità, il realismo strutturale informazionale singolarmente considerato sembrerebbe insufficiente. Si sostiene ciò alla luce di tre argomenti: il primo argomento concerne l'assunto del RS per cui una (specifica) parte della realtà sia conoscibile; il secondo riguarda il limite metodologico della parte di realtà conoscibile; il terzo specifica il limite evidenziato dal secondo argomento.

In primo luogo, si consideri il presupposto che regge la posizione del realista strutturale, ovvero l'assunto per cui siano conoscibili i soli aspetti strutturali dell'oggetto (strutturale) che si indaga. La relazione conoscibile (e, ad un certo livello di astrazione, ontologicamente 'impegnata' a causa del RS ontico non-eliminativista che si supporta), nonostante divenga informazione, implica pur sempre una parte di realtà non conoscibile (questo è il presupposto del realismo strutturale 'generale' che si intende salvaguardare, respingendo la posizione del RS ontico eliminativista). Pertanto, pare ragionevole domandarsi perché postulare *ab initio* la riduzione della realtà a strutture. Sotto questo profilo, il RS epistemico singolarmente considerato sembrerebbe più prudente, giacché essendo una posizione che abbiamo definito 'ontologicamente' debole, permette di 'sospendere il giudizio' in relazione a questo aspetto, senza tuttavia negare che esistano strutture e che esse siano conoscibili[94].

In secondo luogo, si consideri il proposito sotteso al RS informazionale, volto alla creazione di modelli. Emerge qui l'importanza del metodo dei livelli di astrazione, costruito «a partire dai metodi formali propri della scienza informatica. Il loro metodo filosofico comporta la selezione di un insieme di "osservabili" a un dato "livello di astrazione". Attribuendo determinati "comportamenti" agli osservabili, si può costruire un modello dell'ente che si sta analizzando e tale modello può essere messo alla prova delle nostre esperienze, osservazioni ed esperimenti»[95]. Il metodo, pertanto, esige la formalizzazione della realtà per la costruzione del modello, e dunque implica la riduzione a qualche forma altra ed ulteriore dell'osservabile (cioè della variabile) che si intende inserire nel sistema – e che, dunque, possa essere sintatticamente rilevante e rilevabile[96].

---

[93] T.W. Bynum, *Introduzione. Filosofia e rivoluzione dell'informazione*, in L. Floridi, *Infosfera. Etica e filosofia nell'età dell'informazione* (a cura di M. Durante), Torino, 2009, 16, corsivo dell'A.

[94] Come spiega più chiaramente Morganti: «[S]ince we have favourable evidence as regards structures as partially preserved through theoretical change, we can be realist about these, without any commitment to what exists beyond them. That is, given our well established conceptual categories, the advocate of ESR can assume the 'traditional' ontology, based on individuals, as unproblematic while also emphasising the role of structures. But s/he by no means needs to prove that our ontology can't be purely structural. Rather, s/he might opt for some kind of 'suspension of judgement' in relation to ontology» (M. Morganti, *op. cit*., 82).

[95] T.W. Bynum, *op. cit*., 16.

[96] A questo riguardo si impongono due osservazioni. In primo luogo, è opportuno notare che sembrerebbe possibile ovviare a questo problema, inserendo il metodo dei livelli di astrazione in una cornice metodologica più ampia, come propone Russo nel suo lavoro (F. Russo, *Techno-Scientific Practices. An Informational Approach*, cit., 78 ss.). L'Autrice suggerisce, infatti, la cooperazione tra pluralismo metodologico (quindi, l'utilizzo di diversi

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Infine, e strettamente connesso al secondo argomento, si rammenti che le relazioni individuabili a seguito della predisposizione del modello, devono essere, in qualche modo, già identificate[97]. Pertanto, esaurire la vulnerabilità all'interno della categoria informazionale produrrebbe l'effetto di perimetrarla alle sole informazioni disponibili (e dunque alle sole relazioni conosciute) in un dato momento. In altre parole, accogliere soltanto il RS informazionale – similmente alla difficoltà in cui incorre la previsione legislativa di un elenco chiuso di situazioni a cui è riconducibile la vulnerabilità – al fine di identificare quando un soggetto si trovi in una situazione siffatta, implicando la predisposizione *ex ante* di un modello, rischierebbe di tralasciare molteplici, e potenzialmente diverse, forme in cui la vulnerabilità potrebbe manifestarsi.

La fruttuosità del naturalismo liberalizzato consiste nel riuscire a coniugare la posizione del RS informazionale con il realismo ordinario: e, pertanto, ad affiancare, accanto a criteri di individuazione su base informazionale, ulteriori accessi epistemici. Il naturalismo liberalizzato è così una teoria che si pone in grado di contemplare la possibilità di conoscere non solo le strutture dei fenomeni indagati (come invece il realismo strutturale postula, in base al primo argomento poc'anzi esposto); essa teoria, inoltre, non impone la restrizione metodologica implicata dalla costruzione di modelli (come accennato con riferimento al secondo argomento); infine, poiché sprovvista del modello, non è necessaria alcuna preventiva identificazione del fenomeno che si indaga (come emerge dal terzo argomento). Il dolore, quindi, sarà compreso come tale, grazie alla percezione, senza alcuna forma di elaborazione informazionale preventiva.

modelli) e pluralismo dell'evidenza: così inteso, l'apparato metodologico sembrerebbe in armonia con il naturalismo liberalizzato di De Caro. In secondo luogo, contro il secondo argomento, si potrebbe sostenere che l'informazione presupposta dal RS informazionale non sia meramente sintattica ma sia semantica. Scrive Bynum: «Secondo Floridi, l'informazione di cui l'universo si compone è *semantica*, piuttosto che meramente sintattica. Inoltre, è *non-fisica*, nel senso che non obbedisce alle leggi della fisica come la seconda legge della termodinamica. Si tratta di informazione *platonica* […] che comprende strutture di dati non soltanto di oggetti familiari, come tavoli e sedie, esseri umani e computer, ma anche di enti platonici come esseri possibili, proprietà intellettuali e storie non scritte di civiltà sparite» (T.W. BYNUM, *op. cit*., 23, corsivi dell'A.). Orbene, pur essendo tale, tuttavia, la formalizzazione non sembra in grado di riuscire a restituire l'intero significato di una certa esperienza. Come spiega più semplicemente Searle: «Quando dico che il programma implementato di per sé non basta a chiarire la coscienza e l'intenzionalità, questa da parte mia è un'affermazione logica. Per definizione, la sintassi del programma non è costitutiva della semantica dei pensieri reali» (J.R. SEARLE, *Ventun anni nella stanza cinese*, in J.R. SEARLE, *Intelligenza artificiale e pensiero umano. Filosofia per un tempo nuovo*, a cura e trad. di A. CONDELLO, Roma, 2023, 89). Pressoché negli stessi termini, Sartor: «Per ora dai sistemi informatici non hanno accesso, se non in misura molto limitata alla dimensione della semantica. […] Questo aspetto (e limite fondamentale) dell'IA riguarda la fondazione (*grounding*) del significato. Nella comunicazione umana il linguaggio non si limita a combinare parole, esso fa riferimento al mondo fisico e sociale. […] La comprensione piena del linguaggio presuppone infatti l'esperienza del mondo» (G. SARTOR, *op. cit*., 22-23, corsivo dell'A.).

[97] Il metodo dei livelli di astrazione è costruito «a partire dai metodi formali propri della scienza informatica. Il loro metodo filosofico comporta la selezione di un insieme di "osservabili" a un dato "livello di astrazione". Attribuendo determinati "comportamenti" agli osservabili, si può costruire un modello dell'ente che si sta analizzando e tale modello può essere messo alla prova delle nostre esperienze, osservazioni ed esperimenti», T.W. BYNUM, *op. cit*., 16.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

5. Conclusioni

A seguito di una breve introduzione, il primo paragrafo ha cercato di chiarire cosa si intenda per "vulnerabilità", sotto un profilo filosofico-giuridico. Sono così emerse due differenti concezioni: la prima predilige l'identificazione del soggetto vulnerabile con l'appartenenza ad un gruppo categoriale, che però – sia prediligendo definizioni formali che sostanziali – si rivela inadatta all'individuazione concreta di una situazione di vulnerabilità; la seconda concezione, quella della vulnerabilità ontologica, designa una condizione universale che accomuna gli esseri umani in quanto tali. La concezione ontologica di vulnerabilità implica il rifiuto di elenchi tassativi come criteri di individuazione di tale situazione e il ripensamento dell'autonomia individuale di ispirazione kantiana. Dal concetto di vulnerabilità proposto da Pastore emerge, inoltre, che la vulnerabilità ontologica sia profondamente radicata non solo nell'esistenza concreta del singolo essere umano ma che debba essere di volta in volta accertata in base alle condizioni specifiche che attorniano una persona in dato momento storico.

Il terzo paragrafo si è interrogato, a partire dalle indicazioni di Pastore, circa lo statuto ontologico del dolore (par. 3.1.), il quale è stato considerato dalla giurisprudenza italiana come possibile criterio di vulnerabilità. Il dolore è stato definito come una "capacità" incarnata, epperò involontaria, che accade similmente alla maniera dello stupore o della meraviglia; il dolore esiste nella singola persona, particolare, ma è comprensibile in maniera universale. Proprio al problema della comprensibilità è stata, in seguito, volta l'attenzione (par. 3.2.): ci si è quindi domandati come sia possibile comprendere il dolore di un soggetto, o meglio, quali siano i canali epistemici, ad un tempo oggettivi e soggettivi, adeguati a tale scopo. La questione è stata indagata tramite le lenti del naturalismo liberalizzato proposto da De Caro, che, valorizzando in maniera congiunta la posizione del realista ordinario e del realista scientifico, permette di comprendere la varietà e molteplicità di indicatori "oggettivi" e "soggettivi" richiesti per rilevare una situazione di vulnerabilità.

Il quarto paragrafo, nel riportare il fulcro della discussione verso il più ampio tema dell'intelligenza artificiale, ha, infine, avuto ad oggetto una particolare forma di realismo scientifico, quello strutturale informazionale: si è così avuto modo di porre in luce l'inadeguatezza dell'adozione esclusiva di una posizione realista di questo tipo per comprendere una situazione di vulnerabilità, riconducibile, in ultima analisi, ai postulati definitori del realismo strutturale.

L'analisi condotta sembrerebbe, dunque, prediligere il paradigma ontologico di vulnerabilità e confermare, così facendo, le perplessità emerse per quanto concerne la formulazione dell'art. 5 co. 1 lettera b) del Regolamento europeo in materia di intelligenza artificiale: per comprendere la vulnerabilità, in virtù della molteplicità di forme che questa può assumere e della varietà di elementi che essa racchiude, non pare adeguato vincolare la qualificazione giuridica a criteri tassativi. Tuttavia, è altresì opportuno segnalare che tale paradigma ontologico di vulnerabilità appare problematico rispetto al principio di certezza del diritto[98], giacché non risulta sempre precipuo come i singoli giudici siano chiamati ad identificare le situazioni di vulnerabilità. In effetti, è stato notato come «l'unico elemento che accomuna tutti i casi nei quali la Corte utilizza la nozione di vulnerabilità è la necessità di effettuare un esame individualizzato della posizione del ricorrente che tenga conto delle sue peculiarità e

[98] R. CHENAL, *op. cit.*, 52 ss.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

che ciò, a seconda del livello di vulnerabilità riscontrato, conduca a una maggiore o un minore livello di protezione»[99].

Al riguardo, la proposta di Roberto Chenal sembrerebbe convincente: l'Autore propone un modello per presunzioni che incarichi il legislatore di individuare «dei criteri rappresentanti le ragioni per giustificare l'estensione o la riduzione della tutela dei diritti»[100], i quali dovrebbero poi guidare il momento della decisione concreta. Trattasi però di criteri che si collocherebbero in un elenco non tassativo, tale da permettere al giudice, previa motivazione, di discostarsi da essi ogniqualvolta lo ritenesse necessario, dovendo però anche motivare, quando sussista, l'opportuna applicabilità del criterio legislativamente previsto al caso di specie. Insomma, «tali criteri costituirebbero norme prescrittive solo nel senso di indicare delle "linee guida" che devono guidare l'operato del giudice. Essi costituirebbero ragioni che fanno considerare rilevante e fanno prevalere *prima facie* un certo interesse rispetto a un altro»[101]. Il principio di certezza del diritto e la conseguente concezione di vulnerabilità categoriale vengono così, in un certo qual modo, salvaguardate, lasciando tuttavia al giudice non soltanto la facoltà ma anche il dovere di considerare, nella valutazione della vulnerabilità, un più ampio e variegato spettro di realtà.

[99] R. CHENAL, *op. cit.*, 51.

[100] R. CHENAL, *op. cit.*, 54; sul sistema basato su presunzioni Bernardini nota, tuttavia, che «da qualche tempo è oggetto di critiche sempre più severe a causa della sua idoneità a favorire la stereotipizzazione ed essenzializzazione di coloro che sono considerati vulnerabili» (M.G. BERNARDINI, *op. cit.*, 80, 80 ss.). Insomma, un sistema per presunzioni correrebbe sempre il rischio di alimentare stereotipi e quindi, da ultimo, di ridurre l'analisi critica specifica relativa alla situazione concreta.

[101] R. CHENAL, *op. cit.*, 54.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Empowering Vulnerability: Decolonizing AI Ethics for Inclusive Epistemological Innovation

Antonio Carnevale*

ABSTRACT: Recent studies reveal a convergence in the ethical guidelines of AI, emphasising the emergence of 'fundamental principles' for responsible AI. However, dissenting voices argue that these principles are insufficient to address the social impacts of AI, revealing a disconnect between ideals and implementation. This article indirectly explores the necessity of AI ethics. It delves into the complexity of cataloguing discriminatory biases generated throughout the lifecycle of AI systems, analysing various types of causal reasoning for discrimination: technical, counterfactual, and finally, constructivist/genealogical. From this exploration, the article derives two additional arguments. Firstly, a call to move beyond bias-based determinism as a singular approach to evaluating discrimination caused by AI systems, thereby recognising the influence of political and social dynamics, including strong appeals for AI decolonisation. Secondly, there is a need to reconsider advocacy actions for vulnerable subjects not merely as a mere claim of denied or marginalised identities but for their epistemic engagement with the world and with others. In this openness, where machine ethics also resides, vulnerability becomes a central epistemological construct to foster inclusive technological innovation, a decisive element in the context of the growing symbiosis between society and AI systems.

KEYWORDS: Discriminatory-sensitive bias; Algorithmic causality; bias-based determinism; AI justice; AI Decolonization; Vulnerability and empowerment.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

* *Researcher in Moral Philosophy, DIRIUM Department, University "Aldo Moro" of Bari; Co-founder of DEXAI – Artificial Ethics. Mail: antonio.carnevale@uniba.it.* *This work is partially supported by the project FAIR – Future AI Research (PE00000013), spoke 6 – Symbiotic AI, under the NRRP MUR program funded by the NextGenerationEU. The article was subject to a double-blind peer review process.*

1. Introduction: What AI ethics?

Studies comparing existing guidelines found that they converge towards the same principles, even more so in recent times[1]. This level of convergence suggests that we are arriving at a set of 'core principles', which is currently the most favoured approach towards principled RAI[2].
Although there is a niche trend to consider AI ethics as the correlate of a constructivist and socio-technical view of AI[3], approaches that posit that the *ex-ante* incorporation of moral principles – such as respect for human autonomy; prevention of harm; fairness; explicability[4] – into machine design is only one domain of articulation of ethics, broadly the philosophical-scientific debate is addressing a dual set of ideas. On the one hand, the necessity of tools to verify that the AI system actually respects the ethical values[5], and on the other hand, the related thought of framing the engineering of ethics in AI systems as an epistemological and practical issue rather than merely a matter of computer science causing[6]. For example, Morley et al. argue about the need to move from 'what' to 'how', that is, to close the gap between principles and practices by constructing a typology that may help practically-minded developers apply ethics at each stage of the machine learning development pipeline, and to signal to researchers where further work is needed.
But are we confident that this often-speculative rush to find methods and tools to verify how much ontologically and engineering-wise ethical principles incorporated into AI will epistemologically produce fairer, more equitable, and sustainable AI systems?

---

[1] A. Jobin et al., *The Global Landscape of AI Ethics Guidelines*, in *Nature Machine Intelligence*, 1, 9, 2019, 389–399. https://doi.org/10.1038/s42256-019-0088-2 (last visited 29/11/2024).

[2] J. Fjeld et al., *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI*, Berkman Klein Center for Internet & Society, Cambridge (MA), 2020. http://nrs.harvard.edu/urn-3:HUL.InstRepos:42160420 (last visited 29/11/2024).

[3] M. Ananny, K. Crawford, *Seeing Without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability*, in *New Media & Society*, 20, 3, 2018, 973–989. https://doi.org/10.1177/1461444816676645 (last visited 29/11/2024); A. Carnevale et al., *A Human-Centred Approach to Symbiotic AI: Questioning the Ethical and Conceptual Foundation,* in *Intelligenza Artificiale*, 18, 1, 2024, 9–20. DOI: 10.3233/IA-240034.

[4] These are the four ethical principles listed by HLEGAI (High-Level Expert Group on Artificial Intelligence), *Ethics Guidelines for Trustworthy AI*, Brussels, 2018-19. The principles rooted in fundamental rights, which must be respected to ensure that AI systems are developed, deployed and used in a trustworthy manner. «They are specified as ethical imperatives, such that AI practitioners should always strive to adhere to them. Without imposing a hierarchy, we list the principles here below in manner that mirrors the order of appearance of the fundamental rights upon which they are based in the EU Charter» (p. 11.). On an emergent consensus in the international milieu on these principles, see also: A. Jobin et al., *op. cit.*; J. Morley et al., *From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices*, in *Science and Engineering Ethics*, 26, 4, 2020, 2141–2168. https://doi.org/10.1007/s11948-019-00165-5 (last visited 29/11/2024).

[5] I. Van De Poel, *Embedding Values in Artificial Intelligence (AI) Systems*, in *Minds and Machines*, 30, 3, 2020, 385–409. https://doi.org/10.1007/s11023-020-09537-4 (last visited 29/11/2024).

[6] J. Morley et al., *op. cit.*; L. Floridi, *The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities*, Oxford, 2023.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Indeed, some scholars radically argue that AI ethical principles are useless, failing to mitigate AI technologies' racial, social, and environmental damages in any meaningful sense. According to Munn[7], AI ethics are a gap between high-minded principles and technological practice. Even when this gap is acknowledged, and principles seek to be 'operationalised'[8], translating from complex social concepts to technical rulesets is non-trivial. In a zero-sum world, the dominant turn to AI principles is not just fruitless but a dangerous distraction, diverting immense financial and human resources away from potentially more effective activity.

The issue, however, is not just the abstractness and the high-minded principles of ethics but also its opposite: an excessive specialisation on certain aspects, as demonstrated, for example, in the debate on how to best incorporate the concept of 'transparency' in AI system design[9]. Similarly, others suggest that concentrating tightly on bias distracts us from more fundamental and urgent questions about power and AI. The moral properties of algorithms are not internal to the models themselves but rather a product of the social and political systems within which they are deployed. This means that AI ethics should be integrated with AI justice theories[10].

This ambivalence between overly abstract and overly specialised ethics can lead to a series of complications that may further complicate the already challenging governance of relationships between society, humans, and machines. One primary complication is the escalating tension and opposition between a 'hard' and 'soft' variant of digital ethics in AI systems.

As argued by Floridi, hard ethics typically involve discussions of values, rights, duties, and responsibilities – or more broadly, what is morally right or wrong, what should or should not be done – when formulating new regulations or critiquing existing ones. For instance, advocating for good legislation or aiming to improve existing legislation can be considered instances of hard ethics. Hard ethics played a role in dismantling apartheid legislation in South Africa. On the other hand, soft ethics operates within the same normative scope as hard ethics but considers what should or should not be done beyond existing regulations, not in opposition to them, or despite their scope, or to change them. In other words, soft ethics represents post-compliance ethics because the «obligation to do something implies the ability to do that something»[11].

While Floridi's analytical distinction ideally involves dialectical impulses, in reality, it is becoming increasingly marked by a stark and rigid opposition between abstraction and hyper-specialization. This anti-dialectical tension leads, on the one hand, to proposals of supererogatory ethics, meaning requests for something impossible, and on the other hand, to overly permissive ethics proposals, which

[7] L. Munn, *The Uselessness of AI Ethics*, in *AI and Ethics*, 3, 3, 2023, 869–877. https://doi.org/10.1007/s43681-022-00209-w (last visited 29/11/2024).

[8] C. Canca, *Operationalizing AI Ethics Principles*, in *Communications of the ACM*, 63, 12, 2020, 18–21. https://doi.org/10.1145/3430368 (last visited 29/11/2024); J. Morley et al., *op. cit.*; A. Dyoub et al., *Learning Domain Ethical Principles from Interactions with Users*, in *Digital Society*, 1, 28, 2022. https://doi.org/10.1007/s44206-022-00026-y (last visited 29/11/2024).

[9] M. Zalnieriute, *"Transparency Washing" in the Digital Age: A Corporate Agenda of Procedural Fetishism*, in *Critical Analysis of Law*, 8, 1, 2021, 139–153. https://doi.org/10.33137/cal.v8i1.36284 (last visited 29/11/2024).

[10] I. Gabriel, *Toward a Theory of Justice for Artificial Intelligence*, in *Daedalus*, 151, 2, 2022, 218–231. https://doi.org/10.1162/daed_a_01911 (last visited 29/11/2024).

[11] L. Floridi, *op. cit.*, precisely see chapter 6.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

serve to confirm or approve compliance with the existing law. This undermines the internal dynamism of ethics between politics and culture, that is, the faculty to move between strong adherence to or contestation of existing rules (politics) and their transformation based on self-regulation and social praxes (culture).

In such a condition of increasing indecisiveness and indeterminacy, ethics are not uncommon to be distorted and used maliciously. This represents a second type of complication. Indeed, the increasing presence of ethical guidelines, committees, and ethicists in both public and private sectors has led computer and data science researchers to question the role of 'ethics' in the tech industry. Critics argue that companies sometimes use ethics to deflect concerns about their behaviour or political crises. Additionally, ethics can be strategically employed to select principles that impose minimal limits on actions while appearing to contribute to the common good[12].

Finally, the degree of confusion and misleading applicability inherent in this state of division leads to a third complication, a growing frustration among stakeholders[13]. As Hagendorff notes[14], almost all the guidelines that have been produced to date suggest that technical solutions exist, but very few provide technical explanations. As a result, developers are becoming frustrated by how little help is offered by highly abstract principles when it comes to the 'day job'[15]. This is reflected in the fact that 79% of tech workers report that they would like practical resources to help them with ethical considerations[16].

Considering everything, do we really need an ethics of AI?

2. Paper organisation

Throughout this article, I will attempt to address this question indirectly. In the first part of my argumentation, I will examine how conceptually complex and challenging it is to catalogue discriminatory-sensitive[17] biases that might negatively cause alterations and harm in the design and development of

[12] L. Floridi, *op. cit.*; B. Green, *The Contestation of Tech Ethics: A Sociotechnical Approach to Technology Ethics*, in *Practice. Journal of Social Computing*, 2, 3, 2021, 209–225. https://doi.org/10.23919/JSC.2021.0018 (last visited 29/11/2024); G. Van Maanen, *AI Ethics, Ethics Washing, and the Need to Politicize Data Ethics*, in *Digital Society*, 1, 2, 2022, 9. https://doi.org/10.1007/s44206-022-00013-3 (last visited 29/11/2024); B. Wagner, *Ethics as an Escape from Regulation. From "Ethics-Washing" to Ethics-Shopping?*, in E. Bayamlioglu et al. (eds.), *Being Profiled: Cogitas Ergo Sum. 10 Years of Profiling the European Citizen*, Amsterdam, 2018, 84–89. https://doi.org/10.1515/9789048550180-016 (last visited 29/11/2024).

[13] J. Morley et al., *op. cit.*

[14] T. Hagendorff, *The Ethics of AI Ethics – An Evaluation of Guidelines*, in *Minds and Machines*, 30, 1, 2020, 99–120. https://doi.org/10.1007/s11023-020-09517-8 (last visited 29/11/2024).

[15] D. Peters, *Beyond Principles: A Process for Responsible Tech*, in *The Ethics of Digital Experience*, 2 May 2019. https://medium.com/ethics-of-digital-experience/beyond-principles-a-process-for-responsible-tech-aefc921f7317 (last visited 29/11/2024).

[16] J. Morley et al., *op. cit.*

[17] By 'discriminatory-sensitive', I refer to a range of specific quality requirements that, due to space limitations in this contribution, I would equate with (a) the seven ethical requirements defined by HLEGAI, *op. cit.*; and (b) the discrimination categories described in the volume by the European Union Agency for Fundamental Rights, European Court of Human Rights, & Council of Europe, *Handbook on European non-discrimination law*, Strasbourg, 2018. https://data.europa.eu/doi/10.2811/58933 (last visited 29/11/2024).

Downloaded from www.biodiritto.org.
ISSN 2284-4503

AI systems. Against this backdrop, I conduct an examination of various studies that have elucidated discriminatory causality according to three types of reasoning: technical, counterfactual, and constructivist/genealogical.

From this initial exploration of challenges, I derive two additional lines of argumentation to complement my contribution further. In the first of these lines, I argue that we must move beyond an exclusively bias-based determinism to evaluating AI systems' discriminatory-sensitive aspects. Behind each ethical dilemma and every algorithmic process leading causally to biased outcomes lies a complex web of political and social dynamics. These dynamics are influenced by pressing calls for justice, such as those advocating for AI decolonization, reshaping the understanding of causality to be fluid and relational rather than deterministic.

Secondly, within this intricate political landscape, the moral actions of humans and the operationalisation of trustworthy machines extend beyond the mere assertion or protection of marginalized and vulnerable identities or the adherence to binary oppositions. Rather, they represent an epistemic engagement with the world and with others, constituting a cognitive assemblage. AI ethics might play a pivotal role in illuminating and enriching this nuanced discourse, thereby shaping the landscape of digital innovation that lies ahead. Against this backdrop, my aim has been to discern a revitalized notion of vulnerability empowerment. This conception emerges as a central epistemological tenet driving inclusive innovation and ethical governance in anticipation and mitigation of the forthcoming symbiosis between society and AI systems.

3. Algorithmic Bias and Discrimination: A Conceptual Dilemma of Causality

One of the pivotal aspects for ensuring that AI ethics can genuinely transition from the theoretical-conceptual phase ('what') to the pragmatic-orientation phase ('how') is to find convergent epistemic approaches and evaluative measures concerning the thorny issue of *bias* and its *discriminatory causality*. It is now undeniable that AI, especially in variants involving the support of machine learning techniques or extensions of generative AI, inherently revolves around the theme of bias. In certain algorithmic programming paradigms, the practice of bias is not understood in a negative sense – as prejudice – but is used to indicate a 'deviation from a standard', which can, therefore, occur at any stage of the design, development, and implementation process[18].

If it is indeed true that a *design by-bias* cannot be entirely disregarded in the lifecycle of AI systems, how then can one distinguish biases applicable to design from those that may instead engender discrimination? Hence, identifying the causal reasons behind the discriminations produced by AI biases – even considering causality in a thick sense as something constructivist and genealogical[19] – is by no means trivial. The major problem lies in the polyvalent and multi-layered nature of bias manifestations, as they are identifiable (a) both in the replication and reinforcement of cognitive biases already present in historical world data and in those with a higher additional layer of direct responsibility

[18] L. Floridi, *op. cit.*

[19] See: I. Kohler-Hausmann, *Eddie Murphy and The Dangers of Counterfactual Causal Thinking about Detecting Racial Discrimination*, in *Northwestern University Law Review*, 113, 2018, 1163–1228; M. Ziosi et al., *A Genealogical Approach to Algorithmic Bias*, in *Social Science Research Network (SSRN) Electronic Journal*, 2024. https://doi.org/10.2139/ssrn.4734082 (last visited 29/11/2024).

Downloaded from www.biodiritto.org.
ISSN 2284-4503

stemming from interventions that may (b) unveil new associations, highlighting, with tangible results, connections and interdependencies among data never seen before, or (c) synthetically anticipate the formation of new biases, creating hypotheses of future realities that are currently unforeseeable.
In the following, I will consider various research and studies that have attempted to systematise discriminatory causal complexity through different theoretical and methodological proposals: discriminatory causality as (i) *technical reasoning, (ii) counterfactual reasoning, and* (iii) *constructivist and genealogical reasoning*.

Discriminatory causality as technical reasoning. At the groundwork of my inquiry, I posit discriminatory causality as the outcome of technical reasoning concerning the type of modelling and training of the AI system, particularly when employing machine or deep learning techniques. As Pasquinelli argues, within this type of causality, we must identify at least three levels: world, data, and algorithm biases[20]:

- *World bias*: in society, biases like race, gender, and class inequalities are already present, and datasets often reinforce these biases, perpetuating stereotypes. In this context, Crawford distinguishes between two types of harm caused by bias in algorithms: resource allocation harm, such as denying mortgages to minority groups, and social representation harm, like denigration or unfair classification based on race, gender, or class.
- *Data bias* occurs during training data collection, formatting, and labelling, often reflecting outdated and biased taxonomies that distort cultural and scientific realities. This bias becomes ingrained in machine learning algorithms, amplifying existing biases and distorting information further.
- *Algorithmic bias*, resulting from computational errors and information compression, exacerbates inequalities by distorting and amplifying biases present in both the world and the data. This distortion is akin to the anamorphic perspective used in art, where proportions are distorted to maintain shape. This illustrates how machine learning can magnify biases in unexpected ways.

While this approach is helpful in abstracting and defining analytical processes, it tends to overlook the social complexity of the real world[21]. This leads to a dominant mindset in algorithm development, characterised by 'algorithmic formalism', which is adherence to prescribed rules and forms[21]. One potential approach to mitigate this issue involves intentionally excluding certain specific data variables from the training of the algorithmic decision-making process. Indeed, the treatment of statistically relevant sensitive variables or 'protected variables', such as gender or race, is typically restricted or prohibited by anti-discrimination laws and data protection regulations, aiming to mitigate the risks of unfair discrimination. However, this type of intervention raises ethical questions at a higher level than technical reasoning, as we will explore in the subsequent types of discriminatory causal reasoning.

[20] M. PASQUINELLI, *How a Machine Learns and Fails*, in *Spheres: Journal for Digital Cultures*, 5, 2019, 1–17. https://doi.org/10.25969/MEDIAREP/13490 (last visited 29/11/2024).

[21] B. GREEN, S. VILJOEN, *Algorithmic Realism: Expanding the* Boundaries *of Algorithmic Thought*, in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 2020, 19–31. https://doi.org/10.1145/3351095.3372840 (last visited 29/11/2024).

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Discriminatory causality as counterfactual reasoning. As some scholars have emphasised, this kind of reasoning has been found worthy of explanatory conjecture in Judea Pearl's theory of causality[22]. The author articulates causal complexity into three levels, titled 1) association, 2) intervention, and 3) counterfactual.

- The first level is called *association* because it invokes purely statistical relationships defined by the naked data. For instance, observing a customer who buys toothpaste makes it more likely that he/she buys floss; such association can be inferred directly from the observed data using conditional expectation. Questions at this layer are placed at the bottom level of the hierarchy because they require no causal information.
- The second level, *intervention*, ranks higher than association because it involves not just seeing what is but changing what we see. A typical question at this level would be: What happens if we double the price? Such questions cannot be answered from sales data alone because they involve changing customers' behaviour in reaction to the new pricing. Customer choices under the new price structure may differ substantially from those prevailing in the past.
- Finally, the top level is called *counterfactuals*, which is a typical question in "What if I were to act differently?" Thus, it necessitates retrospective reasoning.

Researchers have often applied this reasoning in AI ethics to understand whether a hypothetical intervention to alter a subject's protected characteristic would have changed the outcome[23]. Most notably, Galhotra et al. propose 'probabilistic contrastive counterfactuals', which help quantify a feature's direct and indirect effects on outcomes and provide actionable recourse to individuals negatively affected by such an outcome[24].

This type of reasoning benefits from providing an appreciable logical-argumentative framework within the field of explainable artificial intelligence (XAI), which aims to diminish the opacity of AI-based decision-making systems, enabling human scrutiny and trust. However, as argued by Kohler-Hausmann and Ziosi et al.[25], this model inclines to be flawed. In this way – Kohler-Hausmann claims – «discrimination is detected by measuring the 'treatment effect of race', where the treatment is conceptualized as manipulating the raced status of otherwise identical units (e.g., a person, a neighborhood, a school). […] The counterfactual causal model of discrimination is not wrong because we can't work around the practical limits of manipulation […]. It is wrong because to fit the rigor of the counterfactual model of a clearly defined treatment on otherwise identical units, we must reduce race to only the signs of the category, meaning we must think race is skin color, or phenotype, or other ways we identify group status. And that is a concept mistake if one subscribes to a constructivist, as opposed to a biological or genetic, conception of race. The counterfactual causal model of discrimination is based on a flawed theory of what the category of race references, how it produces effects in

[22] J. Pearl, *Causality: Models, Reasoning, and Inference*, Cambridge (MA), 2000. See also M. Ziosi et al., *op. cit.*

[23] Examples are provided by A.-H. Karimi et al., *Algorithmic Recourse: From Counterfactual Explanations to Interventions*, 2020, arXiv:2002.06278. https://doi.org/10.48550/ARXIV.2002.06278 (last visited 29/11/2024).

[24] See S. Galhotra et al., *Explaining Black-Box Algorithms Using Probabilistic Contrastive Counterfactuals*, 2021, arXiv:2103.11972. https://doi.org/10.48550/ARXIV.2103.11972 (last visited 29/11/2024).

[25] I. Kohler-Hausmann, *op. cit.*; M. Ziosi et al., *op. cit.*

Downloaded from www.biodiritto.org.
ISSN 2284-4503

the world, and what is meant when we say it is wrong to make decisions of import because of race»[26].

Discriminatory causality as constructivist and genealogical reasoning. To avoid protected features like gender, race, disability, etc., being represented as discrete units, existing in isolation rather than in relation, computer scientists and AI ethicists should consider the frontline discussion in social sciences, in which, indeed, many studies are converging on the assumption that no one theory of causation satisfies all scientific domains or specific studies. Accordingly, one should construct an appropriate 'causal mosaic' for each research study to determine what is causally relevant and articulate one's assumptions and approaches for warranting one's causal claim(s)[27]. According to scholars like Ziosi et al., this explains why AI ethics needs to shift the focus to constructive and genealogical conditions rather than the consequences of discriminatory outcomes to emphasise the importance of understanding and preventing algorithmic discrimination. According to Kohler-Hausmann, «Discrimination is a thick ethical concept that at once describes and evaluates the actions to which it is applied, and therefore, we cannot detect actions as discriminatory by identifying a relation of counterfactual causality; we can do so only by reasoning about the action's distinctive wrongfulness by referencing what constitutes the very categories that are the objects of concern»[28].

4. Beyond a Bias-Based Determinism: AI Justice and Decolonisation

Technical and Counterfactual approaches are better suited for observing whether variables like gender, race, disability, etc., are independent factors rather than elucidating the specific role they play in comparison to other factors. However, *observing* a phenomenon does not necessarily equate to understanding it. Increasingly, AI ethics concerns itself with peering into algorithms with the aim of elucidating the opaque mechanisms surrounding inference operations and statistical distribution – to prevent well-known effects such as *over-* or *underfitting* – thereby enhancing the transparency of the AI system. Yet, transparency is a political construct and should not solely be sought *inside* the machinery, but rather, as Ananny and Crawford argue, *across them*: «The implicit assumption behind calls for transparency is that *seeing* a phenomenon creates opportunities and obligations to make it accountable and thus to change it. We suggest here that rather than privileging a type of accountability that needs to look inside systems, that we instead hold systems accountable by looking across them—seeing them as sociotechnical systems that do not contain complexity but enact complexity by connecting to and intertwining with assemblages of humans and non-humans»[29].

In other words, we require theoretical approaches and methodologies qualified for elucidating algorithmic causality beyond the intrinsic rationality inherent in their construction and programming. If we perceive the ethical quandary of a fair, equitable, and reliable AI to lie in rendering its 'statistical

[26] I. Kohler-Hausmann, *op. cit.*, here p. 1163.

[27] R.B. Johnson et al., *Causation in Mixed Methods Research: The Meeting of Philosophy, Science, and Practice*, in *Journal of Mixed Methods Research*, 13, 2, 2019, 143–162. https://doi.org/10.1177/1558689817719610 (last visited 29/11/2024).

[28] I. Kohler-Hausmann, *op. cit.*, here p. 1163.

[29] M. Ananny, K. Crawford, *op. cit.*, here p. 974.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

unconscious'[30], so to speak, as transparent as possible, we risk confusion. Not only are we looking in the wrong place (within the algorithm rather than through it), but we also risk being ensnared by an «enchanted determinism»[31]. For a multitude of reasons, including the nonlinear trajectory from inputs to outputs, we have yet to develop a theory that can explain why deep learning techniques excel at pattern detection and prediction, leading us humans to assert claims about 'superhuman' accuracy and insight while remaining unable to fully explicate the origins of these outcomes.

In the essay *Empiricism and the Philosophy of Mind* (1956), Wilfrid Sellars, in his critique of logical empiricism, demonstrated that knowledge having foundations independent of the linguistic-conceptual dimension is a myth, namely the 'myth of the Given'. The logical empiricist reasoning goes roughly as follows: for knowledge to be meaningful and not merely a play of the intellect, it requires a clear grounding in the empirical realm. Knowledge must be founded on empirical grounds, which must be divorced from any intellectual operation or linguistic-conceptual act to fulfil their role as foundations. On the contrary, Sellars argued that empirical facts only play and can play the foundational role for knowledge because, from their inception, they exist within a specific linguistic and conceptual configuration.

Let us extend Sellars' thought to AI. The determinism we believe inherent in AI's ability to provide a plausible representation of a 'given' reality or even to predict its imminent historical occurrence is not ontologically significant in the strict sense, as it entirely lacks the foundational role played by empirical facts, instead offering regularities and evidence entirely stemming from a statistical configuration of knowledge. Thus, if its foundation lacks empirical facts from the bottom, its knowledge lacks a language that can be spoken, put into practice, externalized, understood, and misunderstood from above. It is a novel mythology, a determinism *doubly insignificant* from an ontological perspective.

Conversely, algorithmic determinism becomes significant when viewed through it, within the *political conditions of its sociotechnical possibilities*. This is what the most advanced studies in critiquing AI ethics, such as AI justice and AI decolonization[32], tell us.

On the one hand, AI justice help to reframe much of the discussion around AI ethics by drawing attention to the fact that the moral properties of algorithms are not internal to the models themselves but rather a product of the social systems within which they are deployed. A scholar like Zalnieriute argues, for example, that the current focus on AI procedural issues like transparency is blinkered, acting as an «obfuscation and redirection from more substantive and fundamental questions about the concentration of power, substantial policies, and actions of technology behemoths»[33]. According

[30] In this context, there are studies that have questioned how machines can have negative conscious experiences, as seen in: L. Dung, *How to Deal with Risks of AI Suffering*, in *Inquiry*, 2023, 1–29. https://doi.org/10.1080/0020174X.2023.2238287 (last visited 29/11/2024).

[31] A. Campolo, K. Crawford, *Enchanted Determinism: Power Without Responsibility in Artificial Intelligence*, in *Engaging Science, Technology, and Society*, 6, 2020, 1–19. https://doi.org/10.17351/ests2020.277 (last visited 29/11/2024). See also K. Crawford, *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*, New Haven, 2021.

[32] See: L. Munn, *op. cit.*; I. Gabriel, *op. cit.*; S. Mohamed et al., *Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence*, in *Philosophy & Technology*, 33, 4, 2020, 659–684. https://doi.org/10.1007/s13347-020-00405-8 (last visited 29/11/2024).

[33] M. Zalnieriute, *op. cit.*, p. 139.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

to Munn[34], if ethical principles are situated within company cultures and broader systems of power, then it makes sense to expand the scope of ethical engagement. Or, put differently, if machine learning reflects, reproduces, and amplifies structural inequalities, then any ethical program must operate intersectionally, considering a wide array of social and political dynamics and questioning the «seductive diversion of 'solving' bias in artificial intelligence»[35].

On the other hand, decolonial theorists recognise parallels between territorial and structural coloniality in the digital era[36]. Digital spaces, akin to physical territories, are susceptible to exploitation and extraction[37], fostering digital-territorial colonialism. This extends to digital-structural colonialism, where colonial power dynamics persist through socio-cultural imaginaries and technological development rooted in unquestioned historical values. Data colonialism and capitalism theories acknowledge data as a resource exploited for economic gain, reflecting the coloniality of technological power. Algorithmic coloniality emerges as algorithms shape resource allocation, societal behaviour, and discriminatory systems, influencing labour markets and geopolitical dynamics[38]. Against this backdrop, Mohamed et al. propose introducing a decolonial foresight taxonomy[39]. It will identify sites of coloniality, such as algorithmic decision systems and ghost work, revealing structural inequalities with historical colonial roots. By recognising these sites, discussions on power and inequality in AI must acknowledge colonial continuities, ensuring a comprehensive understanding of the societal impacts of algorithmic systems.

5. Empowering Vulnerability

«We build material and electronic walls, fences, and dikes to keep out the viruses and dark waters of death. As technological beings, these are the sort of things we humans do. In fact, it is hard to imagine what our material culture would look like without the struggle against vulnerability: technology is our vulnerability guardian, and it is in the guardian's house that we live as technological, risk-phobic beings. We are vulnerable by nature, but we are also vulnerability-averse by nature. We are already rebels. We are the children of Prometheus»[40].

[34] L. Munn, *op. cit*.

[35] J. Powles, *The Seductive Diversion of 'Solving' Bias in Artificial Intelligence*, in *OneZero (blog)*, December 7, 2018. https://onezero.medium.com/the-seductive-diversion-of-solving-bias-inartificial-intelligence-890df5e5ef53 (last visited 29/11/2024).

[36] J. Thatcher et al., *Data Colonialism Through Accumulation by Dispossession: New Metaphors for Daily Data*, in *Environment and Planning D: Society and Space*, 34, 6, 2016, 990–1006. https://doi.org/10.1177/0263775816633195 (last visited 29/11/2024).

[37] N. Couldry, U.A. Mejias, *Data Colonialism: Rethinking Big Data's Relation to the Contemporary Subject*, in *Television & New Media*, 20, 4, 2019, 336–349. https://doi.org/10.1177/1527476418796632 (last visited 29/11/2024).

[38] P. Ricaurte, *Data Epistemologies, the Coloniality of Power, and Resistance*, in *Television & New Media*, 20, 4, 2019, 350–365. https://doi.org/10.1177/1527476419831640 (last visited 29/11/2024).

[39] S. Mohamed et al., *op. cit.*

[40] M. Coeckelbergh, *Human Being@Risk: Enhancement, Technology, and the Evaluation of Vulnerability Transformations*, Dordrecht-New York, 2013, here p. 4.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

I shall begin by specifying that the thesis underlying this final part is not the mere, albeit non-trivial, observation that the empowerment of vulnerability signifies a marked interpretative and ideological shift from a perception of weakness, of fragility to be ashamed of, to one of human dignity to be protected (also with technological aids) and whose assertion makes us stronger—more comprehensively human. A series of studies, including disability, capability approach and feminist studies, have now placed the socio-political issue of vulnerability on this plane[41]. What I would like to highlight, however, is an *epistemic* nuance contained within the dynamics of vulnerability.

To be vulnerable is always to be 'vulnerable to something', something external. This means that being vulnerable describes a situation not inherently one of inferiority but of susceptibility to external inducements. However, let us investigate more closely what this 'being outside' of those things that make us vulnerable entails.

Let us begin by stating that the something to which we are vulnerable is not simply a brute natural fact external to us, which by its presence influences us in some way. That something is an event, it is something that not only lies outside but *comes from outside*. Consider seismic or environmental vulnerability, defined as the propensity to suffer damage because of inducements from an event of a certain intensity. Its mere presence is, therefore, not sufficient. What renders us vulnerable must also possess a certain intensity. Otherwise, the inducements would not trigger, and vulnerability would never transition – to borrow Aristotelian terms – from its nominal potentiality (vulnerability as a noun) to its practical actuality (being effectively vulnerable to that something, i.e., an attribute). Pushing further, one might venture an additional speculation. Precisely because being vulnerable is always 'being vulnerable to something', it could be argued that it is the intensity of external events – hence not the brute facts but the quality of events – that determines the type of inducement, which in turn determines the essence of vulnerability. This leads me to argue that vulnerability is not a causal condition but an epistemic openness to the world[42].

Nevertheless, 'coming from outside' is not the only possible direction of this openness. If we consider some emotional states of individuals, in addition to coming from outside, we must add a second and perhaps more important variant, which is *being put outside*. Indeed, those who are vulnerable are exposed, uncovered, sensitive, and easily hurt. A person with a vulnerable character is easily mortified, offended, or depressed. In this second variant, vulnerability is not an epistemic openness to the world, but to the relationships between oneself and others[43].

[41] Just to mention a few: S.G. HARDING, *The Science Question in Feminism*, Ithaca, New York, 1986; D. HARAWAY, *A Cyborg Manifesto: Science, Technology, and Socialist-Feminism in the Late Twentieth Century*, in *The Transgender Studies Reader*, London, 2013, 103–118; ID., *Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective*, in *Space, Gender, Knowledge: Feminist Readings*, London, 2016, 53–72; A. CARNEVALE, *Robots, Disability, and Good Human Life*, in *Disability Studies Quarterly*, 35, 1, 2015. https://ojs.library.osu.edu/index.php/dsq/article/view/4604 (last visited 29/11/2024); M.J. HAENSSGEN, P. ARIANA, *The Place of Technology in the Capability Approach*, in *Oxford Development Studies*, 46, 1, 2018, 98–112. https://doi.org/10.1080/13600818.2017.1325456 (last visited 29/11/2024); D. CIRILLO ET AL., *Sex and Gender Differences and Biases in Artificial Intelligence for Biomedicine and Healthcare*, in *Npj Digital Medicine*, 3, 81, 2020. https://doi.org/10.1038/s41746-020-0288-5 (last visited 29/11/2024).

[42] A. CARNEVALE, *Tecno-vulnerabili. Per un'etica della sostenibilità tecnologica*, Salerno-Naples, 2017.

[43] L. AMOORE, *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others*, Durham, 2020.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

It is precisely in this going back and forth from and towards the world and from and towards others that I see an epistemic aspect of vulnerability imbued. To echo the words of Coeckelbergh, cited earlier and which may now acquire a broader meaning, «We are vulnerable by nature, but we are also vulnerability-averse by nature. We are already rebels». Being vulnerable to something is neither an ontological condition (something unchosen that we find ourselves saddled with), nor normative (a habitus chosen by law) or rather, more accurately, vulnerability can be both things, but this will depend on the social and political choices we make and which will shape the levels of abstraction with which we define causal nexuses, including the discriminatory causality of algorithms. We are *not* rebels. We are *already* rebels. This implies that AI justice and AI decolonisation are not the politically strongest solutions to AI ethical weakness, but ways of posing the right social and political choices in order to produce different levels of abstraction, thus capable of governing the ethical issue of AI system opacity in a non-hegemonic and mono-ideological manner. And I mean *sociotechnical* opacity, which concerns not only machines but, as Hayles states, the 'cognitive assemblages' of our technosymbioses[44] or our techno-vulnerability[45]. «For example, deciding what areas of autonomy a self-driving car will have is simultaneously a decision about what areas of autonomy a human driver will (and will not) have. Such a system does not exist in isolation. It is also necessary to take into consideration the sources and kinds of information available for the entities in a cognitive assemblage and their capabilities of processing and interpreting it. Humans can see road signs in the visible spectrum, for example, but a self-driving car might respond as well to markers in the infrared region. It is crucially important to realise that the cognitive entities in a cognitive assemblage process information, perform interpretations, and create meanings in species-specific ways»[46].

6. Conclusions

So, revisiting the question posed in the introduction, which kind of ethics of AI do we need?
If we conceive of AI ethics as ensuring that a system, no longer produces biased outcomes – such as when a facial recognition program fails to identify the face of a person of colour – then we would argue against it. We do not require such ethics, as it fails to address the crux of the matter: since the system has the capacity for self-correction, what is needed are engineers who are more attentive and sensitive to revising datasets to include vulnerable individuals and social groups. Similarly, if we regard AI ethics as a rule-driven guideline toward hyper-compliance and meeting demands for greater transparency, explainability, etc., for instance, toward a corporation to disclose its algorithms, once more, we will say no, as well. Such ethics remains abstract, a *petitio principii*, as algorithms are in constant flux as the system learns, rendering transparency at one point means obscurity at another. Such ethics serve no purpose; it is far more advantageous to be supported by jurists and lawyers who at least have the framework of existing laws as a concrete perspective for regulation.

[44] N.K. HAYLES, *Technosymbiosis: Figuring (Out) Our Relations to AI*, in J. BROWNE, ET AL. (eds.), *Feminist AI*, Oxford, 2023 (1st ed.), 1–18. https://doi.org/10.1093/oso/9780192889898.003.0001 (last visited 29/11/2024).
[45] A. CARNEVALE, *Tecno-vulnerabili. Per un'etica della sostenibilità tecnologica*, cit.
[46] N.K. HAYLES, *op. cit.*, p. 14.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Conversely, if we conceive that AI ethics must have some minimal reference to *ethos*, the Greek term from which it originates, and which denoted 'character', signifying the guiding beliefs or ideals that characterize a community, nation, or ideology, then AI ethics must be relevant and attentive to at least two other aspects.

Firstly, behind every ethical challenge and every algorithmic process causing discriminatory biases, there exists a structure of political and social relations upon which strong demands for justice, such as those of AI decolonization, exert influence, rendering the framework of causality fluid, relational, and not the outcome of deterministic inference.

Secondly, any moral and advocacy actions of humans as well as any operationalisation of trustworthy machines happen within this socio-political openness and it is not merely a matter of claiming denied or marginalized identities, of a binary oppositional logic of black or white, but of epistemic positioning in the world and in relation to others, a cognitive assemblage that AI ethics can assist in bringing to light and colouring the digital innovation that is upon us.

In this openness, where machine ethics also resides, vulnerability becomes a central epistemological construct to foster inclusive technological innovation, a decisive element in the context of the growing symbiosis between society and AI systems.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Vulnerabilità. Note sul ruolo del concetto nell'AI Act

Silvia Dadà*

VULNERABILITY. SOME REMARKS ON THE ROLE OF VULNERABILITY IN THE AI ACT

ABSTRACT: The paper aims to analyze the role of the concept of vulnerability in European Artificial Intelligence regulation (AI Act). The article will be developed in two parts. In the first we will conduct a conceptual analysis of vulnerability in its multiple dimensions, distinguishing between a universal and a particular sense, and between a categorizing and a situational approach. To do so, we will discuss the main lines of philosophical, legal and bioethical debate on the topic. In the second we will investigate the role of vulnerability in the AI Act, commenting on the most significant recurrences and major variations. We will adopt a synoptic look at the evolution of the document, from its first version to the final one, passing through the amendments proposed by the European Parliament. We will argue that in this document vulnerability is understood primarily in a particular sense, while the universal sense remains almost absent or not directly expressed. This absence, as we will show, seems only partially compensated for by the risk-based approach.

KEYWORDS: AI Act; Vulnerability; Risk-Based Approach; EU.

ABSTRACT: Il contributo si propone di analizzare il ruolo del concetto di vulnerabilità nella regolamentazione europea sull'Intelligenza Artificiale (AI Act). L'argomentazione sarà sviluppata in due parti. Nella prima ci dedicheremo all'analisi concettuale della vulnerabilità nelle sue molteplici dimensioni, distinguendo tra un senso universale e uno particolare, e tra un approccio categorizzante e uno situazionale. Per fare ciò, discuteremo le linee principali del dibattito filosofico, giuridico e bioetico sul tema. Nella seconda passeremo ad analizzare l'impiego del termine nell'AI Act, commentando le ricorrenze più significative e le principali variazioni. Adotteremo uno sguardo sinottico all'evoluzione del documento, dalla sua prima versione a quella definitiva, passando per gli emendamenti proposti dal Parlamento Europeo. Sosterremo che nel documento la vulnerabilità è intesa principalmente in senso particolare, mentre il senso universale rimane pressoché assente o implicito. Tale assenza, come mostreremo, sembra poter essere solo parzialmente compensata dall'approccio basato sul rischio.

PAROLE CHIAVE: AI Act; Vulnerabilità; Approccio basato sul rischio; UE.

* *Ricercatrice di Filosofia Morale, Università di Pisa. Mail: silvia.dada@unipi.it.* *Contributo sottoposto a doppio referaggio anonimo.*

Downloaded from www.biodiritto.org.
ISSN 2284-4503

1. Introduzione

Il concetto di «vulnerabilità» ha assunto oggi una fondamentale centralità nel dibattito pubblico e accademico[1]. La riflessione filosofica, bioetica, politica e giuridica si sono sempre più interessate a questo termine, aprendo ampi dibattiti sul suo statuto e sul suo significato.

Malgrado tale diffusione nei più svariati ambiti e discipline il termine «vulnerabilità» sembra mantenere ancora un certo grado di indeterminatezza, oscillando tra molteplici interpretazioni e accezioni. Esso necessita, quindi, di un'analisi sistematica che permetta di rendere più chiaro il suo specifico senso nei vari contesti di utilizzo. Primo obiettivo di questo contributo sarà proprio quello di riportare gli aspetti più significativi di tale discorso, per restituire una chiara analisi di questa nozione e delle sue interpretazioni. Nella seconda parte, invece, ci dedicheremo in modo più specifico al ruolo che questo termine gioca nell'attuale contesto tecnologico, dominato dall'IA. Infatti, le crescenti opportunità offerte da questi sistemi vanno di pari passo con un aumento esponenziale dei rischi a cui siamo esposti, rendendoci più vulnerabili. Ciò ha reso necessario un intervento normativo, che ha portato, nel caso dell'Unione Europea, alla redazione e alla approvazione di una specifica regolamentazione in materia di Intelligenza Artificiale (*Artificial Intelligence Act*, d'ora in poi semplicemente *AI Act*[2]). Prenderemo in esame questo documento, per vedere quale ruolo (o *quali ruoli*) giochi il concetto di vulnerabilità al suo interno. Ripercorreremo la sua genesi considerando i cambiamenti e le modifiche dalla prima versione della proposta (2021) passando per gli emendamenti del Parlamento Europeo (2023) al fine di comprendere meglio le scelte adottate nella versione approvata e definitiva.

2. Analisi del concetto tra filosofia, bioetica e diritto

Sebbene dal punto di vista etimologico nella lingua latina non si trovi un corrispettivo esatto al sostantivo «vulnerabilità», l'origine del termine è da collegarsi all'aggettivo *vulnerabilis* e al verbo *vulnerare*. In entrambi i casi si fa riferimento all'esposizione al *vulnus,* ossia alla ferita, intendendo con questo termine il senso corporeo e concreto del danno.

Il quadro semantico della vulnerabilità è perciò principalmente occupato dalle idee di fragilità e finitudine che rimandano alla sfera della percezione e della sofferenza sino a quella della mortalità. Sebbene quindi il riferimento al corpo sia un elemento centrale della nozione, il campo semantico si è ampliato nel suo uso corrente, sino a comprendere anche altri aspetti, tra cui il danno psicologico e morale.

In senso più generale, possiamo dire che l'idea di vulnerabilità è un concetto relazionale, in quanto si manifesta nell'incontro con un'alterità, e si connette alla nozione di *dipendenza*.

[1] Per una completa ricostruzione del dibattito sull'idea di vulnerabilità si rimanda a H. TEN HAVE, *Vulnerability. Challenging Bioethics,* London, 2016.

[2] La versione definitiva del documento risale al 12 luglio 2024.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Si tratta di una caratteristica propria dell'essere umano che non può essere negata se non al costo di indebite semplificazioni o di inverosimili narrazioni: sono rari i momenti della nostra vita in cui possiamo definirci veramente indipendenti ed esenti da fragilità: contro il mito del soggetto autonomo[3] emerge l'immagine più realistica di un soggetto concreto e corporeo, interdipendente ed esposto. Non tutte le forme di dipendenza sono però inevitabili e naturali: alcune sono dovute a degli squilibri di potere che possono essere mitigati o sovvertiti. La dipendenza può infatti finire per declinarsi in termini di esposizione alla violenza, in cui l'essere umano prevarica e sottomette il suo simile, limitandone le proprie possibilità di autodeterminarsi e costringendolo a «vivere alla mercè»[4].
Tutti questi campi semantici toccano e arricchiscono il quadro variegato della vulnerabilità, rivelandoci subito la difficoltà di renderlo in modo sistematico e unificato, col rischio che rimanga un termine troppo ampio e vago[5]. Come dice Samia Hurst, nel tentativo di definire la vulnerabilità ci troviamo come in quella leggenda in cui diversi uomini ciechi toccano un elefante, dando ognuno una prospettiva parziale differente dell'animale, in base alla parte di cui fanno esperienza[6]. Proprio a causa di questo scenario complesso, sono state elaborate numerose tassonomie e classificazioni, al fine di offrire un quadro il più possibile completo e chiaro[7].

2.1 Vulnerabilità particolare e vulnerabilità universale

La principale distinzione che dobbiamo considerare è quella tra un senso *universale* e uno *particolare* di vulnerabilità[8].

[3] M.A. Fineman, *The Autonomy Myth. A Theory of Dependency*, New York, 2004.

[4] E. Ferrarese, *Vivere alla mercè Figure della vulnerabilità nelle teorie politiche contemporanee*, in *La società degli individui*, 38, 13, 2010, 21-33. Sul ruolo politico della vulnerabilità rispetto alla soggettivazione e ai rapporti di potere si veda anche J. Butler, *Vite precarie: Contro l'uso della violenza in risposta al lutto collettivo*, Milano, 2004.

[5] D. Schroeder, E. Gefenas, *Vulnerability: Too Vague and Too Broad?,* in *Cambridge Quarterly of Healthcare Ethics*, 18, 2009, 113-121.

[6] S. Hurst, *Vulnerability in research and health care: describing the elephant in the room?,* in *Bioethics*, 22, 4, 2008, 191-202.

[7] W. Rogers, C., Mackenzie, S., Dodds, *Why bioethics needs a concept of vulnerability*, in *International Journal of Feminist Approaches to Bioethics*, 5, 2, 2012, 11-38; K. Kipnis, *Vulnerability in research subjects: A bioethical taxonomy*, in Aa. Vv., *Ethical and Policy Issues in Research Involving Human Participants*, National Bioethics Advisory Commission, Bethesda 2001.

[8] M.G. Bernardini, *Il soggetto vulnerabile. Status e prospettive di una categoria (giuridicamente) controversa*, in *Rivista di Filosofia del diritto*, 2, 2017, 365-384; S. Pastore, *Semantica della vulnerabilità, soggetto, cultura giuridica*, Torino, 2021.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Si parla, indistintamente, sia di vulnerabilità «universale», «antropologica» o «ontologica»[9] per intendere quella forma di esposizione che ci accomuna tutti in quanto esseri umani e in quanto enti corporei sempre in relazione. Si tratta quindi sia di aspetti endogeni, legati alla natura finita dell'essere umano, sia di elementi esogeni, legati a condizioni sociali ed economiche.

La vulnerabilità particolare, invece, riguarda l'esposizione di persone o gruppi ad un danno addizionale. Questo secondo senso è oggetto privilegiato della riflessione bioetica e giuridica: alla generica tutela dei soggetti in quanto tali si affianca in modo più specifico un interesse per quegli individui che, in determinate situazioni, si trovano a subire a rischio di maggior danno a causa della loro età, della disabilità, di situazioni economiche sfavorevoli, dell'appartenenza a minoranze etniche o religiose, ecc.

Il modo in cui questa vulnerabilità particolare può essere intesa è duplice, e comporta significative conseguenze. Si riscontra infatti un senso di vulnerabilità particolare *categorizzante*, che tende a identificare dei gruppi e dei soggetti vulnerabili in base alle loro specifiche caratteristiche considerandoli *gruppi vulnerabili*[10]. Queste forme di vulnerabilità sono anche definite da Florencia Luna «*labels*», in quanto tendono a "etichettare" i soggetti vulnerabili identificandoli staticamente[11]. Tuttavia, questo primo senso tende a ridurre l'individuo alle sue specifiche vulnerabilità, e rischia di mettere in atto il cosiddetto «paradosso della vulnerabilità»[12]. Se così intesa, infatti, l'identificazione della vulnerabilità, piuttosto che rappresentare un fattore di tutela e protezione addizionale, può divenire, come la definiscono Rogers, Mackenzie e Dodds, «patologica»[13], creando stereotipi[14] e giustificando atteggiamenti paternalistici o addirittura discriminatori. Se infatti consideriamo la vulnerabilità come una proprietà attribuibile soltanto ad alcuni individui, allora emerge una contrapposizione tra uno stato di presunta "normalità", ossia l'autonomia, e dall'altro uno stato di eccezione, da correggere e da eliminare. Il secondo modo per intendere la vulnerabilità particolare, invece, è detto *situazionale*, in quanto si concentra sulle condizioni che rendono un soggetto vulnerabile, spesso molteplici fattori tra loro connessi che concorrono a esporre in modo maggiore alcuni individui rispetto ad altri. In questo se-

[9] È in particolare Martha Fineman a parlare di vulnerabilità universale evidenziandone il complesso rapporto con l'autonomia e le conseguenze pratiche che ciò implica sul terreno giuridico-politico (M. FINEMAN, *Reasoning from the Body: Universal Vulnerability and Social Justice*, in DIETZ, C., TRAVIS, M., THOMSON, M. (eds.), *A Jurisprudence of the Body*, London, 2020). Il riferimento all'aspetto antropologico lo troviamo ad esempio nel lavoro di H. TEN HAVE, *op.cit*. Infine, il piano «ontologico» è richiamato ad esempio da E. PARIOTTI, *Vulnerabilità ontologica e linguaggio dei diritti*, in *Ars Interpretandi*, 2, 2019, 155-170. Tra queste tre proposte prediligeremo quella di *universale*, poiché il riferimento all'ontologia presuppone la possibilità di identificare un'essenza, aspetto che esula le possibilità e gli obiettivi di questo contributo, mentre quello all'antropologia limita il quadro della vulnerabilità al solo terreno dell'umano, tagliando fuori la sfera, ad esempio, animale e più ingenerale quella del mondo vivente (cosa che meriterebbe di essere adeguatamente giustificata).

[10] Per un approfondimento si veda F. MACIOCE, *La vulnerabilità di gruppo. Funzioni e limiti di un concetto controverso,* Torino, 2021.

[11] F. LUNA, *Elucidating the Concept of Vulnerability. Layers not Labels*, *International Journal of Feminist Approaches on Bioethics*, 1, 2009, 120-138.

[12] M. G. FURNARI, *The paradox of vulnerability*, in *Medicina E Morale*, 71, 4, 2021, 425-445.

[13] W. ROGERS, C., MACKENZIE, S., DODDS, *Why bioethics needs a concept of vulnerability,* cit.

[14] E. PARIOTTI, *Vulnerabilità, approccio intersezionale e linguaggio dei diritti*, in *GenIUS,* 2024 (disponibile online: https://www.geniusreview.eu/wp-content/uploads/2024/02/Pariotti_Focus1.pdf); F.J. ARENA, *I due volti degli stereotipi nel diritto*, in *Notizie di Politeia*, 39, 149, 2023, 5-25.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

condo senso la vulnerabilità non costituisce un tratto identitario stabile e univoco, bensì una condizione multifattoriale e graduale, alla cui presenza concorrono fattori ambientali, economici, sociali e politici. Sempre Luna parla, in questo secondo caso, di *«layers»*, ossia di stratificazioni di vulnerabilità, che si uniscono e che si inscrivono in particolari contesti. Secondo questa modalità, quindi, non si tratta di classificare in modo statico gli individui, ma di riconoscere il carattere intersezionale di una serie di fattori che contribuiscono a far emergere la nostra particolare esposizione. Senza distinguere tra soggetti autonomi e soggetti vulnerabili, in questo caso si riconosce una generale vulnerabilità che emerge in modi ogni volta nuovi e peculiari. Non si tratta tanto, quindi, di soggetti vulnerabili, bensì di *relazioni* di vulnerabilità. Come dice giustamente McLean, non bisogna chiedersi *chi* sia vulnerabile, bensì *quando* e *come*[15].

Nel dibattito odierno, soprattutto in ambito bioetico e giuridico, rintracciamo entrambe le possibilità di interpretazione della vulnerabilità particolare, categorizzante e situazionale, anche se risulta ormai chiaro quanto la vulnerabilità categorizzante possa divenire un fattore discriminatorio ed eccessivamente cristallizzante.

Il rapporto tra senso universale e senso particolare (categorizzante o situazionale) può prevedere diverse combinazioni. L'utilizzo della sola categoria universale rischia, a parere di alcuni critici, di rendere tale nozione priva di effettività, eccessivamente astratta e risultando ridondante e inefficace. L'utilizzo esclusivo della vulnerabilità particolare, d'altro canto, favorisce una visione inesatta per cui al soggetto vulnerabile si contrappone il soggetto autonomo. La combinazione che meglio riesce a rappresentare la realtà e allo stesso tempo che rende più proficuo l'impiego del termine è a nostro parere l'unione del senso universale della vulnerabilità con quello particolare *situazionale*[16]. Si dà così un duplice livello, graduale e intersezionale, che garantisce non soltanto una tutela generalizzata di tutti i soggetti, ma anche una specifica per coloro che presentano un maggior grado di vulnerabilità. Ognuno di noi, infatti, si trova, in determinate situazioni o in determinati momenti della propria vita, ad essere vulnerabile. Siamo quindi tutti vulnerabili, ma in misure differenti in base alle nostre specifiche condizioni fisiche o socio-economiche, in base ai contesti e alle situazioni in cui ci troviamo a vivere.

Un esempio di questo specifico senso di vulnerabilità si può trovare in uno dei più importanti atti di organi internazionali in ambito bioetico, ossia la *Dichiarazione sulla bioetica e i diritti umani* redatta dall'UNESCO (2005), in particolare all'articolo 8:

In applying and advancing scientific knowledge, medical practice and associated technologies, human vulnerability should be taken into account. Individuals and groups of special vulnerability should be protected and the personal integrity of such individuals respected.[17]

[15] S. McLean, *Respect for human vulnerability and personal integrity*, in H. ten Have, B. Gordijn (eds.), *Handbook of Global Bioethics,* Dordrecht, 2014, 105–117.

[16] S. Dadà, *La nozione di vulnerabilità in bioetica: tra universalità e particolarità,* in *Il Paradosso*, 1, 2021, 89-104.

[17] UNESCO, International Bioethics Committee (IBC), *Universal Declaration on Bioethics and Human Rights,* Paris, 2005 (disponibile al link https://www.unesco.org/en/legal-affairs/universal-declaration-bioethics-and-human-rights?hub=66535 ultima consultazione 30/11/2024).

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Possiamo osservare qui la compresenza dei due piani di vulnerabilità: quella che accomuna tutti gli individui, che deve essere presa in considerazione e tutelata, e poi una «special vulnerability», che riguarda individui e gruppi in particolari situazioni di svantaggio o di esposizione a danni[18].

2.2 Valore normativo della vulnerabilità

Sebbene si tenda a considerare la vulnerabilità esclusivamente come una *condizione*, e per di più spiacevole e da evitare, tale nozione possiede anche una dimensione positiva, e quindi una funzione normativa[19]. Questo concetto, infatti, non si limita a descrivere degli stati di cose, ma attiva delle risposte a tali situazioni, risposte caratterizzate, in particolare, dalla *responsabilità* e dalla *cura.* Da questo punto di vista, Robert Goodin[20] presenta un'idea di vulnerabilità relazionale ed introduce il cosiddetto «vulnerability model», su cui si fondano gli impegni di ogni membro della società. Ci troviamo da sempre coinvolti in relazioni che esulano dalla semplice assunzione volontaristica di responsabilità: nell'incontro tra un agente morale e un paziente morale, il secondo è esposto alle conseguenze delle azioni del primo. Le responsabilità crescono in base a quanto si è pazienti, il che fa sorgere una gradualità di responsabilità che ogni soggetto è chiamato a sostenere. La protezione dei vulnerabili è una questione pubblica, attorno a cui si devono concentrare gli sforzi egli individui, dei soggetti e delle istituzioni, nell'elaborazione di un sistema adeguato di cura e tutela[21].
Secondo queste prospettive[22], quindi, l'obiettivo di ogni individuo e della collettività deve essere quello di proteggere la vulnerabilità di ognuno, di mitigarne l'impatto, dove possibile, e di resistervi, laddove è generata da contesti di discriminazione e diseguaglianza[23]; ma nello stesso tempo è proprio a partire dalla sua percezione che possono realizzarsi forme creative e solidali di vivere in comune. La vulnerabilità, infatti, è espressione sia delle nostre differenze che di diseguaglianze. Essa ci permette, attraverso l'analisi e la considerazione dei contesti e delle situazioni, di distinguerle, valorizzando le prime e eliminando le seconde identificandone le cause. Essa consente così una tutela specifica e particolare dei soggetti al plurale. Analizzato in base alle sue vulnerabilità, il soggetto si riappropria della sua dimensione materiale e pratica, permettendo un intervento di cura e tutela situato e effica-

[18] UNESCO, International Bioethics Committee (IBC), *The Principle of Respect for Human Vulnerability and Personal Integrity: Report,* Paris, 2009, 1-54; F. LUNA, *La declaración de la UNESCO y la vulnerabilidad, la importancia de la metáfora de las capas,* in M. CASADOS (ed.), *Sobre la Dignidad y los Principios. Análisis de la Declaración Universal de Bioética y Derechos Humanos de la UNESCO*, Pamplona, 2009, 255-266.

[19] S. ZULLO, *Lo spazio sociale della vulnerabilità tra pretese di giustizia e pretese di diritto. Alcune considerazioni critiche*, in *Politica del diritto*, 6, 2016, 488.

[20] R. GOODIN, *Protecting the Vulnerable: A Reanalysis of Our Social Responsibilities*, Chicago, 1985.

[21] Anche nel pensiero femminista, soprattutto nell'ambito della proposta dell'etica della cura, la vulnerabilità e la dipendenza hanno incontrato una rivalutazione positiva, divenendo il perno teorico su cui costruire delle proposte etico-politiche innovative. Sul tema si veda C. GILLIGAN, *Con voce di donna. Etica e formazione della personalità*, Milano, 1991, V. HELD, *The Ethics of Care: Personal, Political, Global,* Oxford 2006; e S. TUSINO, *L'etica della cura. Un altro sguardo sulla filosofia morale*, Milano, 2021.

[22] Oltre a quelli qui citati molte altre correnti e autori hanno evidenziato l'importanza di questo concetto. Per un quadro complessivo si rimanda a P. DONATELLI, *Vulnerabilità e forme di vita*, in *Etica & Politica / Ethics & Politics*, 58, 3, 2016, 59-74 e a S. DADÀ, *Etica della vulnerabilità,* Morcelliana, Brescia 2022.

[23] Sul legame tra vulnerabilità e resistenza si veda S. BARCKE, *Bouncing Back. Vulnerability and Resistance,* in J. BUTLER, Z. GAMBETTI, L. SABSAY (eds.), *Vulnerability in Resistance*, Durham - London 2016, 52-75.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

ce. Quindi, con le parole di Martha Fineman, una teoria della vulnerabilità può divenire «a powerful conceptual tool with the potential to define an obligation for the state to ensure a richer and more robust guarantee of equality than is currently afforded under the equal protection model»[24].

2.3 Vulnerabilità come principio e il rapporto coi diritti umani

Questa categoria, quindi, possiede un'intrinseca potenza concretizzante, poiché permette di riscontrare lo statuto e le specifiche discriminazioni permettendo così di elaborare un'adeguata risposta in base al caso. Proprio pe questo motivo, c'è chi ha parlato di un vero e proprio «principio di vulnerabilità», in quanto orienta l'azione, e favorisce politiche di potenziamento e prevenzione[25]. C'è chi, come Turner, vede nella vulnerabilità il fondamento stesso dei diritti umani: con un'argomentazione "hobbesiana", egli sostiene che la dotazione di tali diritti è direttamente dipendente dalla scoperta della propria fragilità e dalla necessità di contrastarla con un apparato istituzionale e giuridico apposito[26]. Simile la posizione espressa da Francesca Ippolito, che considera la vulnerabilità un principio per il diritto internazionale e per i diritti umani, che lavora in sinergia con altri principi (quali la dignità) e che ha la funzione euristica di interpretazione e orientamento delle norme esistenti, ottimizzandole e concretizzandole in direzione di *empowerment* e responsabilità[27].
Non mancano le critiche a questo tipo di visione. Ad esempio, Michael Kottow sostiene che tale nozione sia un concetto esclusivamente descrittivo, privo di forza normativa intrinseca e che la ottenga solo se associata ad altri principi[28]. Anche Elena Pariotti riconosce al concetto una potenzialità normativa «indiretta», interpretandolo come una categoria euristica utile per concretizzare i principi, e a valorizzare in senso relazionale i diritti umani[29]. Il riferimento alla vulnerabilità, in questo senso, garantisce la concretizzazione e l'individuazione di contesti e condizioni che la causano, favorisce la decostruzione e il ripensamento di categorie quali l'autonomia, e per di più agisce nei punti ciechi dei sistemi giuridici, facendo emergere condizioni che non sono ancora oggetto di tutela da parte dei diritti[30]. Anche Roberto Adorno, criticando Turner, sostiene che il principio fondamentale sia la dignità, mentre la vulnerabilità sia una condizione, ma non la causa della promozione di tali principi[31]. Baldassare Pastore sintetizza una simile posizione in modo chiaro: «I diritti umani, allora, possono essere

[24] M. Fineman, *The vulnerable subject: Anchoring equality in the human condition*, in *Yale Journal of Law & Feminism*, 20, 2008, 1-24.

[25] Un esempio di vulnerabilità intesa come principio si dà nella *Dichiarazione di Barcellona*, del 1998, in cui essa appare al fianco dell'autonomia, della dignità e dell'integrità.

[26] B. S. Turner, *Vulnerability and Human Rights,* University Park, 2006.

[27] F. Ippolito, *La vulnerabilità quale principio emergente nel diritto internazionale dei diritti umani*, in *Ars Interpretandi*, 2, 2019, 63-93.

[28] M. Kottow, *Vulnerability: what kind of principle it is?*, in *Medicine, HealthCare and Philosophy*, 7, 2004, 281-287.

[29] E. Pariotti, *Vulnerabilità ontologica e linguaggio dei diritti*, cit.

[30] Si parla, in questo senso di «fragilità istituzionale» (P. De Stefani, *Conceptualizing Vulnerability in the European Legal Space: Mixed Migration Flows and Human Trafficking as a Test,* in *Frontiers in Human Dynamics*, 4, 2022,861178).

[31] R. Adorno, *Is Vulnerability a Foundation of Human Rights?,* in A. Masferrer, E Garcìa-Sànchez (eds), *Human Dignity of the Vulnerable in the Age of Rights*, Zurich, 2016, 257-272.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

considerati come il risultato della confluenza di due fattori: uno *normativo* (l'intrinseco valore di ogni persona) e uno *fattuale* (la fragilità umana e la suscettibilità del danno)»[32].

3. Vulnerabilità nell'AI Act

Come abbiamo appena visto, quindi, la vulnerabilità presenta un'ampia gamma di sensi, interpretazioni e modalità di utilizzo. Tra di essi, quella che risulta più proficua è l'unione tra un senso universale e uno particolare situazionale, che permetta sia una tutela specifica e particolare di alcuni soggetti o gruppi, ma evitando di contrapporli ad uno standard di normalità basato su un soggetto assolutamente autonomo. Possiamo quindi vedere in questa unione tra universale e particolare un senso ibrido, contestuale e relazionale[33]. Soprattutto oggi, nell'ambiente digitale, ci si rende sempre più conto che un soggetto privo di vulnerabilità non è infatti concepibile: come sostiene Mark Coeckelbergh il rapporto con le nuove tecnologie ha raggiunto un grado di pervasività tale da rendere la vulnerabilità il tratto esistenziale caratterizzante dell'essere umano, tanto da trasformarlo in un vero e proprio «human-being-at-risk»[34].

Come sottolineano Malgieri e Niklas nel loro studio dedicato al Regolamento dell'Unione Europea 679/2016 sulla Protezione dei Dati (GDPR), parlare oggi di «average data subject» informato, consapevole e accorto, distinto da un «vulnerable data subject», risulta altamente inverosimile. Infatti nessuno può effettivamente dirsi al riparo da forme di persuasione, manipolazione e controllo tipiche dell'attuale rivoluzione digitale, così come nessuno può dirsi completamente informato e consapevole, a causa dello squilibrio di informazioni tra coloro che possiedono ed elaborano i dati e i soggetti portatori degli stessi[35]. Questo aumento delle vulnerabilità causato dall'ambiente digitale non è semplicemente un'inaspettata conseguenza, ma un prodotto previsto dalla stessa struttura dei sistemi. Si può per questo parlare, come fanno Helberger e colleghi, di una vulnerabilità «architettonica»[36].

Siamo quindi più esposti a pericoli e rischi a causa delle nuove tecnologie e in particolare dell'utilizzo dell'IA: a tale aumento della vulnerabilità dovrebbe corrispondere un'attenzione particolare a tale tema anche in ambito giuridico.

Dopo aver quindi delineato interpretazioni, significati e usi del concetto, passiamo adesso ad indagare il ruolo della vulnerabilità nell'ambito del digitale, con un focus sulla regolamentazione europea riguardo all'Intelligenza Artificiale e in particolare sul documento di recente approvazione, l'*AI Act.* Si tratta della prima regolamentazione in termini generali dell'IA, esito di un articolato processo pre-

[32] B. Pastore, op.*cit.*, 38. Di simile avviso anche A. Timmer, *A Quiet Revolution: Vulnerability in the European Court of Human Rights,* in M. A. Fineman, A. Grear (eds.), *Vulnerability. Reflection on a New Ethical Foundation for Law and Politics,* Farnham-Blurington, 2013, 147-170.

[33] G. Malgieri, J. Niklas, *Vulnerable Data Subjects*, in *Computer Law and Security Review*, 37, 2020, 105415. Per un approfondimento sullo stesso tema un riferimento indispensabile per chiarezza e completezza è certamente G. Malgieri, *Vulnerability and Data Protection Law,* Oxford, 2023.

[34] M. Coeckelbergh, *Human being@risk: Enhancement, Technology, and the Evaluation of Vulnerability Transformations*, London, 2015.

[35] G. Malgieri, *Vulnerability and Data Protection Law,* cit.

[36] N. Helberger, M. Sax, J. Strycharz, H.-W. Micklitz, *Choice Architectures in the Digital Economy: Towards a New Undertanding of Digital Vulnerability*, in *Journal of Consumer Policy*, 45, 2022, 187.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

paratorio giunto dopo l'emanazione di numerosi atti di *soft law* in materia di intelligenza artificiale.[37] Gli obiettivi che si propone sono quello di abbattere le barriere alla creazione di un mercato unico nell'UE, favorendo sia la produzione e la circolazione di un prodotto sicuro, nel rispetto dei diritti fondamentali dell'UE; e che favorisca l'innovazione tecnologica in quest'ambito (art.1 co.1). Proprio questa esigenza di proporzionalità tra tutela della sicurezza e sviluppo tecnologico ed economico fa sì che vengano introdotte limitazioni e regole differenti secondo un approccio basato sul rischio. In questa architettura, in cui risulta centrale la questione del rischio rispetto alla salute, alla sicurezza e ai diritti fondamentali, ha senso chiedersi quale sia il ruolo della vulnerabilità.
Per questo ci interesseremo sia alle specifiche ricorrenze del termine all'interno della regolamentazione, sia in senso più ampio alle strategie di tutela messe in atto nel documento.

3.1 Ricorrenze del termine

Per svolgere un'indagine sulla ricorrenza, nel documento, del concetto di vulnerabilità, non ci siamo limitati a considerare la sua forma sostantiva («vulnerabilità»), ma abbiamo ricercato anche la presenza di espressioni quali «persone vulnerabili» e «gruppi vulnerabili». Il termine appare 25 volte nella versione approvata dell'AI Act, tenendo conto sia dei considerando (15) che degli articoli del regolamento (10). Di questi riferimenti, 16 sono relativi alla vulnerabilità umana mentre 8 riguardano la vulnerabilità dei sistemi, dei dati o della infrastruttura TIC sottostante e la cybersicurezza[38]. Ci concentreremo qui sulla vulnerabilità umana, pur sottolineando che non possiamo totalmente distinguere i due piani e che un'analisi congiunta offrirebbe ulteriori elementi per un quadro più completo.
Entrando quindi in merito alla vulnerabilità umana, va innanzitutto notato che all'art.3 sono elencate ben 68 definizioni[39], ma nessuna riguarda la vulnerabilità e le espressioni ad essa connesse[40]. Possiamo quindi dedurre da questa assenza che il legislatore europeo non abbia visto in questo termine un concetto chiave di cui fornire indicazioni preliminari sul suo specifico utilizzo, né abbia avvertito il rischio di fraintendimenti e ambiguità rispetto ad esso, lasciando il termine ad un uso spontaneo. Possiamo però prendere in esame altri documenti connessi al regolamento, riscontriamo tale assenza anche nel GDPR, mentre nel Glossario contenuto nelle Linee Guida per un'IA affidabile (aprile 2018), redatto dal gruppo indipendente di esperti ad alto livello sull'intelligenza artificiale (HLEG on AI) istituito dall'Unione Europea[41], troviamo inserita l'espressione «persone o gruppi vulnerabili». Nella de-

[37] Tra questi vanno ricordate le risoluzioni del Parlamento europeo sui principi etici dell'IA, della robotica e della tecnologia correlata e sul regime di responsabilità civile per l'IA (20 ottobre 2020) e sull'uso dell'IA (20 gennaio 2021). Un altro passaggio importante è stato il Libro Bianco sull'Intelligenza artificiale della Commissione (19 febbraio 2020) le cui linee generali sono state discusse attraverso un'intensa fase di consultazioni, conclusasi nel maggio del 2020. Ricordiamo infine le linee guida in materia di Trustworthy AI (8 aprile 2019) elaborate dall'High-Level Expert Group on AI.

[38] Considerando 76 e 110, art. 15 co 5.

[39] Tra queste anche quella assai dibattuta di «sistemi di Intelligenza artificiale»: cfr. C. TRINCADO CASTÁN, *The legal concept of artificial intelligence: the debate surrounding the definition of AI System in the AI Act,* in *BioLaw citazione incomplete,* 1, 2024, 305-44.

[40] AI Act, art.3.

[41] High Level Expert Group on Artificial Intelligence, EU, *Draft Ethics Guidelines for Trustworthy AI*, aprile 2019 (consultabile al link: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai).

Downloaded from www.biodiritto.org.
ISSN 2284-4503

scrizione vengono elencati solo alcuni dei fattori determinati, sostenendo che non esiste una definizione comunemente accettata di questa categoria «estremamente eterogenea». La vulnerabilità viene collegata qui allo specifico contesto, con riferimento alla Carta dei diritti fondamentali dell'Unione Europea. La scelta degli esperti ad alto livello di considerare i gruppi e persone vulnerabili e non la vulnerabilità rivela un'interpretazione in senso *particolare* del concetto. C'è poi da sottolineare che il riferimento al contesto fa propendere per un'interpretazione *situazionale* della vulnerabilità.

Possiamo assumere che anche la regolamentazione europea sull'IA tenga conto di tale descrizione[42]. Anche nell'AI Act si riscontra, infatti, un senso principalmente *particolare* della vulnerabilità. Essa si trova primariamente associata alla persona o al gruppo di persone. I gruppi che vengono citati sono primariamente i minori (considerando 29 e 48, art.9c.9), i disabili (considerando 29 e 165), le persone dipendenti da prestazioni e servizi essenziali (considerando 58), i migranti (considerando 60), minoranze razziali o etniche (considerando 67), persone in specifiche situazione sociale o economica (considerando 29, art.5, co 1 lett.b, art.7, lett.h).

È interessante comprendere se questo senso di vulnerabilità particolare sia declinato in una direzione categorizzante o situazionale. Soffermiamoci sui due articoli dove la vulnerabilità viene tematizzata in modo più ampio, ossia l'art.5 co 1 lett. b, e l'art. 7 co 2 lett. h.

L'art. 5 è dedicato alle pratiche di IA vietate e considerate a rischio inaccettabile. Tra queste pratiche, alla lettera b, troviamo un riferimento alla vulnerabilità:

b) l'immissione sul mercato, la messa in servizio o l'uso di un sistema di IA che sfrutta le vulnerabilità di una persona fisica o di uno specifico gruppo di persone, dovute all'età, alla disabilità o a una specifica situazione sociale o economica, con l'obiettivo o l'effetto di distorcere materialmente il comportamento di tale persona o di una persona che appartiene a tale gruppo in un modo che provochi o possa ragionevolmente provocare a tale persona o a un'altra persona un danno significativo;[43]

Le vulnerabilità di persone e gruppi, quali minori e disabili, inducono a pensare a un approccio categorizzante, simile ai *labels* descritti da Luna. A tal proposito l'elenco risulta limitato e quindi non esaustivo, escludendo vari soggetti e gruppi esposti a discriminazione e altre forme di danno (quali minoranze razziali, etniche religiose e linguistiche, gruppi LGBTQI+ e lavoratori)[44]. Tuttavia, il riferimento alla situazione economica e sociale sposta maggiormente in direzione di un approccio situazionale, in cui la vulnerabilità del soggetto è determinata dalle sue relazioni, andandosi combinare i vari *labels* in base al contesto. Ciò si riscontra in modo ancora più esplicito all'art. 7 comma 2 lettera h. Tra i criteri stabiliti per modificare e ampliare l'elenco dei sistemi ad alto rischio, la Commissione tiene conto di:

h) la misura in cui esiste uno squilibrio di potere o le persone che potrebbero subire il danno o l'impatto negativo si trovano in una posizione vulnerabile rispetto al deployer di un sistema di IA, in

[42] Nel considerando 27 dell'AI Act viene esplicitamente richiamato l'HLEG on AI in riferimento agli orientamenti e ai principi.

[43] AI Act, art.5, co 1, lett. b.

[44] G. MALGIERI, *Human Vulnerability in the EU Artificial Intelligence Act,* in *OUP blog,* 27 maggio 2024 (consultabile al link https://blog.oup.com/2024/05/human-vulnerability-in-the-eu-artificial-intelligence-act/ ultima consultazione 30/11/2024).

Downloaded from www.biodiritto.org.
ISSN 2284-4503

particolare a causa della condizione, dell'autorità, della conoscenza, della situazione economica o sociale o dell'età.[45]

Il riferimento alla *posizione* di vulnerabilità, piuttosto che ai gruppi vulnerabili, rimanda direttamente a un approccio situazionale che è accentuato dal richiamo allo squilibrio di potere, presente anche nei considerando 44 e 59. Tale asimmetria può essere dovuta a vari fattori, sia inerenti che di carattere economico e sociale. A parere di Malgieri l'espressione «in particolare» porterebbe ad interpretare questo riferimento alla vulnerabilità in senso universale[46]. Tuttavia, la specificazione delle cause è riferita alle persone che si trovano già in una posizione vulnerabile rispetto al deployer[47]. Non si tratta quindi di un riferimento alla vulnerabilità universale, ma di una vulnerabilità particolare, situazionale, che può essere generata *in particolare* dalle cause elencate[48].

Vi è un solo riferimento in cui, a nostro parere, si può riscontrare un senso universale di vulnerabilità, connesso alla vulnerabilità particolare in senso situazionale. In conclusione al considerando 29 si parla dei danni causati dalle tecniche di manipolazione basate sull'IA. Tali danni possono sovvertire o pregiudicare l'autonomia, il processo decisionale e la libera scelta:

In aggiunta, i sistemi di IA possono inoltre sfruttare in altro modo le vulnerabilità di una persona o di uno specifico gruppo di persone dovute all'età, a disabilità ai sensi della direttiva (UE) 2019/882 del Parlamento europeo e del Consiglio o a una specifica situazione sociale o economica che potrebbe rendere tali persone più vulnerabili allo sfruttamento, come le persone che vivono in condizioni di povertà estrema e le minoranze etniche o religiose.

La vulnerabilità, quindi, è connessa qui sia a caratteristiche proprie dei soggetti (età e disabilità), che a situazioni che possono rendere i soggetti *più* vulnerabili, il che presuppone che tutti i soggetti siano vulnerabili e alcuni lo siano in maggior grado.

Si può quindi concludere che la vulnerabilità, all'interno del documento, mantiene sia il riferimento a gruppi vulnerabili quali minori e disabili, che alle condizioni e situazioni di vulnerabilità. Senza quindi rinunciare al riferimento ai gruppi e alle categorie vulnerabili, essa introduce un senso situazionale, intersezionale, e relazionale, con una particolare attenzione alle situazioni economiche e sociali, così come agli squilibri di potere. I riferimenti alla vulnerabilità universali sono pressoché assenti o impliciti.

[45] AI Act, art.7 co 2 lett. h.

[46] G. MALGIERI, *Vulnerability and Data Protection Law*, cit.

[47] La posizione dell'autore si giustifica alla luce della sua lettura della vulnerabilità situazionale come di per sé unione di universale e particolare. Sebbene concordiamo sulla maggior compatibilità di universale e particolare situazionale, crediamo che non si possa dedurre il riconoscimento di una vulnerabilità universale dalla presenza di un riferimento situazionale, e che quindi i due concetti non siano necessariamente congiunti, come in questo caso.

[48] Una situazione analoga si incontra nel considerando 60, in cui si fa riferimento alla governance e della gestione dei dati onde evitare distorsioni e discriminazioni «in particolare nei confronti delle persone vulnerabili che appartengono a determinati gruppi […]». Anche in questo caso, sebbene il danno sia potenzialmente universale, la vulnerabilità è chiamata in causa esplicitamente soltanto in riferimento a gruppi particolari.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

3.2 Evoluzione e genesi dalla prima proposta di regolamento

Come è ben noto, l'attuale versione dell'AI Act è l'esito di una lunga fase di consultazione che ha avuto luogo a partire dal 2021 e che ha comportato sostanziali cambiamenti nel testo del documento. Anche il ruolo del concetto di vulnerabilità ha quindi subito delle modificazioni che vale la pena considerare, per comprendere meglio l'attuale quadro.

Osservando in modo sinottico la prima versione della proposta della Commissione Europea del 2021, e la versione finale approvata nel 2024 si possono quindi notare alcune differenze. I principali cambiamenti derivano dagli emendamenti del Parlamento Europeo (2023), e testimoniano un accrescimento del ruolo della vulnerabilità nel documento. Si assiste inoltre all'esigenza espressa dal Parlamento di un passaggio graduale da un senso primariamente categorizzante a uno più situazionale. Tali emendamenti sono stati accolti nella maggior parte dei casi, e sono quindi andati a confluire nella versione approvata.

Innanzitutto riscontriamo un aumento delle ricorrenze: si passa da 11 (di cui 4 relative alla vulnerabilità dei sistemi) della prima proposta alle 24 (di cui 8 dei sistemi) della versione finale approvata[49]. Ciò si spiega solo in parte alla luce dell'espansione del documento, passato da 85 articoli a 113 articoli, suggerendo anche un aumento dell'interesse per tale nozione. È il caso di notare che in questa espansione delle ricorrenze si attua in concomitanza all'accresciuto riferimento all'impatto sui diritti fondamentali (considerando 93 e art.27[50]).

L'evoluzione da un senso maggiormente categorizzante a uno più situazionale e relazionale è attestato dal passaggio dal riferimento alla vulnerabilità dei bambini e dei disabili a uno riguardo alla *posizione di vulnerabilità* causate dall'età e dalla disabilità (considerando 29). Nella prima versione, inoltre, le categorie considerate si limitano principalmente queste due, mentre nelle successive versioni troviamo un riferimento più ampio anche a minoranze etniche e religiose (considerando 29) e anche un generico ad «altri gruppi vulnerabili», lasciando maggior margine di interpretazione (art.9 co 9). In particolare possiamo osservare il cambiamento dell'art. 5 comma 1 lettera a. Nella prima versione ci si riferisce esclusivamente alle vulnerabilità «di uno specifico gruppo di persone dovute all'età o alla disabilità fisica o mentale». Nella versione approvata, invece, si aggiunge il riferimento alla specifica situazione sociale ed economica.

3.3 Vulnerabilità universale e approccio basato sul rischio

Il regolamento adotta un *approccio basato sul rischio*[51], individuando quattro categorie di IA diverse sottoponendole ad altrettanti regimi regolatori (sistemi *a rischio inaccettabile*, sistemi *ad alto rischio*, sistemi con rischio per la *trasparenza*, sistemi a *basso* o a *minimo rischio*).

[49] La vulnerabilità umana viene introdotta specificamente nei considerando 29, 48, 67, 93, 132, 141, 165 e negli artt. 9 co 9, 60 co 4 lett. g, 79 co 2, 95 co 2 lett. e.

[50] Si noti che negli emendamenti all'art. 27 era presente anche un riferimento alla vulnerabilità, mentre nella versione approvata è stato tolto.

[51] M.E. Kaminski, *Regulating the Risks of AI,* in *Boston University Law Review*, 103, 2023, 1347-1411.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Se si accetta la definizione proposta da Luna[52], per cui la vulnerabilità si definisce in termini di rischio come il prodotto della probabilità che si presenti un danno e la dannosità[53], si può pensare che l'approccio basato sul rischio introduca implicitamente una specifica modalità di vulnerabilità, quella universale. Secondo questa argomentazione, se siamo tutti esposti ai rischi provocati dall'IA, allora l'approccio basato sul rischio è anche un approccio volto alla tutela generale delle vulnerabilità.

A nostro parere, invece, per quanto vicine e connesse, tali nozioni non sono del tutto sovrapponibili, e di conseguenza l'approccio basato sul rischio deve essere distinto dalla tutela della vulnerabilità universale. Il rischio, come si è specificato nella definizione 2 dell'art.3, è il prodotto della dannosità e della probabilità che avvenga un determinato fenomeno. Esso ha una dimensione *oggettiva*, in quanto riguarda *ciò* che accade e si rivolge all'oggetto che causa tale danno, in questo caso il sistema di IA. La vulnerabilità, invece, è una nozione *soggettiva,* in quanto pone il focus su colui che subisce il rischio, sui contesti e le specifiche situazioni che causano la sua esposizione al danno[54]. Come infatti sottolinea Maceinate, di fronte allo stesso rischio ognuno è esposto in modo differente e non tutti fanno esperienza dello stesso danno[55]. Per comprendere quindi la differenza tra rischio e vulnerabilità bisogna tenere presente che, con le parole di Kaminski, «the choice to use risk regulation reflects a particular epistemology: the notion that such AI systems are just math, uncovering some ground truth then contingent social facts»[56]. La vulnerabilità invece serve a orientare e a concretizzare il rispetto dei diritti fondamentali che l'approccio basato sul rischio affronta rispetto all'impatto generale dei sistemi sul generico soggetto. Le due nozioni non sono quindi equivalenti, bensì complementari. Porre al centro il rischio significa quindi, come nel caso del AI Act, regolamentare il prodotto[57], per far sì che l'impatto della sua messa in mercato e del suo utilizzo non provochi danni alla salute, alla sicurezza e ai diritti fondamentali. Considerare primariamente la vulnerabilità, invece, non si esaurisce nell'azione regolatrice sul prodotto, ma anche uno sguardo complessivo al soggetto, alle cause della sua posizione di svantaggio, agli aspetti anche contestuali che lo rendono esposto. Per eliminare, mitigare e tutelare la vulnerabilità, insomma, serve un'azione più profonda, in cui l'intervento sul rischio è soltanto un aspetto connesso ad altre misure politiche, economiche e sociali rivolte in modo diretto ai soggetti. Intervenire sul rischio non elimina la vulnerabilità, bensì la sola probabilità del danno provocato dall'IA. L'esposizione del soggetto, poi, rimane in molti casi invariata, in quanto precede l'interazione col sistema e il soggetto, proprio in virtù di una situazione di partenza svantaggiata, si trova a subire un danno addizionale nel contesto dell'IA. L'approccio basato sul rischio non neutralizza la vulnerabilità, ma solo una sua specifica possibilità di acuirsi a causa dell'utilizzo del si-

[52] F. Luna, *Identifying and evaluating layers of vulnerability – a way forward*, in *Developing World Bioethics,* 19, 2019, 86–95.

[53] Questa è l'opinione di G. Malgieri, *Vulnerability and Data Protection Law,* cit.; e È. Gennet, R. Andorno, B. Elger, *Does the new EU Regulation on clinical trials adequately protect vulnerable research participants?*, in *Health Policy,* 119, 7 2015, 925-31.

[54] M. Coeckelbergh, *op. cit.*

[55] M. Maceinate, *The 'Riskification' of European Data Protection Law through a two-fold Shift*, in *European Journal of Risk Regulation*, 2017, 506-536.

[56] M.E. Kaminski, *op. cit.,* 1400.

[57] M. Almada, N. Petit, *The EU AI act: a medley of product safety and fundamental rights?*, in EUI, RSC, *Working Paper*, 59, 2023 (https://hdl.handle.net/1814/75982).

Downloaded from www.biodiritto.org.
ISSN 2284-4503

stema. Non possiamo poi ignorare il fatto che ogni volta che la vulnerabilità è citata nel testo essa richiama al suo senso particolare: per quanto si possa ipotizzare un richiamo implicito o sottointeso alla vulnerabilità universale tramite il riferimento al rischio, l'assenza del termine è significativa e induce a pensare che nel regolamento si preferisca sostituire la vulnerabilità universale con altri concetti (discriminazione, violazione dei diritti, manipolazione e inganno, ecc.).
L'IA Act, quindi, ricorrendo a un approccio basato sul rischio, introduce e si preoccupa della vulnerabilità, ma non possiamo ricavare da questo una completa sovrapposizione tra i due piani.

4. Conclusioni

Abbiamo ricostruito gli aspetti più significativi del dibattito attorno al concetto di vulnerabilità. In particolare abbiamo distinto tra un senso universale e uno particolare, e tra un approccio categorizzante e uno situazionale. In seguito, abbiamo analizzato lo specifico ambito della regolamentazione sull'IA, riscontrando che pur non essendo definita e tematizzata in modo specifico, la vulnerabilità trova uno spazio significativo nell'AI Act. Essa è intesa principalmente in senso particolare, sia con riferimento a specifici gruppi di persone vulnerabili che alla posizione di vulnerabilità dovuta a situazioni economiche e sociali o allo squilibrio di potere. Dall'analisi della genesi del documento, considerando la prima proposta, gli emendamenti del Parlamento Europeo e la versione approvata, abbiamo visto il progressivo passaggio da un senso categorizzante a uno più situazionale. La concezione universale non trova spazio nel documento, se non in modo implicito. Tale assenza sembra poter essere solo parzialmente compensata dall'approccio basato sul rischio, non essendo i due concetti completamente sovrapponibili.
Possiamo, a questo punto, concludere con alcune considerazioni. Il primo aspetto che risulta da questa indagine è la necessità di introdurre una definizione del concetto. La vulnerabilità ha infatti assunto un ruolo sempre più centrale in vari ambiti del diritto, ma con un significato aperto a varie interpretazioni. Senza un'indicazione e un accordo specifici rispetto al suo significato rischia di rimanere un termine eccessivamente vago, destinandolo a un uso ambiguo o spontaneo.
Dall'evoluzione del documento poi, possiamo osservare un progressivo allargamento, nell'AI Act, del ruolo della vulnerabilità in particolare nel suo senso situazionale. Questo ci sembra un risultato apprezzabile, in quanto, come abbiamo visto, permette di evitare esiti discriminatori e fa emergere il ruolo delle criticità contestuali nell'esposizione ai rischi dell'IA. Tuttavia, la mancata integrazione con un senso universale di vulnerabilità limita l'intervento di tutela al solo rischio del prodotto, e lo restringe ai soli casi di vulnerabilità particolare.
Un ulteriore sforzo in termini di accrescimento del ruolo di questo termine, anche in senso universale, permetterebbe interpretare in modo più unitario le varie condizioni di vulnerabilità, elaborando una strategia di contrasto più organica.[58]

[58] Si ringrazia il revisore anonimo per le preziose indicazioni, che hanno contribuito al miglioramento del presente lavoro.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Special issue

The Many Meanings of Vulnerability in the AI Act and the One Missing

Federico Galli, Claudio Novelli*

ABSTRACT: This paper reviews the different meanings of vulnerability in the AI Act (AIA). We show that the AIA follows a rather established tradition of looking at vulnerability as a trait or a state of certain individuals and groups. It also includes a promising account of vulnerability as a relation but does not clarify if and how AI changes this relation. We spot the missing piece of the AIA: the lack of recognition that vulnerability is an inherent feature of all human-AI interactions, varying in degree based on design choices and modes of interaction. Finally, we show how such a meaning of vulnerability may be incorporated into the AIA by interpreting the concept of "specific social situation" in Article 5 (b).

KEYWORDS: Vulnerability; AI Act; AI; Human-Computer Interaction; Specific Social Situation.

SUMMARY: 1. Introduction: Vulnerability and the AI Act – 2. The Underlying Meaning: Vulnerability as a Key Factor in the AIA's Objectives and Risk Analysis – 3. Vulnerability as a Trait or a Situation of Persons/Groups that Can be Exploited – 4. Vulnerability as a Feature of (High-Risk) AI Systems – 5. Vulnerability as a Trait (and a Situation?) of Affected Persons That Can Be Impacted – 6. Vulnerability as a Power Relation – 7. The Missing Piece: Vulnerability Stemming from Human-Computer Interaction – 8. Room for Manoeuvre: Looking at HCI as a "specific social situation" – 9. Conclusion.

1. Introduction: Vulnerability and the AI Act

The uptake of AI and digital technologies, coupled with the increased awareness of their risks to human beings, has revamped the interest in the concept of vulnerability within the legal field. The discussion has taken place both at empirical and regulatory levels.

* *Federico Galli: Research assistant, University of Bologna. Mail: federico.galli7@unibo.it; Claudio Novelli: Claudio Novelli: postdoctoral researcher, University of Bologna. Mail:* claudio.novelli@unibo.it*.* *While the research results are based on a combined effort, Sections 2, 4 and 5 should be attributed to Claudio Novelli, while Sections 3, 6-8 should be attributed to Federico Galli. Introduction and conclusion are shared. Federico Galli was partially supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (GA. 833647) and by the PRIN 2022 PNRR Project DAFNE (P2022R7RS9) under the MUR National Recovery and Resilience Plan funded by the European Union – Next Generation EU. The article was subject to a double-blind peer review process.*

Downloaded from www.biodiritto.org.
ISSN 2284-4503

At the empirical level, research has focused on the impact of the deployment of AI in specific areas, like finance[1], social networks[2], and dispute resolution[3]. Moreover, research has shown that AI systems can exacerbate existing inequalities and disproportionately affect already disadvantaged groups in society, such as gender minorities, low-income individuals, and those with limited competence with technology[4].

At the regulatory level, the discussion is pivoting around a key question: to what extent the vulnerability concept can represent a normative benchmark for different AI-powered contexts and practices, thereby requiring enhanced protection[5]. In other words, this would mean establishing new legal standards and safeguards designed to protect individuals and groups susceptible to harm due to AI technologies.

Some recent EU regulatory initiatives establishing legal frameworks for AI development and use have increasingly referred to the concept of vulnerability[6]. Among these, the recently adopted EU Artificial Intelligence Act (henceforth, AIA)[7] seems to be the one taking the idea of vulnerability most seriously.

While the research results are based on a combined effort, Sections 2, 4 and 5 should be attributed to Claudio Novelli, while Sections 3, 6-8 should be attributed to Federico Galli. Introduction and conclusion are shared. Federico Galli was partially supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (GA. 833647) and by the PRIN 2022 PNRR Project DAFNE (P2022R7RS9) under the MUR National Recovery and Resilience Plan funded by the European Union – Next Generation EU.

[1] E. Mogaji, T.O. Soetan, T.A, Kieu, *The implications of artificial intelligence on the digital marketing of financial services to vulnerable customers*, in *Australasian Marketing Journal*, 29, 3, 2021, 235.

[2] See, among many studies, N. Bol, J. Strycharz, N. Helberger, B. van de Velde, C. H de Vreese, *Vulnerability in a tracked society: Combining tracking and survey data to understand who gets targeted with what content*, in *New Media & Society*, 22, 11, 2020, 1996.

[3] F. Casarosa, *Access to (Digital) Justice: Is There a Place for Vulnerable People in Online Dispute Resolution Mechanisms?*, in *Journal of European Consumer and Market Law*, 13, 3, 2024, 126.

[4] P. Kerrigan, M. Barry, *Automating vulnerability: Algorithms, artificial intelligence and machine learning for gender and sexual minorities*, in P. Aggleton, R. Cover, C.H. Logie, C.E. Newman, R. Parker (eds.), *Routledge Handbook of Sexuality, Gender, Health and Rights,* London, 2023, 164; M. Gilman, *POVERTY LAWGORITHMS A Poverty Lawyer's Guide to Fighting Automated Decision-Making Harms on Low-Income Communities*, Data & Society Research Institute, 2020, in https://datasociety.net/wp-content/uploads/2020/09/Poverty-Lawgorithms-20200915.pdf (last accessed: 29/11/2024); C. Wang, S.C. Boerman, A.C. Kroon, J. Möller, C. de Vreese, *The artificial intelligence divide: Who is the most vulnerable?*, in *New Media & Society*, 24 February 2024, 1-23.

[5] See the discussion around digital vulnerability in private and consumer law, e.g., N. Helberger, M. Sax, J. Strycharz, H.W. Micklitz, *Choice architectures in the digital economy: Towards a new understanding of digital vulnerability*, in *Journal of Consumer Policy*, 44, 4, 2021, 175; M. Grochowski, *Does European contract law need a new concept of vulnerability?* in *Journal of European Consumer and Market Law*, 10, 44, 2021, 133; F. Galli, *Algorithmic Marketing and EU Law on Unfair Commercial Practices*, Berlin/Heidelberg, 2022, 181-207. An equivalent debate has taken shape in the constitutional/administrative law sphere: S. Ranchordas, *Empathy in the Digital Administrative State,* in *Duke Law Journal*, 71, 6, 2021, 1341; S. Ranchordas, *The Invisible Citizen in the Digital State: Administrative Law Meets Digital Constitutionalism*, in C. van Oirsouw, J. de Poorter, I. Leijten, G. van der Schyff, M. Stremler, M. de Visser (eds.), *European Yearbook of Constitutional Law* (forthcoming, 2024).

[6] More or less extensive references to vulnerability are contained in the Digital Services Act, the Digital Markets Act, the Data Act, the Regulation on Political Advertising, and the Cyber Resilience Act. For a comparative review, see M. Sax, N. Helberger, *Digital Vulnerability and Manipulation in the Emerging Digital Framework*, in *Digital Fairness for Consumers*, A report commissioned by BEUC, The European Consumer Organisation, 2024, 11.

[7] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No

Downloaded from www.biodiritto.org.
ISSN 2284-4503

"Vulnerability" is mentioned 19 times, 12 of which are in the Recitals and 7 in the binding part of the text, in several thematic areas. For example, AI systems exploiting some individual's vulnerabilities are classified as a prohibited practice (Article 5(1)(b)). Vulnerability is also a parameter for the European Commission to update the list of high-risk AI systems (Article 7(h)). The extent to which the high-risk AI system impacts minors and other vulnerable groups is one of the steps under the risk-management system (Article 9(9)). Within the context of regulatory sandboxes in the AIA, individuals in a condition of vulnerability due to their age or disability must be appropriately protected (Article 60(4)(g)). When dealing with AI systems presenting risk, market surveillance authorities must pay particular attention to the risks that AI systems present to vulnerable groups (Article 79(2)). The AI Office and Member States should facilitate the drawing up of codes of conduct, inter alia, on assessing and preventing the negative impact of AI systems on vulnerable persons or groups (Article 95). However, the practical implementation of these provisions remains unclear, particularly regarding how the AI Office will fulfil this role and coordinate with other bodies established under the AIA[8].

Despite these many occurrences, the AIA does not provide a unified definition of "vulnerability," thus leaving the term open to interpretation in each instance it is adopted. One may even doubt whether all occurrences of the term refer to the same concept.

In this paper, we review the different meanings of vulnerability contained in the AIA. We show that the AIA follows a rather established tradition of looking at vulnerability as a trait or a state of certain individuals and groups. It also includes a promising notion of vulnerability as a relation, but it does not clarify if and how AI changes this relation. Then, we spot the missing piece of the AIA, namely an idea of vulnerability as a characteristic of all AI-human relations, which manifests depending on different design features and interaction modes. To address this gap, we argue how such a view of vulnerability may be incorporated into the current text of the AIA by interpreting the concept of "specific social situation" contained in Article 5 (b).

2. The Underlying Meaning: Vulnerability as a Key Factor in the AIA's Objectives and Risk Analysis

The absence of an explicit and unified definition of vulnerability in the AIA does not preclude inferring it from the text, where the term is repeatedly used with different referents.

The AIA offers a nuanced account of human vulnerability in interactions with AI systems, as highlighted by the combined normative meaning of Article 5 and various Recitals, notably 5 and 48. These sections emphasise the power, knowledge, and agency imbalances between individuals and AI technology providers. Consequently, the AIA aims to protect individuals who depend on AI systems to fulfil a purpose or exercise a right, acknowledging their potential vulnerability. The AIA's normative references to

168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), OJ 12.7.2024.

[8] C. Novelli, P. Hacker, J. Morley, J. Trondal, L. Floridi, A *Robust Governance for the AIA: AI Office, AI Board, Scientific Panel, and National Authorities*, in *European Journal of Risk Regulation,* 2024, 1-25.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

vulnerability collectively present a multifaceted view that includes factors such as age, health status, financial situation, and level of social inclusion[9].

From this perspective, vulnerability has a dual dimension: it is a *general* condition, where merely possessing fundamental rights increases the risk of negative impacts from AI systems, and a *specific* condition, which depends on the right-holder individual situation (e.g., age, health, education).

As a general condition, vulnerability plays a specific role in achieving the AIA's objectives. As stated in its first recitals, the AIA aims to safeguard fundamental rights such as human dignity, democracy, equality, and the rule of law, which form the bedrock of the EU's approach to AI governance. When AI systems engage with fundamental rights, such as those in employment or law enforcement, the severity of adverse consequences may increase. These consequences can vary significantly based on the specific conditions or traits of the affected individuals or groups. Thus, any interference with fundamental rights resulting from AI deployment must be justifiable.

European legal culture and its case law are heavily influenced by the belief that resolving conflicts between fundamental rights and competing interests (or among rights) – such as those arising from the deployment of AI systems – is inherently complex and requires a balanced approach. This is because they are typically contained in legal principles, which are designed to be open-ended, explicitly value-driven, defeasible optimisation directives that can be realised in various ways and to varying extents (unlike legal rules).[10] They must coexist as far as possible. Thus, conflicts involving these principles are addressed through a proportionality procedure. This procedure facilitates the balancing and trade-offs of these rights in specific situations.

In the AIA, this trade-off procedure takes the form of a risk-based regulation[11]. AI systems are classified according to their varying risk levels. So, for instance, systems that pose unacceptable risks are prohibited because their (prospected) benefits do not outweigh the (potential) harm they may cause to fundamental rights. High-risk systems require more stringent legal safeguards before being brought to market[12].

In this risk-based regulatory architecture, vulnerability constitutes a key component of AI risk. This interpretation aligns with established risk science methodologies and prominent policy reports, such as those by the Intergovernmental Panel on Climate Change (IPCC). In these contexts, vulnerability is a critical factor in evaluating risk magnitude, as it impacts both the likelihood and severity of

[9] Among the most prominent and influential proponents of such a "universal approach to vulnerability" is Martha Fineman. According to Fineman, «human vulnerability arises in the first place from our embodiment, which carries with it the imminent or ever-present possibility of harm, injury, and misfortune». It follows that if vulnerability is embodied, «we can attempt to lessen risk or act to mitigate possible manifestations of our vulnerability» but «the possibility of harm cannot be eliminated». According to this understanding of vulnerability, vulnerable subjects are not the exception; they are the rule. See, M. FINEMAN, *The Vulnerable Subject: Anchoring Equality in the Human Condition*, in *Yale Journal of Law and Feminism*, 20, 1, 2008, 9.

[10] R. ALEXY, *On the Structure of Legal Principles*, in *Ratio Juris*, 13, 2000, 294 ss.; R. ALEXY, *Constitutional Rights, Balancing, and Rationality,* in *Ratio Juris*, 16, 2003, 131-140.

[11] In essence, this is a cost-benefit analysis inspired by the precautionary principle. Given the nature of the regulation, this proportionality procedure is merely outlined in the AIA itself, with the majority of the assessment and balancing to be carried out during the implementation and enforcement phases, primarily by the courts.

[12] Many of these high-risk systems are enumerated in Annex III of the Regulation, reflecting AI applications that align with core European values.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

consequences of a risk event[13]. By factoring the susceptibility of individuals, communities, or regions to adverse effects from hazard sources, alongside other risk components like exposure and response mechanisms, policymakers can develop a more accurate understanding of specific risk scenarios and tailor regulations accordingly. Essentially, the vulnerability in the AIA normative philosophy and architecture is an AI system's risk amplifier.

To illustrate this briefly, consider the AIA's attention to physical and mental disabilities. So, for instance, AI systems used in healthcare may not be designed to accommodate individuals with disabilities, limiting their access and potentially perpetuating bias against those with pre-existing conditions. This is even clearer in cases of malicious intent, such as AI systems designed to exploit emotional triggers and manipulate users into sharing personal information. Individual vulnerability in these cases – in its dual dimension – contributes to signalling the risk level of an AI system and triggers higher standards and increased responsibility.

3. Vulnerability as a Trait or a Situation of Persons/Groups that Can be Exploited

One explicit reference to vulnerability is contained in Article 5, which prohibits certain AI practices. Among the latter, Article 5, lit. b, prohibits «the placing on the market, the putting into service or the use of an AI system that exploits any of the vulnerabilities of a natural person or a specific group of persons due to their age, disability or a specific social or economic situation, with the objective, or the effect, of materially distorting the behaviour of that person or a person belonging to that group in a manner that causes or is reasonably likely to cause that person or another person significant harm».

The prohibition refers to "any kind of vulnerability", but in reality, it scopes out only some sources[14] of vulnerability: 1) age, 2) disability, and 3) a specific social or economic situation. The list of sources seems to be exhaustive. Recital 29 only clarifies that "disability" must be interpreted in line with the notion of "people with disability" contained in Directive 2019/882[15].

This meaning of vulnerability shares many similarities, both in the conceptual framework and in the literal wording[16], with the Directive 2005/29/CE on unfair commercial practices[17]. The Directive

[13] N.P. Simpson, K.J. Mach, A. Constable, J. Hess, R. Hogarth, M. Howden, J. Lawrence, R.J. Lempert, V. Muccione, B. Mackey, M.G. New, *A framework for complex climate change risk assessment*, in *One Earth*, 4, 4, 2021, 489; C. Novelli, F. Casolari., A. Rotolo, M. Taddeo, L. Floridi, *AI Risk Assessment: A Scenario-Based, Proportional Methodology for the AIA*, in *Digital Society*, 3, 13, 2024, 1-29; A.W. Coburn, R.J.S. Spence, A. Pomonis, *Vulnerability and Risk Assessment, Disaster management training programme*, Cambridge, 1994.

[14] We shall also refer to them as "vulnerability drivers".

[15] Recital 29 clarifies that the concept of «disability» must be interpreted in line with the notion of «people with disability» contained in Directive (EU) 2019/882 of the European Parliament and of the Council of 17 April 2019 on the accessibility requirements for products and services, that is, «people who have long-term physical, mental, intellectual or sensory impairments which in interaction with various barriers may hinder their full and effective participation in society on an equal basis with others».

[16] See the in-depth analysis by C. Goanta, *Regulatory Siblings: The Unfair Commercial Practices Directive Roots of the AI Act*, in I. Graef, B. van der Sloot (eds.), *The Legal Consistency of Technology Regulation in Europe*, London, 2024, 71.

[17] Directive 2005/29/EC of the European Parliament and of the Council of 11 May 2005 concerning unfair business-to-consumer commercial practices in the internal market and amending Council Directive 84/450/EEC,

Downloaded from www.biodiritto.org.
ISSN 2284-4503

recognises that factors like age, mental capacity, and credulity can make certain consumers more susceptible to unfair commercial practices. Specifically, it prohibits practices that exploit these vulnerabilities in a way that traders should reasonably anticipate[18].

It seems, however, that the AIA made some important steps forward, which were probably informed by the contemporary debate on vulnerability.

First, the understanding of vulnerability has transformed. It is no longer seen solely as an inherent trait of specific individuals or groups. Instead, it is now recognised as a situational and context-dependent condition that can potentially affect all human beings. This aligns with vulnerability theory, as articulated by scholars like Florencia Luna[19], which emphasises that inherent human vulnerability, stemming from our physical and social nature, is amplified by situational and structural factors. According to Luna, multiple and different layers of vulnerability may overlap. Some of them may be related to problems of knowledge, others to possible violations of human rights, to temporary situations that individuals find themselves in, or to the characteristics of the person involved.

Secondly, among the contextual drivers, the AIA considers both cognitive impairment due to external pressure[20] and socio-economic factors. This move reflects an upgrade in the awareness that vulnerability can arise from broader social and economic contexts, not merely from individual characteristics. Scholars such as Jonathan Herring argue that socio-economic conditions significantly impact an individual's susceptibility to harm, advocating for broader protections[21]. the AIA acknowledges that vulnerability often stems from systemic inequalities and external pressures beyond individual characteristics.

However, Article 5 remains quite generic on the concrete states of vulnerability, i.e., what specific traits or situations categorise individuals and groups as vulnerable in relation to each source. Regarding "specific social or economic situation", Recital 29 only provides two examples, namely "persons living in extreme poverty" and "ethnic or religious minorities". It remains unclear what other types (if any) of a "specific social situation" may result in a vulnerability state, especially whether they include not only enduring situations but also transient states (e.g., temporary unemployment, recent migration, or short-term financial crises). Moreover, no consideration is given on how the vulnerability traits and situations potentially amplify or conversely alleviate in combination with each other. Certain individuals may possess a combination of personal, social and economic vulnerabilities that makes them more susceptible to exploitation (e.g., children living in poverty), while others with similar conditions may

Directives 97/7/EC, 98/27/EC and 2002/65/EC of the European Parliament and of the Council and Regulation (EC) No 2006/2004 of the European Parliament and of the Council.

[18] A similar reference is contained in Directive 2011/83/EU of the European Parliament and of the Council of 25 October 2011 on consumer rights, amending Council Directive 93/13/EEC and Directive 1999/44/EC of the European Parliament and of the Council and repealing Council Directive 85/577/EEC and Directive 97/7/EC of the European Parliament and of the Council.

[19] F. LUNA, *Elucidating the concept of vulnerability: layers not labels*, in *Int J Fem Approaches Bioethics*, 2, 1, 2009, 121, where is argued that the concept of vulnerability should be understood as a complex and multi-layered phenomenon rather than a simplistic label, advocating for a nuanced approach that takes into account the varying degrees and contexts of vulnerability in bioethical discussions.

[20] Article 5, lit. a) AIA.

[21] J. HERRING, *Vulnerability Adults and the Law*, Oxford, 2016.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

have support systems that mitigate these risks (e.g., well-educated children from high-income families).

While the AIA highlights the importance of protecting vulnerable groups, it lacks clear guidelines for defining and identifying such groups. In this context, a group can be broadly defined as a collection of individuals who share certain characteristics or experiences that render them susceptible to exploitation or harm. Relevant characteristics may include demographic factors such as age, disability, and economic status, as well as social conditions like ethnicity, religion, or even transient circumstances such as recent migration or temporary financial crises. Identifying vulnerable groups requires a subtle understanding of how different vulnerabilities intersect and amplify each other. For instance, children living in poverty or elderly individuals with cognitive decline may represent groups with compounded vulnerabilities. However, the AIA falls short in providing specific mechanisms or criteria for recognising such groups or assessing the varying degrees of vulnerability within and across these groups.

Finally, Article 5 does not explain the exact role of AI in exploiting vulnerability. In particular, it needs to be clarified whether exploitation should be understood as an information-based process or whether it suffices for exploitation to manifest that harm to vulnerable individuals and groups occurs. In other words, does Article 5 require that the AI system possesses – either because it is provided with such knowledge or because it was learned by interacting with individuals or groups – information about a vulnerable state and uses it to make a recommendation, decision, etc.?[22] Or, is it enough that the exploitation occurs as a result of the AI system's actions, even if the system does not recognise or process the vulnerability "intentionally"?

For example, consider an AI-driven advertising platform that targets ads for payday loans to users based on their online behaviour and financial data. If the system identifies a user struggling financially and then bombards them with high-interest loan ads, this constitutes information-based exploitation. The AI system is leveraging the user's financial vulnerability to the advantage of the loan company, which profits from the user's desperation and lack of alternatives.

On the other hand, imagine an AI system designed to recommend healthcare services. This system might inadvertently harm financially vulnerable users by recommending expensive treatments without considering their economic constraints. This could lead these individuals to incur debt or forgo necessary care due to cost. Here, the AI system did not specifically leverage the information on economic vulnerability, but the harm still manifests due to the system's actions.

4. Vulnerability as a Feature of (High-Risk) AI Systems

AI systems, like humans, have vulnerabilities that can be exploited. This often-overlooked aspect of AI vulnerability is crucial because these weaknesses can interact with and worsen existing human vulnerabilities.

[22] We are not in any way referring to "mental processes" that imply an intentional state taking place in an AI system.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

There are emerging parallels between human and AI vulnerability[23], and this can also be seen in the AIA. As anticipated, Article 5 portrays human vulnerability as involving susceptibility to harm due to endogenous or exogenous factors, with exploitation involving using these weaknesses to the detriment of the vulnerable party. Similarly, AI systems can be exploited to produce harmful outcomes, such as adversarial attacks (external) or the exploitation of biases (internal). Thus, it is not unlikely that the two concepts may influence each other in the implementation of the AIA. The concept of AI/ICT vulnerability is well-established in cybersecurity[24], where systems are continually assessed for weaknesses that could be exploited by malicious actors. We foresee the possibility that computer scientists look at more human-related accounts of vulnerability in the same vein as technical vulnerability[25].
Moreover, exploiting AI systems can directly impact human well-being, creating a cascade effect. For instance, manipulating an AI used in healthcare can lead to misdiagnoses and harm patients. This is why one of the essential requirements is a vulnerability assessment and mitigation of the systems both for high-risk systems (Article 15(5) and Recital 76) and systemic-risk General-Purpose AI Models (Article 55 and Recital 110). On the other hand, human vulnerabilities can amplify AI vulnerability when vulnerable individuals unknowingly provide data, which the AI system then learns from and perpetuates, leading to even more pronounced biases and errors. This interconnectedness may create feedback loops where human vulnerabilities influence AI outcomes, and flawed AI systems exacerbate human vulnerabilities.
However, it is not clear why AI vulnerability issues and related cybersecurity countermeasures should concern only systems/models classified as having high-risk/systemic risk. All AI systems, regardless of their risk classification, have the potential to harbour vulnerabilities that can be exploited, leading to significant consequences. For example, even low-risk applications (e.g., deep fakes or emotion categorisation systems) can become entry points for broader cyberattacks or can perpetuate subtle biases that have far-reaching implications. Focusing solely on high-risk AI systems may neglect the wider landscape of AI vulnerabilities even in benign or less risky applications.

5. Vulnerability as a Trait (and a Situation?) of Affected Persons That Can Be Impacted

A third meaning of vulnerability can be located in the provider's risk-management system obligation in Article 9, applicable to high-risk systems. Accordingly, the provider of a high-risk AI system must identify, assess and mitigate risks to health, safety and fundamental rights to individuals, including giving due consideration «to whether in view of its intended purpose, the high-risk AI system is likely to have an adverse impact on persons under the age of 18 and, as appropriate, other vulnerable groups»[26].

[23] See, e.g., the reasoning line in the blog post by Chief Research Officer at Women in AI NPO, M. TSCHOPP, *Vulnerability of humans and machines – A paradigm shift*, June 2022, available https://www.scip.ch/en/?labs.20220602 (last visited 27/07/2024).
[24] H. YUPENG, K. WENXIN, Q. ZHENG, L. KENLI, Z. JILIANG, G. YANSONG, L. WENJIA, L. KEQIN, *Artificial Intelligence Security: Threats and Countermeasures*, in *ACM Computing Surveys*, 55, 1, 2021, 1-36.
[25] This may lead to an information-based interpretation of Article 5, lit. b) AIA.
[26] Article 9(9), AIA.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

A similar account of vulnerability is also contained in other risk-oriented provisions of the AIA: Article 60(4) about the conditions for testing high-risk AI systems in the real world outside regulatory sandboxes («…the subjects of the testing in real world conditions who are persons belonging to vulnerable groups…»); Article 79 on investigation activities by market surveillance authorities on systems presenting particular risks («...Particular attention shall be given to AI systems presenting a risk to vulnerable groups»); Article 95 on codes of conduct, which applies to systems other than high-risk AI system (…«assessing and preventing the negative impact of AI systems on vulnerable persons or groups of vulnerable persons…»).

An assessment of the likely impact on vulnerable groups was also part of the deployer's obligation of a fundamental right impact assessment (FRIA), as proposed by the European Parliament in the former Article 29. Interestingly, this version also included "marginalised groups"[27]. The reference to vulnerable and marginalised groups was discarded in the final version of the current Article 27, though it is not implausible that measuring the impact on disadvantaged groups will be in the end part of the content of the FRIA[28].

Regardless of the actors to which these different provisions refer (i.e. providers, market surveillance authorities, Member States and providers' associations, deployers), the vulnerability concept, in this third account, provides the benchmark for ex-ante assessing and mitigating the risk of high-risk AI systems.

This notion builds the underlying meaning of vulnerability presented in Section 2, i.e., vulnerability as a component of AI risk, but in addition, looks at vulnerability as a trait of certain groups, similar to Article 5. Compared to the average affected person, vulnerable groups are indeed expected to suffer greater harm in the event of damage caused by an AI system. Therefore, keeping the probability constant in the overall risk assessment, systems that can harm vulnerable groups are inherently riskier because the severity of the expected harm is greater. It is explained why providers and authorities are encouraged to pay more attention to those risks in the mitigation phase or in enforcement actions.

Unlike Article 5(b), however, vulnerability here is not exploited by the AI provider, but it is "impacted". This means that the harm occurs as a potential side-effect of the AI system's operation rather than as a direct exploitation. For example, an AI system designed for automated hiring processes may inadvertently disadvantage individuals with disabilities by not properly accounting for gaps in employment history related to medical treatment. Although the AI provider does not "exploit" the vulnerabilities of disabled applicants, the system's design and deployment may still result in adverse impacts on disabled groups. The focus of the provider, therefore, should be on preventively recognising and mitigating these unintended effects to ensure that AI systems do not result in disproportionate harm, even in the absence of intentional exploitation.

The idea of vulnerability as something whose impact can be predicted is supported by a risk-based theory of vulnerability, which emphasises the importance of understanding and addressing the specific

[27] Article 29a of the Amendment of the European Parliament to the AIA.

[28] This can happen, for example, if the AI Office guidelines to FRIA (Article 27(5)) will accommodate a broad interpretation of «categories of natural persons and groups likely to be affected by its use in the specific context» (Article 27(1), lit. c).

Downloaded from www.biodiritto.org.
ISSN 2284-4503

risks that different groups face[29]. In technology regulation, this approach means identifying how socio-technological systems might inadvertently harm vulnerable populations and implementing measures to mitigate these risks.

This view of vulnerability is also increasingly influencing legal scholars. For example, Gianclaudio Malgieri explored how GDPR provisions and data protection impact assessment can integrate vulnerability as a critical factor in assessing risks and designing protections[30]. Malgieri argues that recognising vulnerability as a condition that can be impacted by data processing activities allows for a more nuanced and effective regulatory response that goes beyond merely preventing exploitation. A similar discussion is now taking place with the AIA's FRIA[31].

A separate question pertains to whether this different meaning of vulnerability refers to the same vulnerable entities as Article 5. Article 9 refers only to minors and "as appropriate" to other "vulnerability groups"; thus, it does neither include "individuals" nor explicitly refer to "social and economic situations". We can assume that the interpretation of "as appropriate" follows the purpose of the high-risk AI system. For instance, in biometric systems, vulnerable groups may be ethnic minorities, who may be disproportionately misidentified due to biases in the training data; in educational systems, minors and disabled; in employment and worker management systems, women; in justice administration, ethnic groups or already convicted persons. Instead, Article 60 refers only to age and disability, thus reflecting Article 5 only regarding personal traits and not situations. Article 79 generally refers to "vulnerable groups" and not "persons". Article 95 refers to "vulnerable persons and groups, including people with disability".

6. Vulnerability as a Power Relation

Finally, a fourth meaning of vulnerability is contained in Article 7(2) of the AIA. Here, vulnerability features are one of the criteria (lit. h) that the European Commission can consider when amending Annex III on high-risk AI system applications. In particular, the Commission can consider «the extent to which there is an imbalance of power, or the persons who are potentially harmed or suffer an adverse impact are in a vulnerable position in relation to the deployer of an AI system, in particular due to status, authority, knowledge, economic or social circumstances, or age».

This fourth account views vulnerability as a power imbalance where the less powerful entity is more susceptible to harm. Martha Fineman's "universal vulnerability approach" offers valuable insight,

[29] P. Blaikie, T. Cannon, I. Davis, B. Wisner, *At risk: natural hazards, people's vulnerability and disasters*, London, 2004.

[30] G. Malgieri, *Vulnerability and Data Protection Law*, Oxford, 2023, where the Author explores the intersection of vulnerability and data protection, arguing for the incorporation of vulnerability as a key consideration in data protection frameworks, and proposing legal mechanisms to better protect vulnerable individuals in the digital age.

[31] G. Malgieri, C. Santos, *Assessing the (Severity of) Impacts on Fundamental Rights*, 25 June 2024, Available at SSRN, https://ssrn.com/abstract=4875937 (last visited 27/07/2024). See, also, A. Mantelero, *The Fundamental Rights Impact Assessment (FRIA) in the AIA: roots, legal obligations and key elements for a model template*, in *Computer Law & Security Review*, 54, 2024, 1-18.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

emphasizing that while vulnerability is a universal human experience, its extent varies and is shaped by social, political, and relational factors[32].

The idea of vulnerability as connected to power is also explored in socio-political literature. For example, political philosophers like Estelle Ferrarese extensively explored the dynamics of power and its relation to vulnerability. In his work, Ferrarese defines vulnerability as "an exposure to another's power to act"[33] and emphasises how power relations are embedded in social structures and institutions, affecting individuals' ability to protect themselves from harm. On similar lines, Judith Butler's concept of "precariousness" also aligns with this view, highlighting how social structures create conditions of vulnerability for certain groups while privileging others[34].

All this perspective underscores the importance of considering if and how AI systems can perpetuate or exacerbate these power imbalances.

The AIA clarifies that relations of vulnerability may derive from positional differences in terms of status, authority, knowledge, economic and social circumstance, or age. Vulnerability by virtue of age appears as in Article 5, although here, what seems to count is the asymmetry of experience rather than the intrinsic cognitive limitations of minors (and adults?). Socio-economic elements bear relevance too, but reference is made to "circumstances" and not to "specific situations", thus suggesting that transient aspects can also matter.

Finally, reference is made to the concepts of "status", "authority" and "knowledge". The three concepts are not defined. Yet, while "status" typically refers to an individual's social or professional position within a hierarchy or society and "authority" relates to the power or right to give orders and enforce obedience[35], "knowledge" pertains to the information, understanding, and skills that different individuals and organisations possess.

Arguably, examples of vulnerable relations depending on "status" and "authority" can be found in Recitals 58 and 60, which motivate the inclusion of AI systems used in essential services and benefits and in migration and border control management in the high-risk class. Here, different categories of people (namely, citizens and migrants) are deemed vulnerable to public entities (namely, public administration for social security and public authorities for border controls), which means that they can suffer negative consequences depending on the outcome of their decision.

As in the case of Article 5, the role AI plays in the vulnerability relation is not clear. Indeed, the way Article 7 is framed seems to look at the vulnerability condition as a characteristic of the relationship between the deployer and the affected person, *regardless* of the use of AI systems. The examples contained in Recitals 58 and 60 again may provide some clarification. Recital 58 highlights that AI

[32] M. FINEMAN, *The Vulnerable Subject*, *op. cit.*; see also R. GOODIN, *Protecting the Vulnerable: A Re-analysis of our Social Responsibilities*, Chicago, 1985

[33] E. FERRARESE, *Vulnerability and Critical Theory*, Leiden, 2018, 12, where the Author argues that vulnerability, as susceptibility to a harmful event, is, above all, a breach of normative expectations. She demonstrates that these expectations are not mental phenomena but are situated between subjects and must even be conceived as institutions.

[34] J. BURLET, *Precarious Life: The Powers of Mourning and Violence*, London, 2004

[35] See, e.g., H.L.A. HART, *The Concept of Law*, Oxford, 1961, 20, where, based on John Austin, Hart ties the concepts of authority and command: «To command is characteristically to exercise authority over men, not power to inflict harm, and though it may be combined with threats of harm a command is primarily an appeal not to fear but to respect for authority».

Downloaded from www.biodiritto.org.
ISSN 2284-4503

systems used by public administrations for social security services and benefits create a vulnerable relationship where citizens, due to their dependency on these services, are particularly susceptible to the decisions made by these systems. But do citizens depend on social security services only when AI is involved? Recital 60 addresses the use of AI in migration and border control management: are migrants and asylum seekers relying on public authorities to determine their right to enter or remain in a country only when AI is deployed? In our view, these cases clearly suggest that a relational vulnerability does not originate in the use of AI but in specific power relations – which is, in the end, what the same AIA concludes[36].

From a legal point of view, this notion of vulnerability is as innovative in its conceptual underpinning as it is limited in its application. As said above, relational vulnerability can (not shall) only guide the European Commission to review the list of high-risk AI systems in Annex III limited to areas already present. This means that the relational account of vulnerability does not provide any directly actionable protection to people in a vulnerable relationship.

Article 7(2), however, can play an additional role in shaping the implementation of the AIA: it may serve as a hermeneutic key for national courts and other enforcement authorities to interpret the notion of "high-risk" and the respective use cases, possibly using analogy[37]. This means that high-risk systems included in Annex III are there, also because a vulnerable relation is at play between the deployer and the potentially affected person.

7. The Missing Piece: Vulnerability Stemming from Human-Computer Interaction

Overall, among the many meanings of vulnerability, two stand out[38]: vulnerability as a feature of individuals and groups and vulnerability as a relation between organisations and persons. While the first aligns with traditional legal perspectives, the second offers a more nuanced view by considering power dynamics. In both cases, however, the AIA fails to clarify what exact contribution AI does bring into the picture.

We argue that a diverse, albeit essential, account of vulnerability is missing in the regulatory picture of the AIA: vulnerability as an inherent relation between AI systems and humans. This account shifts the focus from identifying and mitigating individual or situational vulnerabilities to evaluating how AI design and interaction paradigms impact human rights and other fundamental values. Interactional versions of vulnerability are currently discussed in the Human-Computer Interaction (HCI)

[36] Cf. also with the various references in the AIA to «power imbalance», such as in Recital 44 regarding the prohibited use of AI systems to infer emotions in the workplace and in Recital 59 on the use of AI by law enforcement authorities.

[37] This interpretation is supported by the fact that the criteria included in Article 7(2) are similar to the risk criteria used by the European Commission in the Impact Assessment accompanying the AIA to provide evidence for the list of high-risk AI systems included in Annex III. See Commission Staff Working Document Impact assessment – Annexes accompanying the proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts, SWD/2021/84 final, 40.

[38] We do not consider here AI vulnerability analysed in Paragraph 3.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

community[39], but they have been largely neglected in mainstream legal-philosophical analysis, which arguably provided the intellectual milieu for the AIA.

The HCI theory suggests that the interactive design features of an AI system are relevant in establishing vulnerability relations. In HCI, "design" is a multifaceted concept encompassing the aesthetic and functional aspects of AI systems and the cognitive, emotional, and social dimensions of user interaction[40]. It involves creating interfaces and interactions that are intuitive, accessible, and responsive to user needs while considering the broader context in which these systems operate.

To understand better this account of vulnerability, we unpack its main design components: (a) the purpose of the system, (b) context of use, (c) autonomy level, (d) interaction modes, and (e) physical appearance.

First, according to the HCI literature, the purpose for which an AI device is designed and its role is crucial in understanding its impact on user vulnerability. The "purpose" has a social meaning: it abstracts away from specific, fixed and predictable uses (as used for Annex III of the AIA) and includes different "types of social interaction" an AI system is expected to engage with[41].

For example, an AI system may have therapeutic or care purposes, such as supporting mental health, emotional well-being, or physical rehabilitation. It may engage in advisory interactions, providing recommendations and guidance in various information-related contexts. AI systems may also be designed for behavioural change, aiming to influence user habits and decisions, or – the distinction may sometimes be subtle – for interactive and engaging purposes, like entertainment and immersive experiences. Other types include assistive interactions that enhance human capabilities, collaborative interactions that facilitate teamwork and productivity, and monitoring and surveillance interactions that ensure security and health monitoring.

In any case, each of these social purposes may entail its consequences in terms of vulnerable relations. For instance, AI systems delivering therapeutic and care purposes, such as in mental health support, present specific vulnerabilities related to emotional manipulation and dependency[42]. Individuals may develop a deep emotional attachment to AI systems, potentially leading to an over-reliance on these devices for emotional support, which can result in neglecting human relationships and becoming more isolated. Additionally, the sensitive personal data shared during therapeutic sessions may be

[39] See, for instance, the 4TU Virtual Symposium on Vulnerability and Human-Computer Interaction, organised by the University of Twente on 2 December 2021. The only exception in the legal community is provided by the 2nd Conference of the Italian PRIN Project DIVE (Digital Vulnerability in European Private Law) dedicated to Human Vulnerability in Interaction with AI.

[40] See, among others, the J. PREECE, H. SHARP, Y. ROGERS, *Interaction Design: Beyond Human-Computer Interaction*, Hoboken, New Jersey, 2015. The HCI's view on design has its roots in the socio-materiality of technology and ecological psychology of James J. Gibson, Donald A. Norman, and Jeff Raskin, who emphasised the importance of affordances and user-centred design principles in understanding and improving human interactions with technology. In this regard, we refer to the foundational reading by D. NORMAN, *The Design of Everyday Things*, New York, 1988.

[41] We follow the approach in C. BURR, N. CRISTIANINI, J. LADYMAN, *An Analysis of the Interaction Between Intelligent Software Agents and Human Users*, in *Minds and Machines*, 28, 2018, 735, distinguishing types of interaction with artificial agents based on different types of goals, such as coercion, persuasion, nudging, trading.

[42] A. FISKE, P. HENNINGSEN, A. BUYX, *Your robot therapist will see you now: ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy*, in *Journal of Medical Internet Research*, 21, 5, 2019, e13216.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

vulnerable to misuse or breaches, raising significant privacy concerns. A virtual therapist providing cognitive behavioural therapy must be designed to safeguard user data and ensure the integrity of the therapeutic process to prevent harm.

Behavioural change AI systems, which aim to influence user habits and decisions, also introduce particular vulnerabilities. These systems often employ persuasive techniques[43] to motivate users towards specific behaviours, such as adopting healthier lifestyles or making environmentally friendly choices. While beneficial, there is a risk that users may feel manipulated or coerced into behaviours they are not fully comfortable with, potentially undermining their autonomy and consent. Furthermore, the continuous monitoring required for these systems to provide feedback and guidance can raise issues of data sensitivity and privacy. For example, a fitness app that tracks physical activity and provides personalised workout plans must handle user data with utmost care to prevent unauthorised access and ensure user trust[44].

Behavioural change seems to be the only "social function" addressed in the AIA. Article 5(a) deals with the manipulative potential of many AI applications and outlaws using subliminal techniques beyond a person's consciousness. The meaning of "subliminal techniques" is unclear, as it is the meaning of "awareness" concerning practices that operate beyond it. One might wonder if, considering this vague terminology, techniques such as the use of digital architectures to induce certain harmful behaviours in users (dark patterns)[45] or personalised and adaptive recommendations that lead individuals to irrational and impulsive choices (so-called hyper nudging) could be included[46]. The risk is that, although commendable in its objective, the ban on manipulation remains a statement of intent.

Secondly, AI-human vulnerable relations may depend on the context of use. AI integrated into private spaces, like homes, can create intimate relationships with users, leading to high levels of dependency and inner bonding[47]. Research shows how smart home devices that control lighting, heating, and security create a seamless and convenient living environment but also pose risks to privacy and data security.

In contrast, AI systems in public or semi-public spaces, such as schools, workplaces, or hospitals, interact with a broader user base and must confront varying levels of trust and dependency, which are

[43] B.J. Fogg, *Persuasive technology: using computers to change what we think and do*, Ubiquity, 5, 2002, 89 ss.. More recently, the edited book by P. de Vries, H. Oinas-Kukkonen, L. Siemons, N.B.D Jong, L. van Gemert-Pijnen (eds.), *Persuasive technology: Development and implementation of personalized technologies to change attitudes and behaviors,* Berlin/Heidelberg, 2017.

[44] See, for instance, E. A. Edwards, J. Lumsden, C. Rivas, L. Steed, L. A. Edwards, A. Thiyagarajan, R. Sohanpal, H. Caton, C. J. Griffiths, M. R. Munafò, S. Taylor, *Gamification for health promotion: systematic review of behaviour change techniques in smartphone apps*, in *BMJ Open*, 6, 10, 2016, e012447.

[45] For an interpretation of Article 5 AIA in light of dark patterns-types of influence, see the recent piece by M. Leiser, *Psychological Patterns and Article 5 of the AIA: AI-Powered Deceptive Design in the System Architecture and the User Interface*, in *Journal of AI Law and Regulation*, 1, 1, 2024, 5.

[46] S. Faraoni, *Persuasive Technology and computational manipulation: hypernudging out of mental self-determination*, in *Frontiers in Artificial Intelligence*, 6, 2023, 1216340.

[47] H.R. Pelikan, M. Broth, *Why that now? How humans adapt to a conventional humanoid robot in taking turns-at-talk*, in *CHI '16: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, May 7-12, 2016), 2016, 4921, in which the Author examines the interaction dynamics and the adjustments humans make in response to the robot's timing and conversational cues, thereby providing insights into the challenges and nuances of human-robot communication in social contexts.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

influenced by the institutional context. For instance, educational AI tools that assist in personalised learning can significantly impact student performance and engagement but also raise concerns about conforming practices to algorithm-driven learning paths and instil an increased feeling of loneliness and a lessened sense of belonging to a learning group[48].

Following an HCI account of vulnerability, a third determinant of vulnerable relations is the degree of autonomy granted to AI systems versus the level of human supervision.

Highly autonomous AI, such as self-driving cars, require users to place significant trust in the system's decision-making capabilities, which can be both empowering and anxiety-inducing[49]. For example, research shows that while autonomous systems can increase efficiency and convenience, they also raise concerns about accountability and the potential for errors[50]. Conversely, AI systems with substantial human oversight, like decision-support tools in medical settings, ensure a higher degree of control and reliability but may also suffer from reduced efficiency and increased cognitive load on human operators, which might also be described in terms of vulnerability[51].

The extent to which such nuances will be considered in the implementation of the AIA is not clear. As known, the definition of "artificial intelligence" in the Regulation contemplates machine-based systems with various levels of autonomy. However, the autonomy level does not seem to be directly correlated with the stringency of the regulatory measures imposed[52]. This raises questions about whether the specific challenges and vulnerabilities associated with highly autonomous systems are being adequately addressed in the regulatory framework.

Fourthly, the modes of interaction between AI systems and users—verbal, visual, physical, or a combination thereof—play a pivotal role in shaping the user experience and associated vulnerabilities. Verbal interactions, facilitated by voice assistants, can create a sense of conversational ease and familiarity but also introduce risks related to misinterpretation and the nuances of human language[53]. Visual interaction modes, such as those used in augmented reality and virtual reality, offer immersive

[48] P. Prinsloo, M. Khalil, S. Slade, *Vulnerable student digital well-being in AI-powered educational decision support systems (AI-EDSS) in higher education*, in *British Journal of Educational Technology*, 5, 2024, 2075 ss.

[49] P.A. Hancock, *Avoiding adverse autonomous agent actions*, in *Human–Computer Interaction*, 37, 3, 2022, 218. The Author offers the metaphor of "isles of autonomy", illustrating how autonomous systems may initially be supported by human operators, but over time, they are expected to become increasingly independent and integrated, reducing the need for human intervention.

[50] R.G. Dutta, X. Guo, Y. Jin, *Quantifying trust in autonomous system under uncertainties*, in *29th IEEE International System-on-Chip Conference (SOCC)*, 2016, 362.

[51] S. Daronnat, L. Azzopardi, M. Halvey, M. Dubiel, *Inferring trust from users' behaviours; agents' predictability positively affects trust, task performance and cognitive load in human-agent real-time collaboration*, in *Frontiers in Robotics and AI*, 8, 2021, 642201.

[52] The only two exemptions are provided by the possibility of the provider to self-exempt from high-risk category pursuant the presumption stated in Article 6(3), for example, when the «AI system is intended to perform a narrow procedural task» (lit. b) and by the possibility granted by Article 7(2) to the Commission to amend Annex III also considering «the extent to which the AI system acts autonomously and the possibility for a human to override a decision or recommendations that may lead to potential harm» (lit. d).

[53] H.A. Voorveld, T. Araujo, *How social cues in virtual assistants influence concerns and persuasion: the role of voice and a human name*, in *Cyberpsychology, Behavior and Social Networking*, 23, 10, 2020, 689.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

experiences that can enhance learning and entertainment but may also lead to over-reliance on virtual environments and potential disconnection from reality[54].

Finally, HCI literature has long stressed that the physical appearance of computational systems bears relevance in determining when a vulnerable human-AI relation exists[55]. The physical appearance of AI systems, whether embodied or disembodied, significantly influences the "social role", reliability, and the bond that individuals form with these systems. Embodied AI, such as humanoid robots, can evoke strong emotional responses and social bonding due to their human-like features[56]. Conversely, disembodied AI, like virtual assistants (e.g., Siri or Alexa)[57] and chatbots (e.g., ChatGPT, Gemini, or Claude), may foster a different type of interaction that relies on the perceived intelligence, responsiveness, and personalisation of the system rather than its physical presence. These systems often employ sophisticated conversational interfaces that create an illusion of understanding and empathy, leading users to engage with them as if they were interacting with a knowledgeable and reliable companion. This form of impersonation, where the AI mimics human-like conversational skills, can generate a sense of trust and emotional connection despite the absence of a physical body[58].

This idea of vulnerability as deception is limitedly expressed in the AIA. Only Article 50, containing transparency obligations for some AI systems considered as "low-risk", accepts that vulnerability may originate from the deceiving effect of human-like, anthropomorphised interactions. The provision requires that users be informed when they interact with an AI system rather than a human being to prevent deception and undue emotional attachment.

At the same time, Article 50 takes an optimistic stance on transparency, which contrasts with insights from HCI literature. This suggests that automating human likeness poses ethical and social questions

[54] D. VAN HEUGTEN-VAN DER KLOET, J. COSGRAVE, J. VAN RHEEDE, S. HICKS, *Out-of-body experience in virtual reality induces acute dissociation*, in *Psychology of Consciousness: Theory, Research, and Practice*, 5, 4, 2018, 346.

[55] L.R. CAPORAEL, *Anthropomorphism and Mechanomorphism: Two Faces of the Human Machine*, in *Computers in Human Behavior*, 2, 3, 1986, 215.

[56] For instance, in marketing studies, anthropomorphism is typically leveraged to entice an empathetic stance over clients and a heightened predisposition to buy. See, P. AGGARWAL, A.L. MCGILL, *Is that car smiling at me? Schema congruity as a basis for evaluating anthropomorphized products*, in *Journal of Consumer Research*, 34, 4, 2007, 468.

[57] We recall, for example, the first announcement by Amazon in September 2019 on the new improvements to Alexa's voice, including the new celebrity-guest-voice skill featuring Samuel L. Jackson's voice. C. GARTENBERG, *All the new features are coming to Alexa, including a new voice, frustration mode, and Samuel L. Jackson*, in *The Verge*, January 2019, https://www.theverge.com/2019/9/25/20883751/amazonalexa-voice-languages-natural-bi-lingual-frustration-support-new-features (last accessed 28/07/2024).

[58] See, among others, E. GO, S.S. SUNDAR, *Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions*, in *Computers in Human Behavior*, 97, 2019, 304. The study highlights several interesting findings, including the "compensation effect", where high anthropomorphic visual cues can make up for low message interactivity and vice versa. It also identifies an "expectancy violation" effect when identity cues are paired with interactive messaging, suggesting that revealing the chatbot's non-human identity can either meet or disrupt user expectations, depending on how it is communicated.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

that go well beyond merely informing users[59]. Transparency could ultimately be counterproductive for AI deployers, as it risks endangering user engagement and trust[60].

8. Room for Manoeuvre: Looking at HCI as a "specific social situation"

The absence of an HCI perspective on vulnerability in the AIA does not preclude the possibility of interpreting and enforcing its provisions through such a lens. In fact, an HCI outlook on vulnerability is compatible with the AIA's fundamental view of AI as a product. Integrating HCI perspectives can enrich the understanding and regulation of AI systems, ensuring that they are designed and deployed in ways that prioritise user well-being and safety. The AIA predominantly treats AI as a product, focusing on the technical specifications, risk management, and compliance measures required to ensure its safe use. In this context, the HCI perspective brings to the forefront the interactions between humans and AI systems, emphasising the importance of design features and user experiences in shaping vulnerability. Consequently, viewing AI through the lens of HCI enriches the product-oriented approach by also considering AI as a service.

In the previous paragraph, we pointed out some provisions of the Regulation that may accommodate an HRI view of vulnerability. We argue now that an opening point in the AIA that allows the re-incorporation of a more structured view of AI-human interaction vulnerability is the concept of "specific social situation" contained in Article 5.

In sociology, a "social situation" is variously referred to as the condition in which individuals interact and form relationships[61]. For example, referring to the social situation of people with a mental health condition in the 60s American society, the famous sociologist Erving Goffman described "social situations" as structured interactions where individuals manage their self-presentation and deal with the expectations of near others[62]. This is part of what Goffman famously coined as the "interaction order", an order which includes the norms that dictate how people present themselves and respond to others in various contexts.

[59] J. PORRA, M. LACITY, M.S. PARKS, *Can Computer Based Human-Likeness Endanger Humanness? – A Philosophical and Ethical Perspective on Digital Assistants Expressing Feelings They Can't Have*, in *Information Systems Frontiers*, 22, 2020, 533.

[60] For example, this may happen in business-consumer relations, where research suggests that undisclosed chatbots are as effective as proficient workers and four times more effective than inexperienced workers in engendering customer purchases and that a disclosure of chatbot identity before the machine–customer conversation reduces purchase rates by more than 79.7%. See, X. LUO, S. TONG, Z. FANG, *Machines vs. humans: The impact of artificial intelligence chatbot disclosure on customer purchases*, in *Marketing Science*, 38, 6, 2019, 937.

[61] It is beyond the scope of this discussion to reference the extensive literature on social situations, which encompasses seminal works such as Harold Garfinkel's ethnomethodological studies, Georg Simmel's analysis of social forms, and significant contributions from scholars including Herbert Blumer, Alfred Schutz, Pierre Bourdieu, and Norbert Elias, among others.

[62] E. GOFFMAN, *Asylums. Essays on the social situation of mental patients and other inmates*, New York, 1961, 144: «By the term social situation I shall refer to the full spatial environment anywhere within which an entering person becomes a member of the gathering that is (or does then become) present. Situations begin when mutual monitoring occurs and lapse when the next to last person has left».

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Perhaps one of the most analytical account of social situations was given by social psychologists Michael Argyle, Adrian Furnham, and Jean Ann Graham[63], who describe social situations as comprising several key features: 1) goals of persons, 2) rules, that is, the beliefs that regulate peoples' behaviours within the situation; 3) roles defining the expected behaviours and responsibilities; 4) repertoire of elements relevant to the goals; 5) sequence of behaviour that need to be completed in a particular order; 6) shared concepts necessary for managing tasks and achieving goals; 7) environmental setting; i.e. physical environment, including boundaries, props, and modifiers, which influences behaviour and interaction in a situation; 8) language and speech, with specific vocabulary and speech patterns that may need to be adapted based on the context; 9) difficulties and skills, i.e. some situations require specific social, perceptual, or linguistic skills, and the challenges faced in these contexts can offer insights into social interaction processes.

Following this framework, an AI-human interaction can be effectively analysed as a social situation.

The AI-human interaction sets roles between the AI system and the human user. For example, an AI might assume the role of an advisor, assistant, or companion, while a human may take on the role of a decision-maker, dependent user, or learner. These roles come with specific expectations and responsibilities, much like roles in traditional social situations, influencing how the interaction unfolds and how the user perceives the AI system's capabilities and trustworthiness.

Humans engage with AI systems to achieve specific objectives like obtaining information or completing tasks through specific interaction sequences. As seen in the previous paragraph, these goals or modes of interaction, as well as the predictability or variability of these users' behavioural sequences, can introduce potential vulnerabilities, especially in those relations where humans become overly reliant on the AI system.

Environmental settings and language and speech are also relevant in AI-human interactions. The virtual environment or interface in which the interaction occurs can affect the user experience, just as the physical setting influences traditional social situations. Language use, including vocabulary and tone, is tailored to the interaction context, whether formal, casual, or technical, and can vary widely depending on the user's expectations and the AI's design.

The reference to a "specific social situation" in Article 5 of the AIA is sufficiently flexible to accommodate a tailored assessment of AI-human interactions. When providers and deployers must comply with the prohibition, namely assess when the AI systems exploit vulnerabilities due to a specific social situation, they may collaboratively consider aspects of vulnerability and assess the level of vulnerability in AI-human relations. This extensive interpretation of Article 5, lit. b) allows the reincorporation of an HCI view of vulnerability into the AIA.

Following such an approach, providers and deployers could be required to adopt measures to mitigate vulnerability by focusing on design features and interaction paradigms. While we cannot elaborate here on the exact nature and details of such measures, they may involve the continuous monitoring and evaluation of AI systems to understand their interactions with persons and the social situations they create. Contextual analysis of the specific environments in which AI systems are deployed and user behavioural patterns should be deemed essential to appropriately tailor design and regulatory responses.

[63] M. Argyle, A. Furnham, J.A. Graham, *Social Situations*, London, 1981.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

9. Conclusion

In this paper, we reviewed the different meanings of vulnerability contained in the AIA. Our analysis reveals that the Act predominantly aligns with risk science literature, where vulnerability is seen as a factor influencing the overall magnitude of risk associated with an event. Additionally, the Act reflects an established tradition of viewing vulnerability as a trait or state of certain individuals and groups. This traditional perspective considers vulnerability as an inherent characteristic of specific demographics, such as the elderly, children, or economically disadvantaged groups, who are more susceptible to harm due to their particular conditions. The AIA also incorporates a promising notion of vulnerability as a relational concept, recognising that vulnerability can arise from the power dynamics between organisations and individuals, such as the dependency of citizens on public administrations for social security services or the precarious position of migrants in relation to border control authorities. However, the AIA falls short of clarifying the specific role AI plays in these interactions and how it may alter the dynamics of vulnerability.

We identified a critical missing meaning in the AIA: vulnerability as an intrinsic feature of all AI-human relations, which manifests depending on different design features and interaction modes. This perspective extends beyond the traditional meaning of vulnerability as merely an inherent trait or a relational dynamic and considers how the design and deployment of AI systems themselves can create or exacerbate vulnerability. Factors such as the purpose of the interaction, the context of use, the mode of interaction, the autonomy of the AI system, and the physical appearance of systems may contribute to determining the extent to which users may become vulnerable when engaging with these systems. Finally, we proposed that this different meaning of vulnerability can be integrated into the current text of the AIA by interpreting the construct of "specific social situation" in Article 5, lit. b) more broadly. By expanding this interpretation to cover the specific contexts and interaction paradigms facilitated by AI systems, the AIA can more effectively address the nuances of vulnerability in AI-human interaction. This holistic approach would not only protect traditionally vulnerable groups but also recognise and mitigate the new forms of vulnerability emerging from constituting relations with advanced AI technologies. In the future, this integration may prove essential for creating norms that ensure equitable deployment of AI systems and pay respect to the inherently weaker human conditions, especially vis-à-vis certain AI advanced technologies.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

La (*seconda*) svolta del 2024. Anche il Consiglio d'Europa decide di regolamentare l'intelligenza artificiale

Costanza Nardocci*

THE (SECOND) 2024 TURNAROUND. EVEN THE COUNCIL OF EUROPE DECIDES TO REGULATE ARTIFICIAL INTELLIGENCE

ABSTRACT: The paper offers a critical investigation of the Council of Europe Convention on Artificial intelligence, human rights, democracy and the rule of law. It, also, aims at disclosing the novelties, similarities and differences compared to the almost contemporary Regulation of the European Union (AI Act).

KEYWORDS: Artificial Intelligence; Council of Europe; Regulation, attempt of; Human Rights; Convention.

ABSTRACT: Il saggio offre un'analisi critica della Convenzione del Consiglio d'Europa in materia di intelligenza artificiale, diritti umani, democrazia e principio di legalità. Il saggio, inoltre, si propone di mettere in evidenza le novità, differenze e somiglianze rispetto al coevo approccio dell'Unione Europea con l'AI Act.

PAROLE CHIAVE: Intelligenza artificiale; Consiglio d'Europa; Regolamentazione, tentativo di; Diritti umani; Convenzione.

SOMMARIO: 1. Considerazioni introduttive: l'accelerazione europea sulla AI *Regulation* – 2. Non solo *AI Act*: il Consiglio d'Europa e il primo trattato globale in tema di intelligenza artificiale e diritti umani – 2.1. *(Segue)* struttura e contenuti – 3. Simili ma non uguali: *ratio*, contenuti e grado di protettività, tra Trattato del CoE e *AI Act* – 4. Riflessioni conclusive sull'impatto (davvero?) globale del Trattato, oltre il Consiglio d'Europa.

> «The Framework Convention on Artificial Intelligence is a first-of-its-kind, global treaty that will ensure that Artificial Intelligence upholds people's rights. It is a response to the need for an international legal standard supported by states in different continents which share the same values to harness the benefits of Artificial intelligence, while mitigating the risks. With this new treaty, we aim to ensure a responsible use of AI that respects human rights, the rule of law and democracy»[1]

* *Professoressa associata in Diritto costituzionale Università di Milano. Mail: costanza.nardocci@unimi.it.* *Contributo sottoposto a referaggio.*

[1] Così, la Segretaria Generale del Consiglio d'Europa, Marija Pejčinović, in occasione della definitiva approvazione della Convenzione in tema di intelligenza artificiale del 17 maggio 2024.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

1. Considerazioni introduttive: l'accelerazione europea sulla AI *Regulation*

Molto si è scritto sugli sforzi profusi nel 2023, e almeno formalmente dal 2021, dall'Unione Europea per raggiungere un accordo, poi confluito, nel marzo del 2024, nell'adozione del c.d. *AI Act*[2].

Meno, in parte, si è detto dei paralleli tentativi promossi dal Consiglio d'Europa che, a partire dal 2019 e tramite l'istituzione di due comitati di esperti *ad hoc* – il CAHAI[3] e il CAI[4] –, è giunto, a due mesi di distanza dalla pubblicazione dell'AI Act, ad approvare il primo trattato di diritto internazionale dei diritti umani in materia di intelligenza artificiale.

L'importanza del trattato del Consiglio d'Europa (CoE), la F*ramework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law*, va colta immediatamente dalla lista degli Stati, membri e non del CoE, che hanno partecipato alle negoziazioni[5]. E l'apertura, allo stato almeno potenziale, del trattato anche oltre i confini del continente europeo[6], suggerisce almeno tre considerazioni.

La prima è che persiste una tendenza degli Stati a preferire che siano organizzazioni internazionali a disciplinare un fenomeno di difficile contenimento su scala nazionale. Ne è dimostrazione la refrattarietà con cui gli Stati membri dell'Unione Europea si sono astenuti da approvare legislazioni nazionali[7], in attesa delle decisioni contrattate dalle istituzioni dell'Unione.

---

[2] Il testo definitivo dell'*AI Act* è consultabile al seguente link: https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf (ultima consultazione 02/12/2024).

[3] Il CAHAI - *Ad hoc Committee on Artificial Intelligence* è stato in carica dal 2019 al 2022. In particolare, si rinvia, in questa sede, al documento con cui il CAHAI ha concluso il proprio mandato, *Possible elements of a legal framework on artificial intelligence, based on the Council of Europe's standards on human rights, democracy and the rule of law*, pubblicato il 3 dicembre 2021 e consultabile al seguente link: https://rm.coe.int/cahai-2021-09rev-elements/1680a6d90d (ultima consultazione 02/12/2024). Sull'operato del CAHAI, si veda, A. MANTELERO, *Regulating AI*, in *Beyond Data. Information Technology and Law Series*, L'Aia, 161 ss.

[4] Il *Committee on Artificial Intelligence (CAI)* è entrato in carica nel 2022 con la finalità di redigere il successivo trattato del Consiglio d'Europa. Le attività del CAI non si sono ancora concluse ed è previsto un meeting dei membri del Comitato per il mese di settembre 2024, dove, in particolare, si discuterà della metodologia preposta alla valutazione di impatto dei sistemi di IA sui diritti fondamentali. La Draft Agenda è consultabile al seguente link: https://rm.coe.int/cai-2024-oj2-draft-agenda/1680b0d65f (ultima consultazione 02/12/2024).

[5] Hanno partecipato quali Stati non membri del CoE alle negoziazioni del trattato: Stai Uniti, Canada, Messico, Giappone, Israele, Ecuador, Perù, Uruguay, Argentina e, da ultimo, anche, l'Unione Europea.

[6] Su cui si veda § 3 dell'*Explanatory Report* sulla decisione del Comitato dei Ministri di ampliare i partecipanti alle negoziazioni del trattato. Dal Report, si legge, così, che: "[t]he Committee of Ministers also decided to allow for the inclusion in the negotiations of the European Union and interested non-European States sharing the values and aims of the Council of Europe – States from around the globe, namely Argentina, Australia, Canada, Costa Rica, the Holy See, Israel, Japan, Mexico, Peru, the United States of America and Uruguay, joined the process of negotiations in the CAI and participated in the elaboration of this Framework Convention as observer States".

[7] In proposito, valga precisare che alcuni Stati dell'Unione Europea hanno avviato alcune iniziative a livello domestico. L'Italia ha presentato un disegno di legge in materia di IA, che, di fatto, ripete l'impostazione del Regolamento dell'Unione Europea. Si tratta del Ddl. *Schema di disegno di legge recante disposizioni e delega al Governo in materia di intelligenza artificiale*, approvato dal Governo il 23 aprile 2024. La Spagna, nel 2023, è stata istituita una apposita Agenzia indipendente, chiamata ad assolvere funzioni di supervisione sull'impiego delle tecnologie di IA, la c.d. AESIA. La Germania ha adottato un proprio *Action Plan* in tema di IA (su cui, si veda il testo al seguente link: https://www.bmbf.de/SharedDocs/Downloads/de/2023/230823-executive-summary-ki-aktionsplan.pdf?__blob=publicationFile&v=1). Oltre i confini dell'Unione Europea può richiamarsi, in questa sede, il caso del Brasile che ha approvato, nel settembre 2023, un proprio quadro normativo in materia di IA. Il

Downloaded from www.biodiritto.org.
ISSN 2284-4503

La seconda è che simile preferenza per la dimensione sovrastatale ha spinto ordinamenti giuridici, per lungo tempo scettici nei confronti di una regolamentazione delle tecniche di intelligenza artificiale, a voltare il proprio sguardo verso chi quella direzione ha deciso, invece, di intraprendere. Il riferimento è, solo per qualche esempio, agli Stati Uniti, ma lo stesso vale per ordinamenti come quello canadese che, oltre a meritorie iniziative locali, ugualmente difetta di una disciplina federale unitaria.
Da ultimo, in terzo luogo, la centralità della *Framework Convention* va inquadrata anche in prospettiva orizzontale e non solo verticale e, cioè, nelle sue relazioni con il dato statale. Il Consiglio d'Europa è, infatti, ad oggi la prima ed unica organizzazione di diritto internazionale dei diritti umani che ha percorso la via del diritto pattizio piuttosto che appoggiarsi a strumenti di *soft law*.
Quest'ultima è stata la scelta delle Nazioni Unite, sì intervenute nel marzo del 2024 con la propria prima risoluzione in tema di intelligenza artificiale, non hanno però seguito il tracciato del Consiglio d'Europa. Inoltre, sempre allo stato attuale, non si danno iniziative analoghe da parte di sistemi di diritto internazionale regionale, primo tra tutti, di quello Inter-americano.
Il saggio si pone due obiettivi: descrivere e analizzare i contenuti del trattato; verificarne il grado di protettività, maggiore o minore, rispetto al coevo *AI Act*, nonché le eventuali possibilità ed opportunità di integrazioni reciproche rispetto all'esigenza, perseguita da entrambi i testi normativi, di assicurare un impiego delle tecnologie di intelligenza artificiale conforme ai diritti fondamentali.

2. Non solo *AI Act*: il Consiglio d'Europa e il primo trattato globale in tema di intelligenza artificiale e diritti umani

Il 17 maggio 2024, il Consiglio d'Europa ha pubblicato il testo ufficiale del proprio trattato in tema di intelligenza artificiale, la già richiamata *Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law*[8].
La Convenzione costituisce l'esito di un percorso durato più di 5 anni, che ha visto la istituzione di due comitati di esperti *ad hoc* incaricati di: redigere, il primo (CAHAI), uno studio sulla fattibilità di un trattato che affrontasse compiutamente le problematiche poste dai sistemi di intelligenza artificiale in

Regno Unito, nel febbraio 2024, ha istituito l'Ufficio per l'Intelligenza Artificiale presso il Dipartimento per la Scienza, l'Innovazione e la Tecnologia ed ha adottato una propria strategia nazionale in materia dapprima nel 2021, per poi aggiornarla nel 2022. Il testo può essere letto al seguente link: https://assets.publishing.service.gov.uk/media/614db4d1e90e077a2cbdf3c4/National_AI_Strategy_-_PDF_version.pdf (ultima consultazione 02/12/2024). Gli Stati Uniti, oltre alcune iniziative a livello statale particolarmente degne di nota (ci si riferisce alla recente normativa approvata dallo Stato del Colorado, CAIA, nel 2024), a livello federale ci si è arrestati al c.d. *Blueprint for an AI Bill of Rights*, consultabile al seguente link: https://www.whitehouse.gov/ostp/ai-bill-of-rights/ (ultima consultazione 02/12/2024).

[8] Il testo del trattato può essere letto al seguente link: https://rm.coe.int/1680afae3c. A commento del testo, si vedano i contributi di A. HARS, *Conceptual Difficulties in the Transformation of Human Rights to the Realm of Artificial Intelligence*, in *Acta Humana*, 2, 2024, 123 ss.; F.P. LEVANTINO, F. PAOLUCCI, *Advancing The Protection Of Fundamental Rights Through AI Regulation: How The Eu And The Council Of Europe Are Shaping The Future*, in P. CZECH, L. HESCHL, K. LUKAS, M. NOWAK, G. OBERLEITNER (a cura di), *European Yearbook on Human Rights 2024*, 2025, Leiden, in corso di pubblicazione e consultabile al seguente link: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4881656 (ultima consultazione 02/12/2024).

Downloaded from www.biodiritto.org.
ISSN 2284-4503

relazione ai valori del Consiglio d'Europa[9]; il secondo (CAI) di avviare e condurre le negoziazioni tra Stati membri e *stakeholders* esterni funzionali alla redazione dei contenuti del futuro trattato[10].
Rinviando al paragrafo successivo una riflessione sui punti di contatto, ma anche, soprattutto, di divergenza tra il testo in esame e l'*AI Act*, pare opportuno mettere qualche punto fermo sui contenuti della Convenzione e sulle finalità che il Consiglio d'Europa si è proposto di soddisfare tramite la sua adozione.
Un primo aspetto da sottolineare di più ampio respiro concerne la chiara scelta di campo del Consiglio d'Europa che precede le Nazioni Unite e altre organizzazioni regionali, nel segno della regolamentazione, anche se di principio, dell'impiego dei sistemi di intelligenza artificiale nello spazio europeo.
Ancora, altrettanto chiara è la *ratio* che sottende al trattato e, cioè, la definizione precisa del quadro di principi o criteri guida, che dovranno orientare i programmatori e gli utilizzatori dei sistemi di intelligenza artificiale in tutte le fasi del loro sviluppo[11].
Se non sono sottovalutate le potenzialità dei sistemi di intelligenza artificiale, è però altrettanto chiaro già nel Preambolo, riletto alla luce del Rapporto Esplicativo, che il Consiglio d'Europa riconosce alcune priorità sul piano dell'impatto delle tecnologie di intelligenza artificiale su alcuni principi fondamentali, identificati con estrema puntualità sin dalle prime righe del testo convenzionale.

2.1. (Segue) struttura e contenuti

Venendo ai contenuti del testo, è l'ampiezza di scopo che costituisce un primo tratto caratterizzante della Convenzione.
L'Articolo 1 si inserisce in questa prospettiva, precisando che le norme del testo convenzionale mirano ad assicurare il rispetto dei diritti umani, del principio democratico e di legalità da parte di tutti i sistemi di intelligenza artificiale in ogni fase del loro c.d. "*lifecycle*".
Questa apertura all'applicazione del testo della Convenzione a tutte le tecnologie di intelligenza artificiale, senza distinzioni, viene poi meglio precisata al paragrafo seguente che risponde alle esigenze, mutevoli e assai dinamiche, dello sviluppo della tecnologia in esame.
Ne discende che, nelle intenzioni degli estensori, le norme convenzionali andranno ad agire secondo un'intensità variabile in dipendenza dalle specificità del sistema e del suo impatto, più o meno severo, sui diritti fondamentali salvaguardati, anzitutto ma non solo, a livello del Consiglio d'Europa.

[9] In particolare, si veda CAHAI, *Possible elements of a legal framework on artificial intelligence, based on Council of Europe's standards on human rights, democracy and the rule of law*, adottato nel Dicembre 2021.
[10] Alle negoziazioni hanno, infatti, partecipato anche 68 NGOs insieme ad altre organizzazioni internazionali tra cui: l'Organisation for Security and Co-operation in Europe (OSCE), l'Organisation for Economic Co-operation and Development (OECD), la United Nations Educational, Scientific and Cultural Organisation (UNESCO) e altri organismi operanti in seno al Consiglio d'Europa. Così, precisa l'*Explanatory Report*, § 4.
[11] Chiaro l'*Explanatory Report*, che precisa che: "[The] Framework Convention focuses on the protection and furtherance of human rights 1, democracy and the rule of law, and does not expressly regulate the economic and market aspects of artificial intelligence systems. Taken as a whole, it provides a common legal framework at the global level in order to apply the existing international and domestic legal obligations that are applicable to each Party in the sphere of human rights, democracy and the rule of law of each Party and aims to ensure that the activities within the lifecycle of artificial intelligence systems by both public and private actors comply with these obligations, standards and commitments".

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Espressione di simile impostazione, di ampio respiro, è anche la definizione di intelligenza artificiale recepita dal testo del trattato e codificata ai sensi dell'art. 2. Si legge, così, che, per intelligenza artificiale, la Convenzione Quadro copre ogni: «*machine-based system that for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations or decisions that may influence physical or virtual environments*»[12].

È il Rapporto Esplicativo che chiarisce le ragioni della scelta: collega la nozione di intelligenza artificiale recepita dal testo convenzionale a quella dell'OECD[13]; e ne circoscrive, poi, i confini, sì da escluderne la riferibilità a tecnologie il cui funzionamento dipenda solo da regole stabilite dal programmatore e che si limitino all'assolvimento di mansioni esecutive. Una nozione, però, che il testo rende volutamente astratta e flessibile, allo scopo di assicurarne la c.d. neutralità tecnologica e, quindi, la ultrattività del trattato[14].

Chiude, infine, il Capo I la norma dedicata all'ambito applicativo del trattato, l'art. 3.

L'aspetto più rilevante va colto in quanto stabilito ai sensi della lettera *b*) del paragrafo 1, dove la Convenzione richiede agli Stati contraenti la stesura di una dichiarazione ("*declaration*") che dettagli le modalità ("*how*") tramite le quali lo Stato intende implementare le obbligazioni del trattato[15].

Accanto alle obbligazioni generali di cui al *Chapter II*, tra gli aspetti più significativi del trattato vi è, però, senza dubbio la scelta di delineare con cura i principi, i relativi corollari e i criteri guida a cui sono chiamati a sottostare gli Stati contraenti in tutte le fasi della vita dei sistemi di intelligenza artificiale e, cioè, la progettazione, l'implementazione, e, da ultimo, l'impiego dei sistemi di intelligenza artificiale da parte dell'utilizzatore finale. Il *Chapter III* del testo convenzionale affianca, cioè, a diritti tradizionali del diritto pubblico, costituzionale e sovranazionale, principi "nuovi" oppure, se così si preferisse definirli, criteri guida che sono venuti affermandosi parallelamente alla discesa delle tecnologie di intelligenza artificiale nello spazio "umano", pubblico e privato.

Gli articoli da 6 a 13 compresi dettagliano, quindi, un insieme di principi che, nelle intenzioni degli estensori del trattato, dovranno governare le multiformi relazioni tra persona umana e intelligenza artificiale. Tra questi, meritano di essere sinteticamente richiamati: la dignità umana, a cui la Convenzione affianca, non così sorprendentemente, il diritto alla c.d. "autonomia individuale"[16]; il principio di

[12] Così l'art. 2.

[13] Sulla definizione di IA, è interessante richiamare l'*Explanatory Memorandum on the updated OECD definition of an AI system*, pubblicato nel marzo del 2024, consultabile al seguente link: https://www.oecd-ilibrary.org/docserver/623da898-en.pdf?expires=1722520019&id=id&accname=guest&checksum=71863626A100B8ADF8E04D28B7EBEC7C (ultima consultazione 02/12/2024). In particolare, si richiama un passaggio del testo che investe l'applicazione della definizione proposta di IA, di cui viene enfatizzata la voluta ampiezza. Si legge, così, che: «[t]he updated definition of AI is inclusive and encompasses systems ranging from simple to complex. AI represents a set of technologies and techniques applicable to many different situations. Specific techniques, such as machine learning, may raise particular considerations for policymakers, such as bias, transparency, and explainability, and some contexts of use (e.g., decisions about public benefits) may raise more significant concerns than others. For that reason, when applied in practice, additional criteria may be needed to narrow or otherwise tailor the definition when used in a specific context».

[14] Si veda, per questo, aspetto, il par. 24 dell'*Explanatory Report.*

[15] Sempre l'art. 3 esclude, che le obbligazioni scaturenti dal trattato possano in qualsiasi modo interferire in materia di sicurezza nazionale.

[16] Cfr. art. 7.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

trasparenza e di controllo umano sul funzionamento delle tecnologie di intelligenza artificiale (il c.d. *oversight*)[17]; il principio di responsabilità (*accountability*)[18]; i principi, classici, di eguaglianza e di non discriminazione[19], di riservatezza e di protezione dei dati[20], per poi chiudere con la c.d. *reliability*[21], cioè la robustezza e affidabilità dei sistemi e la sicurezza (*safe innovation*)[22].

Ciascuno di questi principi meriterebbe un approfondimento a sé stante. In questa sede, ci si limiterà, tuttavia, ad alcune osservazioni sui tratti più caratterizzanti e specifici di ciascuno. L'intento è, soprattutto, evidenziare gli aspetti che più direttamente investono le relazioni tra la "macchina" e l'"umano" e che la Convenzione ha tradotto enucleando la lista di principi sopra tratteggiata.

In questo quadro, primario rilievo riveste la nozione di "autonomia individuale", che il Rapporto Esplicativo definisce corollario del principio di autodeterminazione, enfatizzandone, però, al tempo stesso, i significati più specifici assegnategli nel contesto specifico dell'intelligenza artificiale. Autonomia individuale è, allora, controllo umano sull'operato della macchina, che non deve mai operare finendo con il comprimere le volontà e l'*agere* della persona.

E il Rapporto Esplicativo è ancora più esplicito. La salvaguardia dell'autonomia individuale, si legge, è infatti tanto più importante al cospetto di tecnologie dotate di elevate abilità di imitazione e di manipolazione. Una rilettura del diritto all'autodeterminazione individuale ai "tempi" dell'intelligenza artificiale, che si rivela in definitiva preziosa anche nella prospettiva dell'operato di giudici e Corti, il cui intervento si prospetta sempre più frequente nei prossimi anni.

Il Capo prosegue, poi, occupandosi dei principi di trasparenza e di controllo, cioè di *oversight* sul funzionamento dei sistemi di intelligenza artificiale in tutte le fasi che presiedono il loro impiego.

L'interpretazione autentica di questi due principi, che si ricava ancora una volta dal Rapporto Esplicativo, concorre a posizionare la c.d. *transparency* e lo *human oversight* al centro del reticolo di principi salvaguardati dal trattato. In una nota a piè di pagina del testo, gli estensori della Convenzione hanno, così, la cura di appuntare l'attenzione sullo stretto legame che deve riconoscersi alle relazioni tra i

[17] Cfr. art. 8.

[18] Cfr. art. 9.

[19] Cfr. art. 10. In letteratura, sulle relazioni tra IA e discriminazioni, si vedano F. Zuiderveen Burgesius, *Discrimination, artificial intelligence, and algorithmic decision-making*, Strasburgo, 2018; J. Kleinberg, J. Ludwig, S. Mullainathan, C.R. Sunstein, *Discrimination in the Age of Algorithms*, in *Journal of Legal Analysis*, 10, 2018, 113 ss.; J. Kleinberg, J. Ludwig, S. Mullainathan, A. Rambachan, *Algorithmic fairness*, in *AEA papers and proceedings*, 108, 2018, 22 ss.; J. Gerards, R. Xenidis, *Algorithmic discrimination in Europe: Challenges and Opportunities for EU equality law*, 3 dicembre 2020, in https://www.europeanfutures.ed.ac.uk/algorithmic-discrimination-in-europe-challenges-and-opportunities-for-eu-equality-law/ (ultima consultazione 02/12/2024); P. Zuddas, *Intelligenza artificiale e discriminazioni*, in Aa.Vv. (a cura di), *Liber amicorum per Pasquale Costanzo – Diritto Costituzionale in trasformazione Vol. I – Costituzionalismo, Reti e Intelligenza artificiale*, Genova, 2020, 457 ss.; si consenta, infine, il rinvio a C. Nardocci, *Intelligenza artificiale e discriminazioni*, in *Rivista "Gruppo di Pisa*, 3, 2021, 9-60; in relazione alle specificità delle interferenze tra IA, genere e discriminazioni, si vedano M. D'Amico, C. Nardocci, *Discriminazioni, Donne e Intelligenza Artificiale*, in G. Cerrina Feroni, C. Fontana, E.C. Raffiotta (a cura di), *AI Anthology*, Bologna, Il Mulino, 2022.

[20] Cfr. Articolo 11.

[21] Cfr. Articolo 12.

[22] Cfr. Articolo 13. In dottrina, per un approfondimento sul tema dell'affidabilità dei sistemi di IA in sede di applicazione giudiziaria, si veda S. Penasa, *Intelligenza artificiale e giustizia: il delicato equilibrio tra affidabilità tecnologica e sostenibilità costituzionale in prospettiva comparata*, in *DPCE online*, 1, 2022, 297 ss.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

principi di trasparenza e *oversight*, da un lato, e l'autonomia e autodeterminazione individuale, reinterpretate alla luce delle implicazioni derivanti dai nuovi sistemi di intelligenza artificiale, dall'altro. I contenuti della norma convenzionale sono, infatti, ulteriormente precisati dal Rapporto Esplicativo, che considera partitamente due aspetti: la spiegabilità del sistema di intelligenza artificiale (*explainability*), intesa come capacità del sistema di intelligenza artificiale in esame di fornire spiegazioni sufficienti circa le proprie modalità di funzionamento, nonché delle finalità del suo utilizzo[23]; e l'interpretabilità (*interpretability*), che richiama, viceversa, la possibilità, dall'esterno, di comprendere come la tecnologia di intelligenza artificiale compie le proprie decisioni e previsioni[24].

Agli obblighi di trasparenza e di *oversight* fanno, poi, seguito i principi di cui all'art. 9 e, cioè, la c.d. *accountability* e la responsabilità. Accountability e responsabilità, che dovrebbero assicurare la identificazione del soggetto, individuale oppure collettivo, chiamato a rispondere in ipotesi di effetti del sistema di intelligenza artificiale lesivi dei diritti umani.

Il Rapporto Esplicativo, come sempre, si sofferma sui contenuti della norma, mettendo in rilievo due aspetti[25].

Il primo inserisce i principi in esame all'interno del più ampio discorso sulla *liability* applicata alle tecnologie di intelligenza artificiale, evidenziando l'esigenza (o, forse, l'urgenza), che sussista sempre la possibilità di offrire risposte adeguate, in termini di individuazione del soggetto responsabile, in relazione a ciascuna delle fasi in cui si snoda la vita del sistema di IA considerato. Ciò comporta il recepimento di una concezione ampia e flessibile di responsabilità, che non guarda allo status del programmatore, di colui che implementa il sistema e, infine, dell'utilizzatore, diversificandone a monte e in modo astratto le responsabilità. Piuttosto, si impone agli Stati di garantire, tramite l'adozione di meccanismi *ad hoc*, che sia sempre accertata e accertabile la responsabilità del privato oppure del pubblico in costanza di una violazione dei diritti umani, della democrazia, del principio di legalità.

Il secondo punto attiene, invece, alle relazioni tra la responsabilità, da una parte, e la trasparenza e il controllo umano (*oversight*), dall'altra. Il Rapporto Esplicativo non manca, cioè, di sottolineare come i secondi siano strumentali alla prima: tanto più saranno garantite la spiegabilità e la interpretabilità del sistema di intelligenza artificiale, tanto più agevole si rivelerà, a sua volta, la comprensione della fase della vita della "macchina" a cui addebitare l'effetto lesivo dei diritti fondamentali e, con essa, il soggetto, umano, chiamato a risponderne[26].

Quest'ultimo passaggio è particolarmente significativo nell'impianto della Convenzione.

Se letto unitariamente alla norma che impone la valutazione del rischio del sistema di IA in termini di impatto sui diritti umani, i principi in esame appaiono tanto più importanti nel supportare lo sviluppo di strategie di *assesment* che guardino all'intero sviluppo della tecnologia di IA al fine di distinguerne fasi e sotto-fasi strumentali alla contestuale identificazione del soggetto o dei soggetti responsabili in relazione a ciascuna. In questo senso, muove in modo esplicito anche la *ratio* della Convenzione, che

[23] Cfr. § 60 del Rapporto Esplicativo.
[24] *Ivi*, Cfr. § 61.
[25] Cfr. § 66 ss.
[26] Cfr. § 69.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

pone in relazione l'art. 9 con le obbligazioni positive che scaturiscono dall'art. 16 in materia di valutazione del rischio delle tecnologie di IA[27].

Si inserisce, invece, nel novero dei principi tradizionali del diritto pubblico la salvaguardia dell'eguaglianza e la non discriminazione di cui riferisce l'art. 10 del trattato. Qui, il profilo meritorio attiene alla puntuale delineazione delle cause che, nel quadro delle interazioni tra macchina e persona, sono suscettibili di tradursi in trattamenti irragionevolmente diversificati tra individui e gruppi. Il Rapporto Esplicativo elenca diverse tipologie di *bias*[28], appuntando l'attenzione sulle scansioni temporali della vita del sistema di intelligenza artificiale che possono originare lesioni dei principi di eguaglianza e di non discriminazione. Aspetto di cui si dirà più diffusamente nel paragrafo che segue e che separa l'impostazione euro-unitaria dall'approccio degli estensori del testo convenzionale.

In senso analogo, è anche la disposizione di cui al paragrafo 2, che rimette agli Stati la valutazione circa l'adeguatezza o meno delle normative domestiche vigenti per contrastare la discriminazione originata dall'impiego delle tecnologie di IA.

La Convenzione prosegue, poi, occupandosi di un altro diritto classico, la riservatezza e la protezione dei dati. Enfasi è riposta sull'ovvia centralità che la *privacy* e la tutela della segretezza rivestono nello specifico contesto esaminato e il Rapporto Esplicativo – questo il tratto più interessante – incoraggia gli Stati contraenti ad assicurare quel *surplus* di garanzie attivabili a livello nazionale così come a garantire il coordinamento con la normativa vigente in materia di diritto dell'Unione Europa con un richiamo esplicito al Regolamento sulla protezione dei dati, il c.d. GDPR[29].

Questa parentesi, che aggancia la Convenzione a due dei diritti classici del diritto pubblico maggiormente esposti al rischio di essere oggetto di violazione per effetto del funzionamento dei sistemi di intelligenza artificiale, è chiusa dal riferimento ad un principio, viceversa, "nuovo" o, si potrebbe dire, "*AI-specific*".

---

[27] Cfr. § 70. In tema di valutazione del rischio dei sistemi di AI, la prima proposta degna di nota a livello sovranazionale risale al Joint Technical Committee ('JTC') dell'International Organization for Standardization (ISO) e dell'International Electro technical Commission (IEC), il cui lavoro è sfociato nella pubblicazione dei c.d. standard ISO, *Information Technology—Artificial Intelligence—Artificial Concepts and Terminology*, ISO/IEC 22989:2022(E), luglio 2022 ('ISO/IEC 22989'). A commento, si rinvia al contributo di J.M. BELLO Y VILLARINO, *Global Standard-Setting for Artificial Intelligence: Para-regulating International Law for AI?*, in E. SHIRLOW, D.R. ROTHWELL (a cura di), *The Australian Year Book of International Law Online*, Leiden, 2023, 157 ss.

[28] Cfr. § 75. In particolare, il Rapporto richiama i seguenti: "potential bias of the algorithm's developers; […] potential bias built into the model upon which the systems are built; […] potential biases inherent in the training data sets used […], or in the aggregation or evaluation of data; […] biases introduced when such systems are implemented in real world settings […]; automation or confirmation biases, whereby humans may place unjustified trust in machines and technological artefacts or situations where they select information that supports their own views, in both cases ignoring their own potentially contradictory judgment and validating algorithmic outputs without questioning them […]". Si tratta, da prospettiva differente, dell'insieme delle c.d. fasi in cui può annidarsi il rischio di una deriva discriminatoria del sistema di IA. In letteratura, su questo aspetto, si vedano, diffusamente, S. BAROCAS, A.D. SELBST, *Big data disparate impact*, in *California Law Review*, 104, 2016, 671 ss.

[29] In tema, per un approfondimento sulle relazioni tra intelligenza artificiale e GDPR, si rinvia a G. SARTOR, F. LAGIOIA, *The impact of the General Data Protection Regulation (GDPR) on artificial intelligence (study prepared for the European Parliament)*, Bruxelles, 2020, 15 ss., in https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU(2020)641530_EN.pdf (ultima consultazione 02/12/2024).

Downloaded from www.biodiritto.org.
ISSN 2284-4503

L'art. 12 tratta, cioè, della c.d. *reliability* o affidabilità delle tecnologie di intelligenza artificiale. La norma impone alle Parti contraenti l'obbligo di mettere a punto meccanismi capaci di salvaguardare i corollari del principio in esame: la robustezza, la sicurezza, l'integrità e la protezione dei dati, la sicurezza *cyber*. Ancora una volta, il tema è l'adozione di *standard* che governino il funzionamento dei sistemi di IA sì da certificarne un *agere* conforme al principio di affidabilità. Il Rapporto Esplicativo, però, va oltre la delineazione di una obbligazione generica di risultato e dice qualcosa di più. Arriva, così, a suggerire che gli Stati contraenti si dotino di enti, cioè attribuiscano a *stakeholders* esterni, il compito di verificare la efficacia delle strategie messe a punto dal programmatore, dallo sviluppatore, dall'utilizzatore finale, singolarmente oppure unitariamente considerati, allo scopo di evitare una garanzia solo fallace delle esigenze di affidabilità che, peraltro, evidentemente presentano una forte connessione con i principi di trasparenza e di *oversight*, nonché con la responsabilità di cui, rispettivamente, agli artt. 8 e 9 del trattato[30].

Chiude, infine, il Capo III, la norma dedicata alla "*safe innovation*". Si tratta – lo ribadisce il Rapporto Esplicativo[31] – del principio che, più di tutti, tocca al cuore le dinamiche esistenti tra impiego dell'intelligenza artificiale nella dimensione pubblica e privata, i diritti umani e la democrazia. Il Rapporto Esplicativo non si astiene dall' offrire indicazioni e proposte funzionali a promuovere uno sviluppo dell'innovazione tecnologica che non si traduca in violazione dei principi fondanti lo Stato di diritto. Tra le molte, vi è il richiamo alle "*regulatory sandboxes*", ma, anche, a linee guida specifiche che possano orientare tutti i soggetti coinvolti nella disciplina delle molte fasi di vita della macchina.

La Convenzione tratta, poi, dei rimedi di fronte a violazioni dei diritti fondamentali.

Se l'art. 14 chiama gli Stati ad applicare e a rileggere i meccanismi già vigenti a livello nazionale e di diritto internazionale dei diritti umani, l'art. 15 tratteggia una obbligazione specifica alla luce delle specificità di contesto sottostante all'impiego dei sistemi di intelligenza artificiale e alle loro ricadute sui diritti della persona. La norma stabilisce l'opportunità di disporre garanzie procedurali di *oversight*, operanti sia *ex ante* che *ex post*, in base alle quali vagliare il corretto e non pregiudizievole utilizzo delle tecnologie di IA. Il paragrafo secondo è particolarmente dettagliato e, così, il Rapporto Esplicativo che indugia sugli esempi di meccanismi da implementare a livello domestico in conformità del trattato[32].

Sicuramente centrale nell'impianto della Convenzione del Consiglio d'Europa non è solo la preferenza per un apparato rimediale, che integri gli strumenti già esistenti e "generali" esistenti a livello nazionale e sovranazionale con altri, ancora una volta, "*AI-specific*". Coerente con l'approccio dell'*AI Act* è, così, il Capo V dedicato alla verifica della ricorrenza e, in caso positivo, al contenimento dei rischi derivanti dal ricorso ai sistemi di intelligenza artificiale.

Alla luce dei principi delineati dal Capo III, l'art. 16, rubricato "*Risk and impact management framework*", prevede che gli Stati contraenti si dotino di meccanismi in grado di identificare, valutare,

[30] Cfr. § 84 ss.

[31] Cfr. § 90 ss.

[32] Cfr. § 104. Interessa richiamare l'attenzione con cui il Rapporto Esplicativo si preoccupa di chiarire gli obblighi degli Stati contraenti ogniqualvolta sussista un'interazione tra la persona umana e la "macchiana", precisando che: "persons interacting with an artificial intelligence system should be duly notified that they are interacting with an artificial intelligence system rather than with a human". Si tratta di una affermazione che si ricollega evidentemente al principio del c.d. *right to know*, qui declinato come diritto di chi sia soggetto al funzionamento di dette tecnologie e non di chi faccia uso delle medesime.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

prevenire e contenere i rischi scaturenti dall'impiego delle tecnologie di intelligenza artificiale valutandone l'impatto, anche potenziale, sui diritti umani, sul principio democratico e di legalità. Il paragrafo 2 suggerisce alcune strategie di intervento tra cui merita richiamare la possibilità di testare *ex ante* il sistema di IA prima della sua messa in funzione.

La norma non circoscrive la valutazione di impatto a certe tipologie di sistemi di intelligenza artificiale e rimette alla discrezionalità delle Parti contraenti l'eventuale introduzione di moratorie oppure di divieti laddove talune tecnologie si rilevino particolarmente lesivi dei diritti fondamentali[33].

Chiudono la Convenzione i Capi VI, VII e VIII, dedicati, rispettivamente, alla implementazione del trattato, ai meccanismi di cooperazione tra gli Stati contraenti sino alle clausole finali, a cui si dedicherà qualche riflessione in chiusura del presente paragrafo soprattutto con riferimento alle norme in tema di ratifica, riserve e risoluzione delle controversie relative all'applicazione del trattato.

Per quanto attiene al Capo VI, la Convenzione Quadro ha la cura di precisare che l'implementazione del trattato: non può, in alcun caso, risolversi in una violazione dei principi di eguaglianza e di non discriminazione; deve tenere conto delle esigenze dei minori e delle persone con disabilità, favorire l'acquisizione di competenze informatiche da parte di tutti e tutte, non pregiudicare *surplus* di tutela garantiti dalle legislazioni nazionali, né comprimere diritti salvaguardati a livello domestico[34].

Interessante è la previsione di cui all'art. 19, "*Public consultation*", che impone agli Stati parte di assicurare il coinvolgimento, tramite consultazioni pubbliche, di *stakeholders* esterni in ogni aspetto che attenga alle implicazioni sociali, economiche, giuridiche, etiche, ambientali dei sistemi di intelligenza artificiale.

Si tratta di norma da guardare con particolare favore per almeno due ragioni.

La prima è che la scelta del Consiglio d'Europa di incoraggiare scambi con associazioni non governative e altri enti, pubblici e privati, recepisce una prassi che è emersa nel corso delle negoziazioni dell'*AI Act*. Al di là di come il Regolamento dell'Unione Europea abbia, poi, accolto o meno i suggerimenti ricevuti, è però evidente che, dal 2021 al 2024, il testo è stato oggetto di alcuni importanti *Statement* che ne hanno sottolineato fragilità e punti critici[35]. La seconda ragione, connessa alla prima, si coglie nel ruolo delle associazioni non governative, che il trattato valorizza, mettendo a sistema l'esigenza di supportare la partecipazione "dal basso" nel processo che guarda alla implementazione delle disposizioni convenzionali.

Sui meccanismi di cooperazione, ci si limita, in questa sede, a riprendere quanto stabilisce l'art. 23, che invita gli Stati contraenti a riunirsi periodicamente in un'apposita Conferenza allo scopo, tra gli altri, di discutere della implementazione della Convenzione, della opportunità di apporvi emendamenti, di definire eventuali controversie tra le Parti, nonché di favorire lo scambio con *stakeholders* esterni.

Oltre alla necessaria cooperazione tra gli Stati, la Convenzione, seguendo in questo l'impostazione euro-unitaria, non manca di sollecitare l'istituzione di appositivi meccanismi di controllo e di *compliance* delle norme del trattato, secondo quanto stabilisce l'art. 26. Meccanismi di controllo, che –

[33] Sul carattere eterogeneo della norma rispetto all'art. 27 dell'*AI Act*, si veda, *infra*, il paragrafo successivo.

[34] Così l'art. 18.

[35] A commento della prima versione del testo, si rinvia, diffusamente, a M. VEALE, F. ZUIDERVEEN BORGESIUS, *Demystifying the Draft EU Artificial Intelligence Act—Analysing the Good, the Bad, and the Unclear Elements of the Proposed Approach*, in *Computer Law Review International*, 4, 2021, 105 ss.

Downloaded from www.biodiritto.org.

ISSN 2284-4503

precisa il paragrafo 2 – devono rispondere ai canoni di indipendenza e di imparzialità, così come lo Stato dovrà dotare tali enti di poteri adeguati a rispondere in modo efficiente alle funzioni assegnategli. La Convenzione Quadro non impone la creazione di enti *ad hoc* quale unica strategia per rispondere agli obblighi derivanti dal trattato. Tuttavia, laddove decida di avvalersi di enti di previa istituzione oppure di articolare l'apparato di supervisione in una molteplicità di organismi, lo Stato parte è chiamato a disciplinare che tali enti non operino in modo frammentato, assicurandone il coordinamento.
Da ultimo, si collocano le disposizioni finali, che contengono qualche punto meritevole di sintetica notazione.
Anzitutto, la Convenzione è stata aperta alla ratifica anche da parte di Stati non membri del Consiglio d'Europa. Lo prevede esplicitamente l'art. 31, che dispone che il Comitato dei Ministri potrà invitare Stati non membri del Consiglio d'Europa, che non abbiano preso parte alle negoziazioni, a firmare e ratificare il trattato secondo quando stabilito a norma dell'art. 20, lett. *d*) dello Statuto del Consiglio d'Europa[36].
In secondo luogo, e venendo alla sua forza precettiva, il trattato non potrà essere fatto oggetto di riserva da parte degli Stati contraenti, escludendo, così, compressioni o limitazioni dell'ambito applicativo della Convenzione. Una scelta meritoria soprattutto, perché, così facendo, il Consiglio d'Europa blinda il Capo III, sancendo l'inderogabilità per tutti gli Stati parte dei principi *ivi* contenuti.
Seguendo la tradizionale impostazione dei trattati di diritto internazionale e per concludere, anche la *Framework Convention* disciplina l'istituto della denuncia del trattato e, quindi, la possibilità per lo Stato parte di sottrarsi alle obbligazioni a cui aveva acconsentito di vincolarsi quale effetto della ratifica.

3. Simili ma non uguali: ratio, contenuti e grado di protettività tra Trattato del CoE e AI Act

Oltre l'analisi delle previsioni del trattato, interessa interrogarsi sulle relazioni intercorrenti tra la *Framework Convention* del Consiglio d'Europa e il Regolamento dell'Unione Europea, il già più volte evocato *AI Act*, secondo una prospettiva che ne evidenzi frizioni, se ve ne sono, oppure contenuti non del tutto sovrapponibili se non, addirittura, divergenze.
Si tratta di un tema di sicuro rilievo se si considerano due dati fattuali preliminari. Il primo è che, dal punto di vista temporale, le attività del Consiglio d'Europa si sono svolte quasi parallelamente al dibattito sviluppatosi in seno alle istituzioni dell'Unione Europea, cosa che avrebbe potuto propendere per una influenza biunivoca tra le due organizzazioni internazionali. Il secondo concerne, invece, la partecipazione dell'Unione Europea alle negoziazioni della *Framework Convention*. Elementi, che avrebbero potuto suggerire una maggiore comunanza di intenti sin dall'inizio tra le organizzazioni internazionali e che, ad oggi, si è tradotta nell'adozione di due testi che, forse, costituisce sì l'esito di un'influenza reciproca che, però, ha portato ad esiti più positivi per l'*AI Act* di quanto non si sia invece riscontrato in relazione al trattato del Consiglio d'Europa[37].

[36] Il riferimento è alla maggioranza qualificata dei 2/3, a cui dovrà, eventualmente, sottostare l'approvazione della possibilità che Stati non membri del Consiglio d'Europa possano firmare e ratificare la Convenzione Quadro.
[37] Non a caso, il testo definitivo del trattato è stato fatto oggetto di critiche da parte di un gruppo di associazioni non governative, che hanno reso pubbliche le proprie perplessità in una *open letter* pubblicata il 5 marzo 2024.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Sebbene le divergenze di *ratio* e di impostazione fossero inizialmente più marcate – l'Unione Europea più incline ad occuparsi di una regolamentazione sufficientemente protettiva del diritto alla riservatezza e alla protezione dei dati ed un Consiglio d'Europa più attento alle ricadute delle tecnologie di IA sui diritti individuali –, persistono alcuni elementi che separano il Consiglio d'Europa dall'Unione Europea nel rispettivo approccio alla regolamentazione dell'intelligenza artificiale.

Alcune scelte, in particolare quelle che insistono sugli aspetti che hanno occupato una posizione di primo piano nel dibattito precedente all'approvazione dell'AI Act nel marzo del 2024, hanno per lungo tempo caratterizzato l'operato del Consiglio d'Europa, sino alla pubblicazione del *Consolidated Working Draft* nel luglio del 2023.

Inizialmente divergenti rispetto alla direzione intrapresa dall'Unione Europea, alcune soluzioni proposte dal Consiglio d'Europa si sono in seguito ricomposte, in occasione della stesura definitiva del testo del trattato che si è progressivamente avvicinato all'*AI Act*. Altre, che, originariamente divergenze non erano ma che, anzi, esprimevano un'impostazione condivisa tra le due organizzazioni internazionali, sono state invece, per così dire, "assorbite", cioè rese non più esplicite dal trattato, tanto che, ad oggi, non ne è del tutto pacifica l'interpretazione in termini di omogeneità oppure di eterogeneità rispetto alle opzioni accolte dal legislatore dell'Unione.

Un esempio delle prime, cioè delle c.d. "divergenze originarie" poi risolte, è costituita dalla definizione di intelligenza artificiale. Se quella contenuta nelle versione vigente del trattato è perfettamente coincidente con la nozione, più ristretta, di cui all'*AI Act*[38], non così è stato almeno sino al 2024. La

Il testo integrale può essere consultato al seguente link: https://docs.google.com/document/d/19pwQg0r7g5Dm6_OlRvTAgBPGXaufZrNW/edit (ultima consultazione 02/12/2024).

[38] Anche in seno all'AI Act, la definizione di intelligenza artificiale è stata protagonista di vicende alterne. Si è passati da una nozione ampia di intelligenza artificiale, secondo cui un sistema di IA equivarrebbe a qualsiasi: «software that is developed with one or more of the techniques and approaches listed in [the AI Act] and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with»; definizione criticata per la sua incapacità di tracciare una distinzione netta tra semplice *softwares* e sistemi di IA veri e propri. A questa prima opzione, ne è seguita una seconda, più ristretta, suggerita dal Consiglio dell'Unione Europea, che così qualificava le tecnologie di IA: «[AI system] means a system that is designed to operate with elements of autonomy and that, based on machine and/or human-provided data and inputs, infers how to achieve a given set of objectives using machine learning and/or logic- and knowledge based approaches, and produces system-generated outputs such as content (generative AI systems), predictions, recommendations or decisions, influencing the environments with which the AI system interacts». L'approccio del Consiglio, poi più in linea con la definizione conclusiva, mirava a circoscrivere la nozione di IA ai sistemi di *machine learning*, *logic*, o *knowledge-based*. La definizione più circoscritta è stata oggetto di critiche, in modo particolare, da alcune associazioni non governative, tra cui si richiamano, in questa sede, *Algorithmwatch* e *AccessNow*, nel documento *EU policy makers: Protect people's rights, don't narrow down the scope of the AI Act!*, 23 novembre 2021. Valga ricordare che, almeno originariamente, la definizione di partenza e di riferimento per le azioni successive tanto dell'Unione Europea quanto del Consiglio d'Europa era stata proposta dall'*Independent High-Level Expert Group on Artificial Intelligence* nel 2019, che così si esprimeva: «[w]e propose to use the following updated definition of AI: Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions».

Downloaded from www.biodiritto.org.
ISSN 2284-4503

definizione del *Consolidated Working Draft* era, infatti, più ampia e rispondeva in modo più convincente al tema dei tempi dell'innovazione tecnologica, che non sono i tempi del diritto, ammettendo interpretazioni oppure aggiustamenti della nozione di intelligenza artificiale proposti dagli Stati parte proprio con la finalità di rispondere tempestivamente al dinamismo delle nuove tecnologie. Accanto a questo aspetto, se da più parti è stata per lungo tempo messa in discussione la decisione dell'Unione Europea di abbracciare una nozione ristretta di intelligenza artificiale, circoscritta pressoché soltanto ai sistemi di *machine learning*, e che si scontrava con l'impostazione viceversa più onnicomprensiva del *Consolidated Working Draft*, oggi tale frizione non esiste più con il Consiglio d'Europa, come già detto, che ha fatto un passo indietro negli ultimi mesi ripiegando sulla stessa proposta definitoria dell'Unione.

Tre le seconde – casi, cioè, di convergenza tra le due organizzazioni sfumati nelle versioni definitive dei testi –, si inserisce l'opportunità o meno di classificare le tecnologie di IA in base al criterio del rischio, chiarendo che cosa si intenda per rischio.

Se l'*AI Act*, condivisibilmente, chiarisce che cosa costituisce un "rischio" nella prospettiva della regolamentazione dell'Unione Europea[39], il Consiglio d'Europa ha intrapreso una strada parzialmente differente. Nonostante la *ratio* del trattato sia sovrapponibile a quella dell'*AI Act*, nel senso di guardare alla legittimità del ricorso ai sistemi di intelligenza artificiale soltanto laddove non costituiscano un rischio per la tenuta dei diritti individuali, la *Framework Convention* non menziona più, come viceversa faceva il testo del *Consolidated Working Draft*, la propria adesione al criterio del rischio. Il *Consolidated Working Draft*, ancora una volta secondo una impostazione preferibile perché meglio delineava i requisiti a cui subordinare l'impiego dei sistemi di IA, dedicava infatti una disposizione apposita al c.d. criterio del rischio[40]. Disposizione che, viceversa, scompare nel testo del trattato, che, peraltro, non segue nemmeno l'impostazione finale dell'*AI Act*, omettendo qualsiasi riferimento esplicito anche soltanto alla definizione della nozione di "rischio".

I due esempi sono interessanti, perché evidenziano come la contaminazione tra le due organizzazioni che, per prime, hanno abbracciato un approccio favorevole alla *regulation* dei sistemi di IA si sia poi tradotta in esiti non del tutto conformi, sebbene uno scambio tra Unione Europea e Consiglio d'Europa sia più che evidente, nel bene e nel male[41]. Nel male, se si guarda alla nozione più ristretta di intelligenza artificiale, su cui ha ripiegato anche il Consiglio d'Europa e, ancora, alla non più espressa previsione di una norma in materia di valutazione del rischio. Nel bene, se si considera invece lo spostamento dell'Unione Europea verso una impostazione più *human rights-based* rispetto alla versione originaria del testo pubblicata nell'aprile 2021.

In relazione a quest'ultimo aspetto, sembra che sia stato più il Consiglio d'Europa ad influenzare l'Unione Europea che non viceversa. Una influenza, però, che si è risolta in un più ampio spazio

[39] Così, l'art. 2.

[40] All'art. 2 si leggeva, infatti, quanto segue: «[i]n order to give full effect to the principles and obligations set out in this Convention, each Party shall maintain and take such graduated and differentiated measures in its domestic legal system as may be necessary and appropriate in view of the severity and probability of occurrence of adverse impacts on human rights and fundamental freedoms, democracy and the rule of law during design, development, use and decommissioning of artificial intelligence systems».

[41] Per una analisi delle tendenze globali sulla *governance* dell'IA, si rinvia a H. ROBERTS, E. HINE, M. TADDEO, L. FLORIDI, *Global AI governance: barriers and pathways forward*, in *International Affairs*, 3, 2024, 1275 ss.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

riservato a diritti ulteriori rispetto alla riservatezza e alla protezione dei dati – si pensi, tra tutti, al richiamo alla Carta dei diritti fondamentali dell'Unione nel Preambolo – e che, tuttavia, non è scevra di perplessità. La principale risiede nell'omesso riferimento alle Direttive in materia di diritto anti-discriminatorio dell'Unione[42], che l'*AI Act* non menziona nemmeno nel Preambolo, lasciando aperto più di un interrogativo circa la riferibilità o meno di tale *corpus* normativo anche alla *AI-derived discrimination*. Qui, viceversa, la *Framework Convention* presenta una risposta più efficace, dedicando due norme ai principi di eguaglianza e di non discriminazione che agiscono secondo una prospettiva duplice: quale principio guida o generale a cui subordinare, da un lato, la creazione e l'impiego delle tecnologie di IA e, dall'altro, come diritto a cui guardare nella fase di implementazione dei sistemi in esame.

Sotto altra prospettiva, nonostante i testi abbraccino la stessa nozione di intelligenza artificiale, una differenza importante investe l'ambito applicativo della c.d. valutazione di impatto dei sistemi di intelligenza artificiale sui diritti fondamentali, che i due testi disciplinano, rispettivamente, ai sensi degli artt. 27, l'*AI Act*, e 16, la *Framework Convention.* Il legislatore dell'Unione ha, infatti, scelto di circoscriverne l'operato ai soli sistemi c.d. "ad alto rischio"[43], laddove, all'opposto, il Consiglio d'Europa non delimita la valutazione a tipologie più o meno "pericolose", allargando così la valutazione a tutti i sistemi di IA e non introduce divieti circa l'impiego di talune tipologie di tecnologie.

La differenza, sebbene importante, va, però, inserita nel quadro della eterogenea vincolatività dei due strumenti.

In altre parole, la perplessità è che il carattere meno vincolante del trattato del Consiglio d'Europa rispetto al Regolamento dell'Unione Europea finisca con il rendere solo potenzialmente più protettivo il trattato, residuando ampio margine di apprezzamento agli Stati sulle modalità attraverso le quali darvi attuazione. L'art. 16 del trattato del Consiglio d'Europa, rischia, cioè, di rimanere una norma solo sulla carta più garantista, non aggiungendo in definitiva nulla a quanto dispone il vincolante a tutti gli effetti *AI Act*.

Da ultimo, così come la *Framework Convention* si astiene dal delimitare la valutazione di impatto ai sistemi "ad alto rischio", allo stesso modo, il trattato omette di suggerire qualsiasi classificazione dei sistemi di intelligenza artificiale come, viceversa, propone, pur con tutti i suoi limiti, il Regolamento dell'Unione Europea. Ne discende, che il trattato non introduce deroghe al ricorso a sistemi di intelligenza artificiale ritenuti con maggiore frequenza causa di violazioni di diritti fondamentali – si pensi, per tutte, alle tecnologie di riconoscimento facciale –, anche se la previsione di cui all'art. 3, paragrafo

[42] Il riferimento, in particolare, è alla c.d. *Race Directive*, Direttiva 2000/43/CE del Consiglio, del 29 giugno 2000, *che attua il principio della parità di trattamento fra le persone indipendentemente dalla razza e dall'origine etnica*, e alla Direttiva Direttiva 2000/78/CE del Consiglio, del 27 novembre 2000, *che stabilisce un quadro generale per la parità di trattamento in materia di occupazione e di condizioni di lavoro*.

[43] Quelli, cioè, elencati all'*Annex* III del Regolamento e richiamati a norma dell'art. 6, § 2, dell'*AI Act*. In particolare, quanto ai sistemi di riconoscimento facciale, si rinvia a quanto precisa il *Recital* n. 54, dove si legge quanto segue: «[a]s biometric data constitutes a special category of personal data, it is appropriate to classify as high-risk several critical-use cases of biometric systems, insofar as their use is permitted under relevant Union and national law. Technical inaccuracies of AI systems intended for the remote biometric identification of natural persons can lead to biased results and entail discriminatory effects. The risk of such biased results and discriminatory effects is particularly relevant with regard to age, ethnicity, race, sex or disabilities. Remote biometric identification systems should therefore be classified as high-risk in view of the risks that they pose».

Downloaded from www.biodiritto.org.
ISSN 2284-4503

4, che esclude dall'ambito applicativo della *Framework Convention* ogni questione inerente la difesa nazionale, potrebbe essere letta in senso conforme all'eccezione introdotta dall'*AI Act* circa il legittimo impiego di sistemi di IA "ad alto rischio" in presenza di esigenze di sicurezza pubblica.
Vero è che questa norma del trattato pare criticabile al punto da minare in radice le potenzialità del trattato, comprimendone a monte la già precaria vincolatività a livello domestico.
In proposito, il precedente *Consolidated Working Draft* proponeva una formulazione preferibile. Il testo dell'allora art. 7, preposto all'elenco dei principi di cui all'attuale Capo III, possedeva una ratio differente, chiarendo meglio la finalità di eventuali e ammissibili eccezioni. Si stabiliva, cioè, che nessuna deroga ai principi salvaguardati dal trattato avrebbe potuto essere introdotta dagli Stati contraenti se non per ragioni, tra le altre, di sicurezza nazionale, difesa, sanità.
È vero che il Rapporto Esplicativo giustifica la deroga rifacendosi all'art. 1 dello Statuto del Consiglio d'Europa, precisando che la norma non esclude l'assoggettamento dei sistemi di IA al diritto internazionale. Tuttavia, tali affermazioni paiono in ogni caso contraddire ed indebolire la portata del trattato, legittimando scelte discrezionali, eterogenee e, eventualmente, meno garantiste di quegli stessi principi su cui si regge la Convenzione Quadro, minando il consolidamento di obiettivi comuni tra gli Stati Parte.
A voler tirare le fila di quanto appena descritto, sembra che a beneficiare degli scambi tra le due organizzazioni internazionali sia stato più l'*AI Act* rispetto al trattato del Consiglio d'Europa, che la versione approvata nel maggio 2024 ci restituisce debole, sicuramente meno pregnante e innovativa di quanto le versioni precedenti avessero fatto sperare. I due testi, cioè, si sono quasi scambiati le parti. Dalla *Zero Draft*[44], alla *Revised Zero Draft*[45] sino al *Consolidated Working Draft*, il Consiglio d'Europa sembrava, infatti, occupare una posizione di primo piano, almeno quanto alla esigenze di tutela dei diritti fondamentali e, invece, sul finire ha ceduto il passo all'Unione Europea senza che, purtroppo, nessuno dei due testi si dimostri attualmente pienamente soddisfacente. Di questo, lo scarso peso della non discriminazione nell'impianto dell'*AI Act* è solo un esempio, non però poco importante.

4. Riflessioni conclusive sull'impatto (davvero?) globale del Trattato, oltre il Consiglio d'Europa

Le settimane, che hanno preceduto l'entrata in vigore del trattato, sono state caratterizzate da alcuni eventi degni di nota.
Al di là dell'approvazione dell'*AI Act*, di cui si è già detto, merita ricordare che, sebbene optando per uno strumento di *soft law*, anche le Nazioni Unite hanno scelto di intervenire "dicendo la loro".
La prima risoluzione dell'ONU in materia di intelligenza artificiale è stata, così, pubblicata l'11 marzo 2024[46], poco dopo il Regolamento dell'Unione e due mesi prima l'adozione della Convenzione Quadro.

[44] Il testo della *Zero Draft* può essere consultato al seguente link: https://www.statewatch.org/media/3697/coe-artificial-intelligence-convention-zero-draft-30-6-22.pdf (ultima consultazione 29/11/2024).
[45] Su cui si veda il testo integrale al link: https://rm.coe.int/cai-2023-01-revised-zero-draft-framework-convention-public/1680aa193f (ultima consultazione 29/11/2024).
[46] Qui il link al testo della Risoluzione: https://documents.un.org/doc/undoc/ltd/n24/065/92/pdf/n2406592.pdf?token=ArxLTgaVGOUYe5P1e5&fe=true (ultima consultazione 29/11/2024).

Downloaded from www.biodiritto.org.
ISSN 2284-4503

La Risoluzione non è, però, il solo evento da richiamare e che, in breve, testimonia la volontà di riportare i diritti al centro, legando – questa è sicuramente una novità –, per la prima volta, il tema dello sviluppo sostenibile all'intelligenza artificiale[47].

Maggiore peso riveste, però, nella prospettiva che qui interessa, la lettera aperta, sottoscritta da numerose associazioni non governative, che sollecitava il Consiglio d'Europa a ritornare sui suoi passi, emendando il testo del trattato, poi approvato, reo di essersi tramutato in un documento privo di contenuti e adeguatamente protettivo dei di diritti esposti alle ricadute anche pregiudizievoli dei sistemi di intelligenza artificiale.

Il riferimento a questa lettera aperta supporta alcune riflessioni conclusive sull'esito delle negoziazioni e sulle aspettative, tradite, rispetto all'operato iniziale del Consiglio d'Europa, almeno sino all'adozione del *Consolidate Working Draft* di cui si è detto.

Con le parole "*No to the abdication of our rights*", si apriva la lettera indirizzata dalle associazioni non governative firmatarie, nel gennaio del 2024, alle istituzioni dell'Unione Europea, nonché agli Stati membri parti attive della negoziazione affinché ostacolassero alcune delle criticità che il testo ha poi ratificato, anche sulla scorta delle pressioni, per tutti, degli Stati Uniti d'America, invitati a prendere parte al processo di definizione dei contenuti del trattato.

I due aspetti, su cui appuntavano le proprie preoccupazioni le associazioni non governative firmatarie dello *Statement*, sono condivisibili e vanno ripresi in chiusura a testimonianza della torsione del trattato e dell'indebolimento delle garanzie che le precedenti versioni del testo avevano, viceversa, tratteggiato in modo convincente.

Ci si riferisce, per prima, all'esenzione senza eccezioni dall'applicabilità del trattato e, quindi, alla completa esclusione dal suo ambito applicativo di tutte le ipotesi di impiego dei sistemi di intelligenza artificiale qualora venga impiegati per ragioni di difesa nazionale e di sicurezza. La deroga appare tanto più grave se si considera che i pochissimi casi sinora giunti con successo dinanzi alle Corti, nazionali e sovranazionali, hanno, non così casualmente, riguardato proprio l'utilizzo, talvolta massiccio, delle tecnologie di intelligenza artificiale per ragioni di sorveglianza pubblica, sfociando in condanne delle autorità di pubblica sicurezza per le conseguenze discriminatorie scaturenti da tali sistemi.

Non meno severe sono, in secondo luogo, le critiche relative all'altra pesante esenzione, che taglia fuori tutti gli enti privati, le *big tech* per prime, dal rispetto delle norme convenzionali. La scelta è evidentemente politica, ma non può certo posarsi simile retrocessione delle garanzie di salvaguardia dei diritti fondamentali di fronte ad esigenze di mercato.

Deludente, in definitiva, il trattato che ci consegna il Consiglio d'Europa.

Deludente quanto alle aspettative iniziali, e, soprattutto, se si pensa alle speranze che potesse sopperire o, almeno, contenere alcune delle debolezze del coevo *AI Act*.

Si chiudono, allora, queste riflessioni con due passaggi dello *Statement* di cui sopra, che meglio di ogni altro esprimono le criticità della Convenzione Quadro:

[47] In questo senso, muove il § 3, punto *d*), dove così si legge: «[e]nhancing the ability of developing countries, in particular the least developed countries, to address major structural impediments and lift obstacles to accessing the benefits of new and emerging technologies and artificial intelligence innovation to achieve all 17 Sustainable Development Goals, including through scaling up the use of scientific sources, affordable technology, research and development, including through strengthened partnerships».

Downloaded from www.biodiritto.org.
ISSN 2284-4503

«*[w]e have never given mandate to our elected representatives to abdicate our rights through the conclusion of a Convention that is ironically supposed to safeguard them. A hollowed-out Convention will provide little meaningful protections to individuals who are increasingly subject to powerful AI systems prone to bias, human manipulation, and the destabilisation of democratic institutions. As for public activities, some states are pushing for a blanket exemption with regard to national security and defence. Nothing justifies the unconditional waiving of the safeguards set in international, European and national law that usually apply in these fields. These attempts also turn a blind eye to the geopolitical context characterised by an increasing prevalence of dual use AI that put our lives and freedom at risk*».

Se il futuro saprà riservare di meglio, è un augurio che, in questa sede, si ritiene di potere e dovere auspicare.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

La vulnerabilità degli utenti *in rete*

Luca Di Majo*

THE VULNERABILITY OF NETWORKED USERS

ABSTRACT: The digital platforms express particularly pervasive forms of manifestation of economic and social power. The platform economy exploits users' passions and penetrates the most vulnerable recesses of people. The European Union has launched a strategy aimed at regulating the most problematic aspects of the relationship between users and platforms, which to date have been embodied in a *digital regulatory package*. The regulatory acts are concerned with protecting – with some criticality – the vulnerable categories that interface with the network.

KEYWORDS: Digital platforms; digital constitutionalism; artificial intelligence; State, vulnerability.

ABSTRACT: Le piattaforme digitali manifestano espressioni di potere economico e sociale particolarmente pervasive, sfruttando le dinamiche della c.d. *platform economy*. Lo sfruttamento delle passioni, dei bisogni e degli interessi degli utenti consente alle piattaforme di *leggerne* le aspirazioni, farle proprie, e così sfruttarne al massimo la vulnerabilità degli utenti *on line*. L'Unione europea ha avviato una strategia volta alla regolazione degli aspetti più problematici del rapporto tra utenti e piattaforme, ad oggi concretizzatasi in un vero e proprio *pacchetto digitale*. La regolazione europea si occupa di tutelare – con qualche criticità – le categorie vulnerabili di utenti che si interfacciano con la rete.

PAROLE CHIAVE: Piattaforme digitali; costituzionalismo digitale; intelligenza artificiale; Stato; vulnerabilità.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

* *Ricercatore t.d., lett. b), di Diritto costituzionale e pubblico, Università della Campania Luigi Vanvitelli. Mail: luca.dimajo@unicampania.it.* *Contributo sottoposto a doppio referaggio anonimo.*

1. Premessa

Vulnerabilità è una «parola contenitore»[1], poliforme[2], concernente situazioni individuali e collettive[3]; indica la presa di coscienza della fragilità umana come «destino dei singoli e della specie […], elemento costitutivo della condizione antropica»[4].

Il concetto di vulnerabilità evoca l'idea di una «umanità fragile»[5] e di una condizione di sottomissione[6] che si manifesta in modo dinamico, legata al mutamento del contesto in cui la persona è collocata[7], e che dà conto della «precarietà, della fragilità, dell'insicurezza, delle minacce, che caratterizzano l'epoca contemporanea e che incidono sulla vita degli individui»[8].

La prevalenza dell'aspetto tecnologico sulla vita umana, poi, accentua la natura fragile di un soggetto esposto, negli ecosistemi digitali, «alle azioni e alle scelte di qualcuno»[9]: algoritmi, meta-stati, antisovrani.

La rete digitale, come ogni forma di progresso scientifico-tecnologico, esprime forme di eterogenesi di fini, talvolta accompagnando l'uomo in un percorso di affrancamento da una condizione di minorità psico-fisica, talaltra divenendo «mezzo senza fine»[10], contraria al significato più profondo della dignità umana.

È il paradosso dell'età moderna: l'uomo scopre la tecnologia (la rete) ma la utilizza contro sé stesso, in violazione dei diritti fondamentali della persona.

Tra le opportunità e i rischi di reti e algoritmi, ormai divenuti convitati di pietra della quotidianità umana, si colloca il tema classico del rapporto tra potere e autorità[11], declinato nella relazione tra co-

[1] L. Re, *Introduzione. La vulnerabilità fra etica, politica e diritto*, in M.G. Bernardini, B. Casalini, O. Giolo, L. Re (a cura di), *Vulnerabilità: etica, politica, diritto*, Roma, 2018.

[2] K. Brown, K. Ecclestone, N. Emmel, *The many faces of Vulnerability*, in *Social Policy & Society*, 3, 2017, 497 ss. Distingue «cinque figure della vulnerabilità», B. Pastore, *Vulnerabilità situata e risposte alle vulnerazioni*, in *Etica & Politica*, 1, 2020, 283-291.

[3] B. Pastore, *Vulnerabilità, diritto, ragionamento giuridico*, in F. Mancuso e V. Giordano (a cura di), *Ombre del diritto*, in *Teorie e Storia del Diritto Privato*, num. spec., 2022, *passim*.

[4] M. Luciani, *Le persone vulnerabili e la Costituzione*, Intervento di discussione della *Lectio magistralis* del Presidente della Corte europea dei diritti dell'uomo, Prof. Roberto Spano, *Diritti e umani e persone vulnerabili*, Università degli Studi di Roma La Sapienza, Facoltà di Giurisprudenza, 22 aprile 2022.

[5] S. Rossi, *Forme della vulnerabilità e attuazione del programma costituzionale*, in *Rivista AIC*, 2, 2017, 1.

[6] Secondo E. Ferrarese, *Vulnerability and Critical Theory*, Leiden-Boston, 2018, 24, «una persona alla mercè di altri», o, «precaria», secondo J. Butler, *Precarious Life. The powers of mourning and violence*, Londra, 2004.

[7] F. Luna, *Elucidating the Concept of Vulnerability: Layers non Labels*, in *International Journal of Feminist Approachs to Bioethics*, 1, 2009, 121 ss.

[8] B. Pastore, *op. cit.*, 3.

[9] G. Goodin, *Protecting the Vulnerable*, *A reanalysis of our social responsibilities*, Chicago, 1985.

[10] P.D. Omodeo, *L'aut aut di fatticità scientista e relativismo postmoderno quale semplificazione ideologica del problema epistemologico di* expertise *e populismo post-veritativo*, in G. Ienna, F. D'Abramo, M. Badino (a cura di), Expertise *ed epistemologia politica*, Milano, 2022, 45.

[11] N. Bobbio, *Liberalismo e democrazia*, Milano, 1995.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

stituzionalismo e sovranismo digitale[12], connaturato a quello della condizione di fragilità dell'utente sottoposto alle prassi dominanti di «sovrani dalla corona di silicio»[13].

L'agire pratico degli algoritmi nell'ecosistema digitale consegna la persona al conformismo sociale, ne detta il modo di vivere, pensare, immaginare, desiderare, presuppone l'alienazione da ogni campo della vita, la perseguita negli anfratti più remoti della sua vulnerabilità.

L'interazione tra persona e tecnologia è continua: gli algoritmi contribuiscono alla costruzione di un'artificialità sociale, determinano l'immaginario della persona, ne delineano il volto e il modo con cui è percepita all'esterno, la dominano dall'alto, costruiscono il nesso identità-esperienza attraverso la captazione di bisogni, impulsi, aspirazioni, penetrando nei luoghi più reconditi delle sue passioni, segnando il passo della sua natura fragile.

Anzi sono proprio studi offerti dalla «psicografia biometrica»[14] a mettere in luce come la raccolta e l'utilizzo di dati biologici – funzionali a rilevare dettagli intimi su simpatie, antipatie, preferenze, interessi di un utente – possano condurre a metodi ancora più opachi e invadenti di profilare, categorizzare e manipolare soprattutto i gruppi più vulnerabili[15].

Si tratta, a ben vedere, di un metodo *standard* di profilazione intrusiva che crea «dipendenza da *internet*»[16]: la perdita del controllo di una persona fragile e indifesa, soggetta continuamente ad attività di *marketing* digitale aggressivo e di sorveglianza capillare, destinata ad essere catalogata – attraverso tecniche di profilazione e di *hypernudging* – in singole *personalità upload*, in individualità *datificate* e omogeinizzate, in categorie segmentate e facilmente manipolabili.

In molti casi, la personalizzazione algoritmica si basa, infatti, sulla comparabilità e sulla somiglianza (per alcune categorie semplificate) dell'utente con altri (*collaborative filtering* o *filtraggio sociale*). Così, la personalizzazione nega paradossalmente l'unicità individuale attraverso un'omogeneizzazione intelligente che annichilisce la diversità del genere umano[17] e rende impalpabili le differenze che pure esistono tra utenti più vulnerabili e utenti meno vulnerabili.

Costruendo, manipolando e rafforzando queste categorie omogeneizzanti, la c.d. profilazione di gruppo crea aggregati governabili e sorvegliabili secondo criteri conosciuti esclusivamente dai gestori delle piattaforme.

Gli individui diventano facile bersaglio di pratiche manipolative e/o discriminatorie[18].

---

[12] Ne ho parlato in L. Di Majo, F. Paruzzo, *Nuovi aspetti del potere nel Metaverso*, in M. Calamo Specchia (a cura di), *Processi politici e nuove tecnologie*, Torino, 2024, 200 ss.

[13] A. Venanzoni, *Cyber-costituzionalismo: la società digitale tra silicolonizzazione, capitalismo delle piattaforme e reazioni costituzionali*, in *Rivista italiana di informatica e diritto*, 1, 2020, 5 ss.

[14] Electronic Frontier Foundation (EFF), *Virtual Worlds, Real People: Human Rights in the Metaverse*, 9 dicembre 2021, in https://www.eff.org/deeplinks/2021/12/virtual-worlds-real-people-human-rights-metaverse (ultima consultazione 02/12/2024).

[15] Cfr. S. Rodotà, *Tecnopolitica. La democrazia e le nuove tecnologie della comunicazione*, Roma-Bari, 2004.

[16] Ivan Golberg propose di introdurre la sindrome *Internet Addiction Disorder* (IAD) nel Manuale diagnostico e statico dei disturbi mentali DSM, per la forte analogia dei segni e sintomi al gioco d'azzardo patologico.

[17] Cfr. I. de Vivo, *Towards an Algorithmic Public Opinion?*, in R. Andò (a cura di), *New Journalism(s) in Theory and Practices Learning from Digital Transformations*, Roma, 2023, 81 ss.

[18] Sul tema, si veda ampiamente M. Hildebrandt, S. Gutwirth, *Profiling the European citizen: cross-disciplinary perspectives*, New York, 2008.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

L'identità della persona negli ecosistemi digitali diventa il risultato di un legame anfibio e probabilmente irreversibile, almeno nella *metà della mela* in mano ai sovrani digitali che difendono interessi economici privati (i loro) senza essere garanti di una comunità organica che è mossa secondo la soggettività di chi detiene un potere di condizionamento economico, e che implica un'influenza anche sociale e culturale[19].

2. Dalla libertà della rete alla vulnerabilità nella rete. La condizione soggettiva degli utenti nello sconfinato spazio digitale

Lo spirito originario della rete, prima ancora dell'era algoritmica, era improntato ad una esaltazione della libertà individuale[20], materializzatosi come uno spazio di libertà senza confini[21], per gli utenti e degli utenti.

Era il periodo in cui maturava la svolta del *web 2.0*, un nuovo modo di concepire l'*agorà* digitale come contesto di partecipazione e produzione di contenuti personali.

L'avvento del *web 2.0* ha rafforzato l'accesso alla rete, ha prodotto immediatamente la nascita di una cultura partecipativa[22], ha incrementato una *peer production*, certamente, ma ha rappresentato anche la capitalizzazione di una massa collettiva scaturita dalla partecipazione degli utenti, per un verso indistinta, per altro verso connotata da individualismi, ognuno dei quali portatori di un capitale umano ben presto trasformatosi in capitale economico a favore di coloro i quali diventeranno, poi, «i sovrani dalla corona di silicio»[23].

Su questo terreno sono germogliati i semi del capitalismo digitale e di un prosumerismo[24] estremo: la produzione di beni diventa consustanziale al consumo, non potendo l'una prescindere dall'altro[25]. I due processi si compenetrano: l'azienda è chiamata a svolgere attività di impresa – diventa essa stessa attore sociale – unitamente a chi, in passato, aveva contribuito a formare il popolo della rete: l'utente-consumatore.

Anzi, era prevedibile come la cultura partecipativa finisse con l'assumere sembianze sempre più commerciali[26], tanto che la vera innovazione è consistita nell'inserimento, tra l'azione dei produttori e dei consumatori, di terzi soggetti: le piattaforme digitali.

L'ascendenza delle piattaforme sulla persona è dovuta, in gran parte, alle tecnologie *disruptive* dell'intelligenza artificiale che, rispetto al percorso di massimizzazione dei diritti, hanno svelato il volto fragile della persona, per lo meno per quanto attiene alla condizione di sottomissione al potere economico/sociale esercitato dai sovrani digitali.

[19] In tal senso, A. Gentili, *La vulnerabilità sociale. Un modello teorico per il trattamento legale*, in *Rivista critica del diritto privato*, 1, 2019, 41 ss.

[20] T. Detti, G. Lauricella, *Le origini di Internet*, Milano, 2013.

[21] M. Castells, *Comunicazione e potere*, Milano, 2009, e dello stesso A., *Galassie Internet*, Milano, 2011.

[22] H. Jenkins, R. Purushotma, M. Weigel, K. Clinton, A.J. Robison, *Culture partecipative e competenze digitali. Media education per il XXI secolo*, Milano, 2010.

[23] A. Venanzoni, *op. cit.*, 5 ss.

[24] G. Fabris, *Customer Knowledge Marketing*, in *Consumatori, Diritti e Mercato*, 1, 2008.

[25] A. Toffler, *The Third Wave*, Londra, 1980.

[26] B.E. Duffu, D.B. Nieborg, T. Poell, M. Fax, *Platforms and Cultural Production*, New Jersey, 2022.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Lo spirito originario della rete, libero e a disposizione della comunità mondiale, si è articolato in una struttura a quattro strati[27], ove ha iniziato a muoversi l'*homo oeconomicus* e dove valori e regole sono definite da «non-Stati»[28] che hanno sviluppato un modello capitalistico-oligopolistico incentrato principalmente sulle logiche e le dinamiche della *platform economy*, segnando il passaggio dall'antropocentrismo al *datocentrismo*: tutto si *datifica*, tutto è dato.

Le piattaforme «spia[no], afferra[no] e incorpora[no] la preda»[29] attraverso una tecnica apparentemente blanda che fa leva sulla volontà di trascendenza della persona, sui simboli e sulle rappresentazioni esperienziali che annullano l'autonomia di utenti attratti dalle fascinazioni suadenti del potere economico.

L'estrazione del valore economico dalla persona avviene attraverso il progressivo utilizzo della rete, dentro la logica organizzativa di un mondo rigido, racchiuso in un codice binario e caratterizzato dalla «diminuzione della capacità di agire e di poter fare»[30]: più l'utente naviga, più diventa produttore; più diventa produttore, più è consumatore; più consuma, più viene consumato; più è consumato, più perde quell'autocoscienza che lo rende impermeabile ad una logica profittevole fondata sull'acquisto e la *ricommercializzazione* di gusti e abitudini, dietro il *tranello* della gratuità dei servizi.

L'utente viene continuamente bersagliato nei punti più fragili che ne disvelano la personalità e i tratti caratterizzanti del suo essere: le ambizioni, i desideri, le passioni e le aspirazioni diventano l'*humus* delle piattaforme digitali che si nutrono di tutto ciò che filtra dall'essere umano per trasformarlo in occasioni di profitto attraverso la movimentazione continua dei dati da parte di algoritmi e *blockchain*.

La rete modella l'utente e lo sottomette alle regole autoimposte dai sovrani digitali, mirando alla massimizzazione del profitto a detrimento delle solide garanzie dei diritti fondamentali: un ecosistema a sé stante, uno spazio nel quale si manifestano nuovi e inediti centri di potere e gestito da chi, attraverso procedure opache, estrae dati, manipola comportamenti, influenza dinamiche private, penetra nella sfera dei diritti e delle libertà individuali, alterando i tradizionali meccanismi di funzionamento della dialettica democratica, sfruttando così la naturale fragilità umana che si abbandona ai bisogni e alle aspirazioni individualiste.

La concentrazione che le piattaforme digitali realizzano delle sfere economiche, sociali, culturali, politiche – storicamente rappresentate come spazi di libertà di fronte al potere politico –determina for-

[27] Secondo M. Castells, *Galassie Internet*, cit., meritocratico, *hacker*, comunitario/virtuale e imprenditoriale.

[28] N. Bobbio, *Il problema del potere. Introduzione al corso di scienza della politica*, Torino, 1966, 37. La letteratura, sul punto, è amplissima. Si riferiscono alle piattaforme come «poteri ibridi» o «para-statali», G. De Gregorio, *Digital Constitutionalism in Europe. Reframing Rights and Powers in the Algorithmic Society*, Cambridge, 2022; A. Simoncini, E. Longo, *New technologies and the rise of the algorithmic society*, in H-W. Micklitz, O. Pollicino, A. Reichman, A. Simoncini, G. Sartor, G. De Gregorio (a cura di), *Constitutional Challenges in the Algorithmic Society*, Cambridge, 2021; O. Pollicino, *L' "autunno caldo" della Corte di giustizia in tema di tutela dei diritti fondamentali in rete e le sfide del costituzionalismo alle prese con i nuovi poteri privati in ambito digitale*, in *Federalismi.it*, 19, 2019; O. Grandinetti, *Le piattaforme digitali come "poteri privati" e la censura online, in Rivista italiana di informatica e diritto*, 1, 2022, 175-188; M. Betzu, *I poteri privati nella società digitale: oligopoli e antitrust*, in *Diritto Pubblico*, 3, 2021.

[29] E. Canetti, *Massa e potere*, Zurigo, 1960, trad. it. a cura di F. Jesi, Milano, 2024, 243.

[30] P. Ricoeur, *Sé come un altro*, Milano, 1993, 286. Nello stesso senso, M. Cuniberti, *Tecnologie digitali e libertà politiche, in Diritto dell'informazione e dell'informatica*, 22015, 278.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

me di diseguaglianza, subordinazione e vulnerabilità che disvelano, sul piano della struttura sociale, un'amplificata asimmetria tra il *dover essere costituzionale* e l'*essere* effettivo del diritto, tra la potenza di poteri (privati) e l'impotenza di contro-poteri (pubblici) garantisti[31].

Ciò esclude, in radice, ogni forma di rivendicazione di autonomia di utenti «sottomessi alla logica proprietaria»[32], vincolati alle dinamiche concorrenziali[33], in un contesto di diritti negati e di significativa limitazione della capacità di azione autonoma di un individuo condizionato in «pensiero, scelte e intenzioni»[34], sottoposto a forme di controllo capaci di produrre effetti rilevanti nella sua sfera giuridica e personale[35] e che lo incasellano in una «gabbia costruita da altri»[36], creandone, attraverso algoritmi di tipo probabilistico, una proiezione futura che preconizza non solo quel che ha già manifestato di essere, ma anche ciò che si ritiene potrà essere.

L'utente è pienamente consapevole di quanto la rete amplifichi l'accesso a beni, servizi e contenuti digitali ma, allo stesso tempo, non percepisce quanto tali tecniche possano essere così invasive da orientarne e attirarne l'attenzione in misura crescente, anche attraverso le «generali strategie persuasive correlate al *design* dei prodotti dei servizi digitali»[37].

In questo modo, l'utilizzo di tecniche ad alto impatto invasivo[38] si fa prassi in ambienti che proseguono nel percorso erosivo del ruolo dello Stato e dell'affievolimento dei diritti, mostrando la persona ancora più fragile di quanto non lo sia già nel mondo analogico.

E, invero, la vita digitale fa emergere un profilo di *doppia* vulnerabilità, legata, da un lato, ad una «condizione primaria di necessità e bisogno»[39], che esorbita al di là di ogni aspetto di recrudescenza delle libertà fondamentali e dell'autodeterminazione; dall'altro, alla dissolvenza dello Stato in luogo

---

[31] F. Paruzzo, *I sovrani della rete. Piattaforme digitali e limiti costituzionali al potere privato*, Napoli, 2022, *passim*.

[32] S. Rodotà, *Vivere la democrazia*, Roma-Bari, 2018, 94.

[33] F. Pizzolato, *Mutazioni del potere economico e nuove immagini della libertà*, in *Costituzionalismo.it*, 3, 2017, 3, ma anche F. Pizzolato (a cura di), *Libertà e potere nei rapporti economici. Profili giuspubblicistici*, Milano, 2010. Allo stesso modo, vi è tornato L. Ferrajoli, *Principia juris. Teoria del diritto e della democrazia*, II, Roma-Bari, 2007, 19 ss.

[34] C. Pinelli, *Pluralismo e democrazia nella società digitale. È tempo di regole*?, Seminario Astrid, 29 gennaio 2021. Per un'ampia trattazione delle modalità di interferenza dell'*agency* algoritmica e della datificazione che presuppone sui processi di costruzione identitaria, si veda S. Tiribelli, *Identità personale e algoritmi. Una questione di filosofia morale*, Roma, 2023. L'Autrice sottolinea come il fattore algoritmico, ampliando o riducendo gli orizzonti epistemici e comunicativi, agisca direttamente sui processi di costruzione identitaria riplasmando tutte le sue dimensioni: l'identità epistemica, l'identità socio-relazionale, l'identità morale, dove per quest'ultima s'intende la possibilità di *costruzione autoriale* del proprio progetto identitario. In merito, cfr. altresì I. de Vivo, *Il sé allo specchio dell'algoritmo. Libertà epistemica e identità individuale*, in A. Sterpa (a cura di), *L'ordine giuridico dell'algoritmo*, Napoli, 2023, 24-33.

[35] L. Ferrajoli, *La costruzione della democrazia. Teoria del garantismo costituzionale*, Roma-Bari, 2021, 205 li definisce, nel cap. 3 par. 2, «diritti civili di autonomia privata».

[36] S. Rodotà, *Il diritto di avere diritti*, Roma-Bari, 2012, 30.

[37] I. Geraci, *Minori e pubblicità mirata*, in *Diritto, Mercato, Tecnologia*, 24 gennaio 2022, 2.

[38] Diffusamente, A. Jablonowska, A.M. Maciej Kuziemski, H.W. Novak, P. Micklitz, P. Palka, G. Sartor, *Consumer law and artificial intelligence Challenges to the EU consumer law and policy stemming from the business' use of artificial intelligence*, in *EUI Working Paper LAW*, 11, 2018.

[39] J. Butler, *Violenza, lutto, politica*, in *Vite precarie*, Roma, 2004, 52.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

di uno «spazio senza confini»[40], nel quale il diritto rincorre «l'ovunque»[41] e dove la persona è estromessa da ogni «legame territoriale o di sangue»[42].

Ciò ha fatto sì che lo Stato medesimo abdicasse al monopolio della sovranità interna per mezzo della quale ha, sì, talvolta esercitato la triplice essenza del potere[43], ma allo stesso tempo ha assunto impegni e obblighi non limitati al riconoscimento delle tradizionali libertà liberali, ma estesi all'accoglimento delle pretese del popolo, dai diritti sociali ai nuovi diritti[44], annoverando la protezione delle categorie tradizionalmente fragili: i minori, i diversamente abili[45], le donne, insomma, tutti coloro che si trovano in una posizione deteriore rispetto alla maggioranza del corpo sociale e che – non per ciò solo – sono tanto più vulnerabili rispetto ai poteri pubblici e alle prevaricazioni dei soggetti privati.

È questa, probabilmente, la caratteristica principale dell'ultima evoluzione dello Stato contemporaneo – «ambivalente»[46] per certi aspetti – chiamato ad esercitare la forza *sul* popolo e *per* il popolo che ne riconosce la funzione di arbitro, garante, anche fustigatore – se necessario a preservare l'interesse collettivo all'ordine pubblico – ma nel prisma dei limiti sanciti dal costituzionalismo, inteso quale strumento di garanzia, protezione e riequilibrio delle disparità nel tracciato dei diritti inviolabili dell'uomo che lo impegnano ad elevare la condizione di fragilità umana a dignità della persona[47], di modo che «il potere arresti il potere»[48].

Lo spazio digitale, al contrario, segna il progressivo arretramento dello Stato dinanzi ad un mondo che manifesta i soli caratteri negativi della triplice concezione statuale weberiana[49] e fondato su logiche prettamente economiche, riluttante ai limiti all'esercizio di poteri che le Carte costituzionali, frutto del patto tra governanti e governati, pongono[50].

---

[40] D. Di Sabato, *Il ruolo delle piattaforme digitali nello svolgimento delle attività economiche in rete,* in *Annali della Facoltà Giuridica dell'Università di Camerino – Studi –*, 9, 2020, 5.

[41] N. Irti, *Le categorie giuridiche della globalizzazione*, in *Rivista di Diritto Civile*, 5, 2002, 625 ss.

[42] S. Rodotà, *Relazione 2002. Discorso del Presidente Stefano Rodotà*, Roma, Garante per la protezione dei dati personali, 2003, 7.

[43] G. Dumézil, *Jupiter, Mars, Quirinus*, Torino, 1955.

[44] D. Morana, *I diritti costituzionali in divenire. Tutele consolidate e nuove esigenze di protezione*, Napoli, 2022.

[45] Sul tema, cfr. G. Arconzo, *I diritti delle persone con disabilità. Profili costituzionali*, Milano, 2020.

[46] F. Dei, C. Di Pasquale (a cura di), *Stato, violenza, libertà. La «critica del potere» e l'antropologia contemporanea*, Milano, 2018.

[47] Si rinvengono tracce di questa idea di Stato nell'impostazione di P. Bordieu, *Sullo Stato Corso al Collège de France*, trad. it. a cura di M. Guareschi, Milano, 2021: ordine, sicurezza, giustizia, interesse generale.

[48] C. de Montesquieu, *Lo spirito delle leggi,* libro XI, capitolo IV.

[49] «associazione di predatori» (M. Weber, *Alcune categorie della sociologia comprendente (1913)*, in Id., *Il metodo delle scienze storico-sociali*, Torino, 1958, 273), «pretesa di monopolio della coercizione fisica» (M. Weber, *Economia e società*, I, Milano, 1961, 53), «associazione di dominio» (M. Weber, *La politica come professione*, Roma, 1998, 182).

[50] La riflessione sul costituzionalismo digitale transnazionale, che scinde diritto e politica, è stato indagato da G. Teubner, *Societal Constitutionalism; Alternatives to State-Centred Constitutional Theory?*' in C. Joerges, I.J. Sand, G. Teubner (a cura di), *Transnational Governance and Constitutionalism. International Studies in the Theory of Private Law*, Hart 2004; Id., *Constitutional Fragments: Societal Constitutionalism and Globalization,* Oxford, 2012; Id. *The project of constitutional sociology: Irritating nation state constitutionalism,* in *Transnational Legal Theory,* 4, 2013, 44-58; Id. *Nuovi conflitti costituzionali,* Milano, 2012, 69; Id. *Il costituzionalismo della società transnazionale, in Quaderni costituzionali,* 1, 2014, 196.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

In un ecosistema come quello digitale, per vocazione transnazionale, che si caratterizza appunto per la cesura di quell'accoppiamento strutturale tra diritto e politica, che nelle Costituzioni nazionali ha trovato la propria sintesi[51], il c.d. costituzionalismo digitale[52] (o *neocostituzionalismo*) è emerso e si è affermato, infatti, dal basso e sulla spinta di soggetti che operano al di fuori di un contesto strettamente politico, come le organizzazioni non governative e le comunità epistemiche[53].
Lo Stato appare pertanto sempre più in affanno nel mantenere la storica ascendenza sui privati e tra i privati nella dimensione relazionale-digitale, «scarnificato»[54], come è, dalla progressiva azione di «scoronamento»[55] della globalizzazione (prima) e del progresso tecnologico (poi).

3. La regolazione delle piattaforme digitali. Dai profili economici alle ambizioni di tutela degli utenti

L'utente, continuamente bersagliato dalle logiche travolgenti e persuasive della *platform economy*[56], in assenza della tutela statuale, come può difendersi dal potere di sovrani operanti in una dimensione profondamente diversa da quella analogica, nella quale le individualità scompaiono, tutti i soggetti sono apparentemente uguali davanti alle regole imposte, certo, ma lo sono a prescindere dalle peculiarità di ciascuno?
Se, da un lato, è necessario rispondere alla descritta metamorfosi della natura di una forza *esterna* in grado di limitare e vincolare l'autonomia e la libertà del singolo – non più soltanto pubblica, ma privata o parapubblica –, dall'altro è necessario prendere atto dell'esistenza di una forza autonoma, forse meno visibile, ma altrettanto dirompente: gli algoritmi[57] detengono una capacità talmente pervasi-

[51] Cfr. G. Teubener, *Global Bukovina. Legal pluralism in the World Society*, in G. Teubner (a cura di), *Global law without a State*, Dartmouth, 1997.

[52] Cfr., *ex multis*, O. Pollicino, *Di cosa parliamo quando parliamo di costituzionalismo digitale?*, in *Quaderni Costituzionali*, 3, 2023.

[53] Sul tema del costituzionalismo digitale come forma di costituzionalismo spontaneo (*societal constitutionalism*), si veda G. Teubner, *Societal Constitutionalism; Alternatives to State-Centred Constitutional Theory?'*, in C. Joerges, I.J. Sand, G. Teubner (a cura di), *Transnational Governance and Constitutionalism. International Studies in the Theory of Private Law*, Hart 2004; Id. *Constitutional Fragments: Societal Constitutionalism and Globalization*, Oxford, 2012; Id., *The project of constitutional sociology: Irritating nation state constitutionalism*, in *Transnational Legal Theory*, 4, 2013, 44-58,3; Id., *Nuovi conflitti costituzionali*, Milano, 2012, 69; Id., *Il costituzionalismo della società transnazionale*, in *Quaderni Costituzionali*, 1, 2014, 196; A. jr. Golia, G. Teubner, *Societal Constitutionalism: Background, Theory, Debates*, in *ICL Journal*, 4, 2021, 357-411. Per una visione critica, si veda M. Betzu, *op. cit*. Sul rapporto tra progetto di sovranità digitale europea come «terza via» alla regolazione da leggersi nel *framework* del costituzionalismo digitale, si veda M. Santaniello, *Sovranità digitale e diritti fondamentali*, in *Rivista italiana di informatica e diritto*, 1, 2022, 49; E. Celeste, *Digital Sovereignty in the EU: Challenges and Future Perspectives*, in F. Fabbrini, E. Celeste, J. Quinn (a cura di), *Data Protection beyond Borders: Transatlantic Perspectives on Extraterritoriality and Sovereignty*, Hart, 2021.

[54] A. Mastropaolo, *Fare la guerra con altri mezzi*. *Sociologia storica del governo democratico*, Bologna, 2023, 18.

[55] G. Capograssi, *Saggio sullo Stato*, Torino 1918, poi in Id., *Opere*, a cura di M. D'Addio e E. Vidal, Milano 1959.

[56] A. Gawer, *Platforms, Markets and Innovation*, Northampton (MA), 2010.

[57] I. de Vivo, *Il potere d'opinione delle piattaforme-online: quale ruolo del "regulatory turn" europeo nell'oligopolio informativo digitale?*, in *Federalismi.it*, 2, 2024, 45-75.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

va di influenzare e sostituire l'autodeterminazione individuale dall'interno tale da sovvertire profondamente il concetto di autonomia, indipendenza e capacità di agire sotteso al modello liberale[58].

Quanto è vulnerabile la persona in un contesto de-statualizzato è chiaramente percepibile nella misura in cui quelle garanzie costituzionali, parte del patto tra lo Stato e il popolo, vengono meno per il venir meno del patto stesso, nel quale il sovrano incontra i suoi limiti, certo, ma è di converso sottoposto a precisi obblighi di tutela della persona, oltre ogni «legame territoriale o di sangue»[59], per il solo fatto che i diritti umani sono tali, e quindi inviolabili, a prescindere dal legame politico con lo Stato.

D'altronde, l'uomo è forte nei confronti del potere statuale quanto più il livello di garanzia dei suoi diritti nelle Costituzioni – «espressione di uno stadio evolutivo culturale, un mezzo di autorappresentazione culturale del popolo, lo specchio di un patrimonio culturale e fondamento delle sue speranze»[60] – è lungo, largo e profondo[61].

Così, nell'esigenza di mantenere una separazione tra il dogma e il rifiuto del progresso, bisogna individuare la fonte di un «diritto faticoso»[62] in grado di preservare libertà e diritti della persona nello spazio digitale e consentire il ritorno dello Stato dentro la dimensione «*on life*»[63], a partire da un nucleo duro di garanzie poste proprio a tutela di quegli aspetti più vulnerabili della persona (analogica e digitale), preservando quei princípi «storicamente e materialmente indisponibili ai poteri stessi»[64] nella dimensione dominante-dominato[65] (finanche nei rapporti interprivati[66]), ogni volta in cui si determina un'asimmetria riconducibile a forme di cogenza simile agli schemi di potere statuale[67].

Quella fonte, un po' per vocazione strutturale e un po' per ragioni di competenza, risiede nel diritto europeo (non che le Costituzioni e la legislazione ordinaria degli Stati membri sia indifferente a quanto accade nel mondo virtuale e rispetto a tutte quelle situazioni in cui la libertà di iniziativa economica privata deve fare i conti – e come non potrebbe – con il nucleo duro dei valori concernenti la dignità umana, la libertà di scelta, l'autodeterminazione).

---

[58] R. Bodei, *Dominio e sottomissione. Schiavi, animali, macchine, Intelligenza Artificiale*, Bologna, 2019. Secondo l'A., dal «capitalismo algoritmico» (che utilizzerà l'intelligenza artificiale e la robotica per legare sempre più l'economia e la politica ad alcune forme di conoscenza) ha origine un nuovo potere «occulto» in cui «il *logos* umano sarà sempre più soggetto a un logos impersonale».

[59] S. Rodotà, *Relazione 2002. Discorso del Presidente Stefano Rodotà*, Roma, Garante per la protezione dei dati personali, 2003, 7.

[60] P. Häberle, *Costituzione e identità culturale*, Milano, 2006, 11.

[61] E. Mounier, *Rivoluzione personalista e comunitaria*, Milano, 2022, 90 ss.

[62] S. Rodotà, *Dal soggetto alla persona. Trasformazioni di una categoria giuridica*, in *Filosofia politica*, 3, 2007, 375 ss.

[63] L. Floridi, *The Onlife Manifesto. Being Human in a Hyperconnected Era*, Berlino-Heidelberg, 2015.

[64] F. Paruzzo, *op. cit*.

[65] L. Ferri, *Nozione giuridica di autonomia privata,* in *Rivista Trimestrale di Diritto e Procedura Civile*, 1947, 129.

[66] R. Alexy, *Teoria dei diritti fondamentali*, Bologna, 2012, 570-571 e G. Lombardi, *Potere privato e diritti fondamentali*, Torino, 1970, 88.

[67] Secondo R. Conti, F. De Stefano, O. Pollicino, *L'algoritmo e la nuova stagione del costituzionalismo digitale: quali sfide per il giurista(teorico e pratico)?*, in *Giustizia insieme*, 15 aprile 2021, «se le coordinate del rapporto tra autorità e libertà mutano geometria, e si spostano da una dimensione verticale ad una orizzontale, relativa al rapporto tra piattaforme e utenti, anche il costituzionalismo, in accordo con la sua vocazione evolutiva, dovrebbe essere in grado di cambiare prospettiva».

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Certo, l'oggetto della tutela assume una connotazione di natura economica per il progressivo mutamento delle relazioni interpersonali che, nella dimensione digitale, hanno assunto per lo più aspetti negoziali, come in un qualsiasi mercato. Non poteva essere altrimenti: quando si parla di piattaforme digitali, inevitabilmente la tutela della persona si sposta in una dimensione strettamente economica.

Che le piattaforme avessero un ruolo dominante sulla vita delle persone, lo si poteva avvertire all'indomani dei primi *software* di scambio *peer to peer* e delle controversie legate alle violazioni del *copyright*. Per percepirne pericoli, tuttavia, il legislatore e gli utenti hanno dovuto comprendere quanto i dati e le informazioni potessero essere soggette ad uno sfruttamento economico senza precedenti[68].

È in questa fase, prima ancora dell'avvento *disruptive* dell'intelligenza artificiale – solo recentemente regolato dall'*AI Act* (reg. UE 2024/2689) – che il legislatore ha predisposto misure incisive tanto per la tutela dei dati sensibili (reg. 2016/679), quanto per i diritti economici, come il diritto autorale su tutti (dir. CE 2001/29, reg. UE 1383/2003, dir. UE 2008/95).

In Italia, in esempio, si è posto sovente il quesito se l'azione dei privati non contrastasse con norme costituzionali di riferimento (artt. 41 e 47 Cost.) per individuare la competenza decisionale delle autorità pubbliche di regolazione del mercato. L'art. 41 Cost. attribuisce una situazione giuridica che corrisponde ad un vero e proprio *status* di fronte all'ordinamento, a tal punto che l'iniziativa economica incontra un limite in quella espressione dell'«utilità sociale» (c. 2) riconducibile alla meritevolezza delle garanzie che deve caratterizzare un qualsiasi rapporto negoziale interprivato.

Se, dunque, è vero che l'iniziativa economica del singolo diviene, nel suo concreto esercizio, attività economica privata, perché quest'ultima sia in contrasto con la stessa norma costituzionale deve essere contraria all'utilità sociale. Detto altrimenti, il rapporto tra utilità individuale (basti pensare all'istituzione e allo scambio su piattaforme) e utilità sociale (il risparmio, la stabilità economica) impone di guardare la prima in maniera doverosamente recessiva, ossia tenendo conto del principio di proporzionalità, come misura in cui il valore costituzionale può sottrarre o deve sottrarre uno spazio decisionale al mercato, atteso che quando si parla di piattaforme digitali, inevitabilmente la tutela della persona si sposta in una dimensione strettamente economica e, conseguentemente, privatistica.

A partire da tali principi, il legislatore italiano, in esempio, ha disciplinato recentemente le piattaforme di scambio di criptoattività, tutelando l'informazione nel mercato, regolando così ciò che viene offerto in termini anche monetari e l'accesso ad alcuni servizi che è reso subordinato a condizioni oggettive e soggettive poste a garanzia dell'utente. Da ultimo, con il d.l. n. 25/2023 (c.d. *decreto Fintech*) sulla circolazione degli strumenti finanziari digitalizzati, anche le imprese innovative, come quelle che gestiscono le piattaforme *metaversanti*, sono sottoposte al rispetto della dettagliata normazione secondaria di settore[69].

La legislazione europea tenta di compiere un passo ulteriore, sebbene la condizione di vulnerabilità sia stata sempre colta – anche dalle Corti[70] – in astratto e in modo «indefinito»[71], in una dimensione

[68] Cfr., su questi temi, A. Patroni Griffi (a cura di), *Bioetica, diritti e intelligenza artificiale*, Mimesis, 2023 e G. Cerrina Feroni, C. Fontana, E.C. Raffiotta (a cura di), *AI Anthology. Profili giuridici, economici e sociali dell'intelligenza artificiale*, Bologna, 2022.

[69] M. Passaretta, *Nuove traiettorie societarie nell'era del FinTech: la digitalizzazione delle partecipazioni sociali*, in *MediaLaws – Rivista del diritto dei media*, 1, 2024.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

esclusivamente relazionale[72] e gruppocentrica[73], sebbene ricondotta alla garanzia dei diritti[74], quale «fondamento dei diritti umani»[75], a partire da un innalzamento del livello di tutela di un soggetto che il costituzionalismo digitale non può che considerare ontologicamente fragile fin dall'approccio con il virtuale.

Il modello europeo di regolazione delle piattaforme esprime, nelle premesse, la massima convinzione della natura vulnerabile dell'utente, il quale è travolto da una spirale di potere illimitato che penetra negli anfratti più remoti della sua personalità, leggendone le passioni, interpretandone i bisogni e così muovendolo come *pedina in uno scacchiere.*

Da una lettura complessiva dei due principali regolamenti che attualmente disciplinano l'universo delle piattaforme, emergono luci, certo, ma soprattutto ombre in relazione al profilo concernente il livello di tutela medio/alto degli utenti maggiormente vulnerabili, almeno in linea con i considerando del reg. UE 2022/2065 (*Digital Services Act – DSA*) e del reg. UE 2022/1925 (*Digital Markets Act – DMA*).

4. Le garanzie in favore degli utenti vulnerabili nel reg. UE 2022/2065 (Digital Services Act) e nel reg. UE 2022/1925 (Digital Markets Act). Luci e ombre

La delicatezza del tema è stata progressivamente colta dal legislatore fin dall'«affermazione e ascesa»[76] del diritto alla protezione dei dati personali, a partire dal quale l'Unione europea ha individuato l'elemento maggiormente problematico delle questioni nell'utilizzo profittevole delle informazioni proiettate negli spazi digitali.

La consapevolezza, da parte del legislatore europeo, di quanto l'utente si abbandoni ad una condizione ontologica di vulnerabilità la si percepisce sin dai preamboli dei regolamenti concernenti la *Strategia digitale dell'Unione europea*[77], concretizzatasi fino ad ora con il *Digital Markets Act* (reg. UE 2022/1925), il *Digital Services Act* (reg. UE 2022/2065), l'*Artificial Intelligence Act* (reg. UE

[70] Corte di Giustizia (Grande Sezione), 13 maggio 2014*, Google Spain SL e Google Inc.* c. *Agencia Española de Protección de Datos (AEPD)* e *Mario Costeja González*, Causa C-131/12.

[71] R. Chenal, *La definizione della nozione di vulnerabilità e la tutela dei diritti fondamentali*, in *Ars Interpretandi*, 2, 2018.

[72] M. Ainis, *I soggetti deboli nella giurisprudenza costituzionale*, in *Politica del diritto*, 1, 1999, 26 ss.

[73] L. Peroni, A. Timmer, *Vulnerable Groups: The Promise of an Emerging Concept in European Human Rights Convention Law*, in *International Journal of Constitutional Law*, 4, 2013, 1056 ss.

[74] R. Chenal, *op. cit.*, 51 ss., ma anche D. Poletti, (voce) *Soggetti deboli*, in *Enciclopedia del Diritto Annali*, Milano, 2014, 964 ss.

[75] È la tesi di R. Andorno, *Is Vulnerability the Foundation of Human Rights?*, in A, Masferrer, E. Garciìa-Sanchez (a cura di), *Human Dignity of Vulnerable in the Age of Rights. Interdisciplinary Perspective*, Cham, 2016, e di B.S. Turner, *Vulnerability and Human Rights*, Pennsylvania, 2004.

[76] L. Califano, *Privacy: affermazione e pratica di un diritto fondamentale*, Napoli, 2016.

[77] Risoluzione del Parlamento europeo del 25 novembre 2020 *Verso un mercato unico più sostenibile per le imprese e i consumatori* (2020/2021 INI). L'obiettivo del nuovo *corpus* regolamentare (*Digital Services Package*) sarebbe quindi quello di «abilitare proceduralmente» una forma di costituzionalismo «autopoietico» al fine di rendere effettiva e giustiziabile la protezione dei diritti fondamentali al di là dei confini territoriali (I. de Vivo, *Sfide esistenziali e resilienze identitarie nella geopolitica informazionale: l'identikit europeo tra sovranità e costituzionalismo digitale*, in *Diritto Pubblico Europeo Rassegna online*, spec. 1, 2024).

Downloaded from www.biodiritto.org.
ISSN 2284-4503

2024/2689), il *Data Act* (reg. UE 2023/2854) e il *Data Governance Act* (reg. UE 2022/868), questi ultimi due fondamentali per una nuova stagione politica eurounitaria in materia di monetizzazione e tutela dei dati.

Con il reg. UE 1689/2024 sembra essersi *chiuso il cerchio* di una disciplina concernente i diversi profili di sfruttamento economico dei dati degli utenti nell'immensità dello spazio digitale attraverso gli algoritmi.

È proprio la regolazione europea – che opera in virtù dell'attrazione della materia alla propria sfera di competenza – ad aver assunto la piena consapevolezza (pur con alcuni *caveat*, di cui si dirà) che la condizione ontologicamente vulnerabile dell'utente non è il prezzo da pagare ai *gatekepeer* o ai fornitori di servizi per *aprire le soglie* dei mondi digitali, quanto piuttosto il profilo tramite cui misurare il grado di tutela che le piattaforme sono obbligate a riservare ai «destinatari» (art. 3, par. 1, lett. *b)*, *DSA*, art. 2, par. 20, *DMA*), andando oltre una disciplina fondata sul mero rilascio del consenso informato[78].

Pur nella prospettiva *alta* di tutela di tutte quelle situazioni che possono generare dei rischi significativi e che trovano fondamento nelle linee guida adottate dalla *Dichiarazione europea sui diritti e i principi digitali per il decennio digitale* (2023/C 23/01), la regolazione europea segna talvolta luci, talaltra ombre.

Tra gli atti normativi del pacchetto digitale europeo, il *Digital Services Act* e il *Digital Markets Act* sono stati emanati allo scopo di disciplinare le piattaforme *online*, prevedendo nuovi obblighi per i *provider* e nuovi diritti per gli utenti: si tratta di una regolazione che ha avuto il pregio di arricchire le procedure sanzionatorie (art. 51 *DSA* e art. 30 ss. *DMA*), di garantire adeguata informazione e trasparenza (artt. 15, 27, 39, 42, *DSA* e *Considerando* 45, 58, 71, 72, *DMA*), rappresentando il frutto di una impostazione radicalmente diversa dalla disciplina precedente concernente i soli *host provider*, sottratti a diverse forme di responsabilità, salvo il solo commercio dei falsi o la diffusione di contenuti terroristici (reg. UE 2021/784).

La classificazione delle piattaforme quali spazi neutrali di intermediazione di contenuti (*hosting-providers*) ne aveva consentito lo sviluppo in un ambiente protetto dalla regolamentazione statale, beneficiando di una sostanziale immunità determinata dal principio di esenzione della responsabilità per le conseguenze sociali e giuridiche delle pubblicazioni di terzi che avvenivano per loro tramite (*providers exemption from liability*).

Il principio del c.d. *safe harbor* – corollario del regime transnazionale di *industry self-regulation* introdotto nella legislazione statunitense, con sezione 230(c)(I) del *Tele -communications Act* del 1996 – ha trovato il suo gemello europeo nella *E-Commerce Directive* 2000/31 ed è rimasto sostanzialmente immutato fino all'emanazione del nuovo regolamento europeo *Digital Services Act* che introduce – per la prima volta a livello paneuropeo – il cosiddetto meccanismo di *notice and takedown* (art. 16), imponendo alle piattaforme il dovere giuridico di analizzare rapidamente e, se necessario, rimuovere i contenuti segnalati come illeciti[79].

[78] Sul modello *consensualistico,* cfr. le critiche di Cfr. S. RODOTÀ, *Tecnologie e diritti*, Bologna, 1995, 82; A.M. GAMBINO, *Big data e fairness. Il ruolo delle authorities*, in *Nuovo diritto civile*, 2020, 298 ss.; G. FINOCCHIARO, *Il quadro d'insieme sul Regolamento europeo*, in ID., (diretto da), *Il nuovo regolamento europeo sulla privacy e sulla protezione dei dati personali,* Bologna, 2017, ss.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Si tratta di un ulteriore stadio evolutivo della regolazione che, pur restando sbilanciata maggiormente su aspetti economici – ma ciò, come detto, è fisiologicamente legato alla novità e alla natura delle circostanze regolate – accetta la sfida posta dalle piattaforme che si nutrono dei dati *immagazzinati* dalla rete telematica per costruire una modalità di lettura e attrazione dei bisogni più capillare e fondata principalmente su algoritmi generativi.

Il *Digital Services Act* e il *Digital Markets Act* aspirano a rendere talune tipologie di mercato e di servizi più sicure per gli utenti, senza per ciò solo sacrificare la libertà di impresa, predisponendo un uniforme ambiente regolativo nel mercato unico europeo[80], dove piattaforme e utenti si relazionano in un ambiente equilibrato (almeno in teoria), e dove le garanzie approntate ad alcune categorie *speciali* muovono dalla considerazione dell'esistenza di una condizione di maggiore vulnerabilità legata per lo più all'età dei naviganti.

Le preoccupazioni espresse nelle numerose determinazioni finalistiche che dominano le premesse di entrambi i regolamenti lasciano intendere, tuttavia, un'affermazione più *alta* di alcune categorie speciali sottoposte con maggiore facilità a pratiche discriminatorie, violente e aggressive.

Chi si è occupato del tema ha già manifestato alcune perplessità per una regolazione certamente difensiva, ma non particolarmente sensibile verso quegli utenti più facilmente esposti a pratiche discriminatorie (i disabili), truffe (gli anziani), violenze (donne), concentrandosi maggiormente sull'unico profilo vulnerabile espressamente tutelato – i minori – quando, invero, la collocazione al centro del modello regolatorio anche delle altre categorie *fragili* avrebbe consentito di fortificare quella paratìa invalicabile oltre la quale la penetrazione di *software* e algoritmi deve essere vietata o talmente stringente da prevedere una serie di obblighi e responsabilità[81] in capo al titolare del trattamento dei dati (diffusamente nel reg. UE 679/2016), al *gatekepeer* (artt. 5, 6 e 7, reg. UE 2022/1925), ai prestatori di servizi intermediari e ai prestatori di memorizzazione di informazioni (diffusamente, nel Reg. UE 2022/2065) in tutte quelle attività considerate ad alto rischio, concernenti l'immissione sul mercato, la messa in servizio o l'uso di un sistema di intelligenza artificiale che «sfrutta le vulnerabilità di uno specifico gruppo di persone, dovute all'età o alla disabilità fisica o mentale, al fine di distorcere materialmente il comportamento di una persona che appartiene a tale gruppo in un modo che provochi o possa provocare a tale persona o a un'altra persona un danno fisico o psicologico» (art. 5, reg. UE 2024/2689).

Ed invero, proprio dalla lettura dei *Considerando* del *DMA* e del *DSA,* densi di *alert* e grondanti di determinazioni finalistiche, probabilmente ci si sarebbe atteso qualcosa in più rispetto ad un modello sostanzialmente procedurale[82].

---

[79] Su questi aspetti, *amlius*, I. DE VIVO, *The "neo-intermediation" of large on-line platforms: Perspectives of analysis of the "state of health" of the digital information ecosystem*, in *Communications*, 3, 2023, 11.

[80] Su profili più ampi, si rinvia a A. MANGANELLI, A. NICITA, *Regulating Digital Markets. The European Approach*, Cham, 2022.

[81] Cfr. M.R. ALLEGRI, *Il futuro digitale dell'Unione europea: nuove categorie di intermediari digitali, nuove forme di responsabilità*, in *Rivista italiana di informatica e diritto*, 2, 2021, 8 ss.

[82] Emerge dal complesso meccanismo di contrasto alle pratiche illegali, come evidenziato da E. BIRRITTERI, *Contrasto alla disinformazione,* Digital Services Act *e attività di private enforcement: fondamento, contenuti e limiti degli obblighi di compliance e dei poteri di autonormazione degli operatori*, in *MediaLaws – Rivista del diritto dei media*, 2, 2023.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Mentre il *Digital Markets Act* si rivolge principalmente alle piattaforme di *e-commerce* (delineando la disciplina dei *gatekeeper* che operano per lo più nella dimensione degli scambi commerciali *B2B* e *B2C* in un regime di mercato caratterizzato da concorrenza leale, trasparenza, equità, disciplinando la portabilità dei dati[83]), è il *Digital Services Act* ad accendere maggiormente la *spia* sugli utenti a seguito dell'applicazione di pratiche dannose, in specie con riferimento a destinatari finali vulnerabili come i minori (parr. 63, 95, 104 del preambolo, art. 28).

Il *DSA* regola i servizi di intermediazione digitale («*mere conduit», «caching», «hosting*») offerti dalle piattaforme e dai motori di ricerca *on line* (art. 3, par. 1, lett. *i)*) in funzione dei ruoli e della dimensione *di scala* delle piattaforme (artt. 11-13), riservando obblighi di diligenza asimmetrici rispetto alle *Very Large Platforms*, *Online Platforms*, *Hosting Services*, *All Intermediares* (art. 33 *DSA*): più la dimensione delle piattaforme è *large*, maggiori sono gli adempimenti supplementari (artt. 14, 15, 24 *DSA*) legati alla trasparenza[84] e ai divieti mirati ad impedire lo sfruttamento della vulnerabilità degli utenti[85] che fanno del *risk-based approach* una delle *best practices* concernenti la valutazione annuale dei rischi, a cui sono soggetti soprattutto quei profili maggiormente esposti alla manipolazione della sfera psicologica (art. 34, parr. 1 e 2, lett. *d)*, invitando ad attuare «misure mirate per tutelare i diritti dei minori, compresi strumenti di verifica dell'età e di controllo parentale, o strumenti volti ad aiutare i minori a segnalare abusi o ottenere sostegno, a seconda dei casi» (art. 35, par. 1, lett. *j)*), a cui si affianca un dovere di diligenza sancito dall'art. 44 *DSA*.

Pur in assenza di una determinazione soggettiva di quali possano o non possano essere catalogati come *utenti a rischio* – salvo affidarsi ad una ricostruzione giurisprudenziale[86] pregressa – il *DSA* prende posizione a favore solo di una delle categorie speciali (i minori, *ex* art. 28), vietando ogni forma di pubblicità basata sulla profilazione di dati *ex* art. 9, par. 1, reg. UE 2016/679[87].

In merito alla profilazione dei minori, l'art. 28 irrigidisce il divieto di pubblicità se si è consapevoli, «con ragionevole certezza», che il destinatario sia un utente minore di età, ed estende il divieto di profilazione, così come definito *ex* art. 4, par. 4, GDPR, a prescindere dalla natura «speciale» dei dati utilizzati.

La norma non obbliga i fornitori di piattaforme *online* a trattare dati personali ulteriori per valutare se il destinatario del servizio sia minore, piuttosto si preoccupa – per le piattaforme *VLOP* e *VLOSE* – di

[83] Il regolamento sui mercati digitali mira a garantire la contendibilità dei mercati digitali combattendo *ex ante* le pratiche anticoncorrenziali dei *controllori dell'accesso* e a correggere gli squilibri causati dal loro dominio sul mercato digitale europeo. Il regolamento sui servizi digitali mira a responsabilizzare i fornitori di servizi intermedi (ISP) e a combattere la distribuzione di contenuti illegali. A tal fine, prevede responsabilità differenziate e obblighi di diligenza, come ad esempio obblighi di informazione e trasparenza.

[84] Trasparenza nella presentazione degli annunci sulle piattaforme online: marcature evidenti, identità dell'inserzionista, spiegazione del motivo per cui l'utente viene mostrato annuncio; marcature evidenti per i contenuti sponsorizzati (ad esempio, gli *advertorial* degli *influencer*); obbligo di mitigare i rischi sociali legati alla pubblicità, ad esempio nel monetizzare disinformazione o nel discriminare determinate categorie di utenti.

[85] Divieto di pubblicità mirati ai minori basati sulla profilazione; divieto pubblicità mirata basata sulla profilazione che includa categorie speciali di dati personali, come l'etnia, le opinioni politiche, i dati sulla salute, l'orientamento sessuale, ecc.

[86] Cfr. G. Goffredi, V. Lorubbio, A. Pisanò (a cura di), *Diritti umani in crisi? Emergenze, diseguaglianze, esclusioni*, Pisa, 2021.

[87] M. Iaselli, *Digital Services Act e privacy*, in *Diritto di internet*, 1, 2023.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

imporre valutazioni periodiche del rischio concernente l'impatto dei servizi sui diritti fondamentali e sulla diffusione di contenuti illegali, prevedendo misure mirate per tutelare i diritti dei minori, compresi gli strumenti di verifica dell'età e di controllo parentale, o quelli volti ad accompagnare i minori a segnalare abusi, ottenere sostegni, implementando le misure di contrasto alle *fake news*[88], ai *deep fake* (art. 35, par. 1, lett. *k)*, *DSA*)[89] e, più in generale, di ogni contenuto illegale ai sensi dell'art. 9 *DSA*, attraverso le procedure stabilite dal successivo art. 10, in modo da limitarne la profilazione ai sensi degli artt. 22, 38, 71, reg. UE 679/2016.

In particolare, l'art. 14, par. 3, *DSA*, espressamente dispone che se un servizio intermediario è principalmente destinato a minori o è utilizzato in prevalenza da questi (si pensi, i.e. ai *baby influencer*)[90], il prestatore deve illustrare, in modo chiaro per i minori medesimi, le condizioni e le restrizioni che si applicano all'utilizzo del servizio.

Quella ex art. 28 *DSA*, tuttavia, è l'unica categoria espressamente tutelata. Eppure, la ragionevolezza avrebbe potuto anche suggerire di apprestare ulteriori strumenti a garanzia di altri soggetti che si prestano più facilmente a pratiche di bullismo (i minori, certo, ma anche i disabili e gli autistici) o di aggressione sessuale (le donne) e discriminatorie dal punto di vista religioso (le donne e gli uomini di altre religioni)[91].

Talune categorie *speciali* – pur potendosi immaginare un'applicazione estensivamente analogica della regolazione – vengono talvolta *dimenticate* quando si tratta di rafforzarne la protezione nei confronti di manifestazioni preoccupanti di *cyberbullismo* ad alto tasso impattante sullo sviluppo emotivo e sul benessere dei soggetti fragili. Una *dimenticanza*, quest'ultima, che rischia di provocare conseguenze ancor più deleterie[92]. Non è chiaro il motivo per cui alle «persone con disabilità», pur espressamente definite come categoria speciale (art. 3, par. 1, lett. *v)*, *DSA*), sia stata negata una protezione supplementare, diversamente dai minori (art. 28 *DSA*) ai quali viene dedicato ampio spazio e rispetto ai quali la lettura dei *Considerando* lascia immaginare una tutela più pervasiva di quanto in realtà non lo sia.

---

[88] Su cui, già in passato, F. Pizzetti, *Fake news e allarme sociale: responsabilità, non censura*, in *MediaLaws – Rivista del diritto dei media*, 1, 2017.

[89] Cfr. M. Cazzaniga, *Una nuova tecnica (anche) per veicolare disinformazione: le risposte europee ai* deepfakes, in *MediaLaws – Rivista del diritto dei media*, 1, 2023.

[90] S. Van Der Hof, V. Verdoodt, M. Leiser, *Child Labour and Online Protection in a World of Influencers*, in *SSRN Electronic Journal*, 2019.

[91] Tale ultimo aspetto è disciplinato dalla direttiva UE 2024/1385, primo atto legislativo eurounionale che tratta in modo sistematico il problema della violenza di genere, definendo in modo specifico la violenza online, neanche prevista dalla convenzione ratificata dal Consiglio UE nel 2014 (*Convenzione di Istanbul*). Nel tentativo di innalzare *standard di tutela* e armonizzare quadri nazionali, prevede quindi l'introduzione di norme minime per la definizione delle fattispecie di reato online connesse al genere quali lo *stalking* (art. 6), le molestie (art. 7) e l'istigazione alla violenza o all'odio online con specifico riguardo al genere (art. 8), per i quali prevede la possibilità di emanare ordini giuridici vincolanti a carico dei prestatori di servizi intermediari di rimozione di tale materiale o di disabilitazione dell'accesso, andando ad integrare la nozione di illecito rilevante ai sensi del regolamento generale su servizi digitali (art. 23).

[92] In questo senso, M.T.M. Francese, *Cyberbullismo e genere. Modificazione antropologica dei nuovi adolescenti*, in *Ricognizioni. Rivista di lingue, letterature e culture moderne*, 18, 2022, 253-264, E. Corbo, B.E. Palladino, E. Meresini, *Bullismo e disabilità. Una revisione della letteratura*, in *Psicologia clinica dello sviluppo*, 2, 2021, 191-216.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Non mancano, poi, ulteriori profili di criticità che saranno probabilmente *sciolti* dalla sedimentazione della giurisprudenza eurounionale, in particolare nella parte in cui il *DSA* sembrerebbe poco coerente con le caratteristiche tecniche, in esempio, di servizi di memorizzazione di informazioni – come *cloud* e *hosting* – la cui attività principale non concerne divulgazione al pubblico di servizi, ovvero si configuri come una caratteristica minore o meramente accessoria connessa intrinsecamente al servizio principale (art. 2, par. 2, *DSA*).

Si tratta di un aspetto di debolezza in relazione a quelle manifestazioni di *body shaming*, odio, incitamento alla violenza, divulgazione di *fake news* poste in essere, in esempio, nei commenti alle notizie di cronaca diffuse da una piattaforma che eroga servizi di informazione, oppure a mezzo dei servizi di comunicazione interpersonale come *Whatsapp*, *Messenger*, *Telegram* che, in base alla direttiva UE 2018/1972, non rientrerebbero nell'ambito di applicazione della definizione di «piattaforma online», poiché sono utilizzati per la comunicazione interpersonale tra un numero limitato di persone stabilito dal mittente. Eppure sono proprio tali mezzi ad essere utilizzati dai più per *adescare* le categorie vulnerabili e generare, attraverso un primo approccio remoto, una spirale di sfruttamento e di violenza molto complessa da scovare e reprimere.

Tuttavia, gli obblighi previsti dal *DSA* per i fornitori di piattaforme *online* sono comunque applicabili ai servizi che consentono la messa a disposizione di informazioni a un numero potenzialmente illimitato di destinatari, non stabilito dal mittente della comunicazione, come ad esempio attraverso gruppi pubblici o canali aperti (*Considerando* 14), a mezzo dei quali sono sempre più diffuse quelle condotte penalmente rilevanti come il *revenge porn* che mette in pericolo l'integrità morale, fisica e psicologica soprattutto dei minori e delle persone fragili[93], oppure attività di *marketing* aggressivo che erodono la capacità di resistenza degli utenti che rischiano di cadere nella dinamica profittevole dei prestatori di servizi.

Ulteriormente, se uno dei problemi principali – se non il principale – delle piattaforme è la predisposizione a rifuggire (se non un vero e proprio atteggiamento riluttante) ad ogni tentativo di regolazione eteronoma – tanto che la sede stabilita in Paesi extra-UE aveva, in origine, garantito una tendenziale immunità di azione –, tale *aterritorialità* viene temperata dall'ambito di applicazione transfrontaliero della regolazione, estesa tanto ai rapporti interni tra Stati quanto alle relazioni con i Paesi terzi.

Sia l'art. 1, par. 2, *DMA*, che l'art. 2, par. 1, *DSA* conferiscono efficacia alle rispettive discipline a prescindere dalla tipologia dei servizi ed «indipendentemente» dal luogo in cui opera il *gatekeeper* o il prestatore di servizi – dentro o fuori l'Unione europea – talvolta ancorando la disciplina all'esistenza di uno stabilimento in Europa, talaltra attribuendo rilevanza alla destinazione delle attività svolte (artt. 2, 3, *DSA*).

Ricondurre la giurisdizione *dentro* l'Unione europea, anche coinvolgendo le autorità giurisdizionali ed amministrative nazionali (artt. 2, par. 6, *DSA*, art. 6, par. 4, *DSA*, art. art. 18, par. 1, *DSA* art. 51 *DSA*,

[93] Un problema simile si era posto per i *Large Language Models* come *ChatGPT*. Secondo Hacker, Engele Mauer, *Regulating ChatGPT and other Large Generative AI Models*, si tratta di tipologie di servizi non attratte nella disciplina del *DSA* poiché le informazioni generate sono scambiate direttamente con l'utilizzatore e non si tratta di un motore di ricerca. Tali algoritmi generativi sono oggi disciplinati dall'*AI ACT*. Sul rapporto tra *DSA* e *AI Act*, cfr. le prime riflessioni di S. TOMMASI, *Digital Services Act e Artificial Intelligence Act: tentativi di futuro da armonizzare*, in *Persona e Mercato*, 2, 2023, 279 ss.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

art. 23, par. 10, *DMA*)[94] è una scelta condivisibile poiché elimina due rischi: in primo luogo, che la competenza sia attratta da Paesi illiberali dove il livello di tutela dell'utente è minimo e dove i diritti per alcuni soggetti vulnerabili (come le donne) non sono poi così affermati come nelle democrazie occidentali; in secondo luogo, che la protezione di beni costituzionalmente garantiti sia almeno equivalente[95] e comunque superiore rispetto a quella apprestata dalle Costituzioni degli Stati membri che già si prendono cura del *volto fragile* di minori, donne e disabili.

Ciononostante, non può non rilevarsi come l'assunta esclusività di giurisdizione non faccia i conti con le norme di diritto internazionale che trattano anche delle *Alternative Despute Resolution*: l'ambito oggettivo molto ampio del foro competente convive solo nel *DSA* con la clausola di salvaguardia ex art. 1, par. 4, lett. h).

Sarà ovviamente l'interpretazione più o meno restrittiva del Regolamento Bruxelles I-*bis* a sciogliere i conflitti di giurisdizione concernenti, in esempio, le clausole *anti-steering* (artt. 4 *DMA*), i *dark patterns* (art. 25 *DSA*), le *pubblicità nascoste* (art. 25 *DSA*), le campagne di disinformazione o discriminatorie (Considerando 69 *DSA*) che nel digitale colpiscono, sì, tutti i *naviganti*, ma certamente molto più le categorie *speciali*.

Si tratta, a ben vedere, di un aspetto alquanto oscuro di una regolazione che potrebbe subire inevitabili restrizioni laddove, l'omessa riserva di giurisdizione a favore dell'Unione europea, in deroga alle clausole che ammettono la designazione di un Tribunale di un Paese terzo, potrebbe condurre ad un processo incardinato in un Paese nel quale lo spettro di tutela fondato sulla condizione di vulnerabilità del soggetto leso è meno ampio di quello previsto dal *pacchetto digitale*.

5. Conclusioni

La prevalenza di profili economico-negoziali del modello europeo non equivale ad una sottovalutazione della vulnerabilità della persona, sebbene la regolazione si presenti per lo più immatura rispetto ad ulteriori categorie di utenti fragili che, al pari dei minori, pretendono una maggior tutela a fronte degli ordinari tranelli del *web*.

La condizione ontologica vulnerabile aggravata dallo stato di minorazione fisica, psichica o di età, ovvero dalla condizione di sesso o di religione, avrebbe dovuto suggerire una riflessione maggiore per un'analisi di impatto della regolazione più coerente con i complessivi «indicatori qualitativi e quantitativi delle situazioni di discriminazione, subordinazione, dominazione, violenza»[96], in modo da riservare loro uno spazio più ampio nel *corpus* normativo del *Digital Markets Act* e del *Digital Services Act*.

[94] R. Sabia, *L'*enforcement *pubblico del* Digital Services Act *tra Stati membri e Commissione europea: implementazione, monitoraggio e sanzioni*, in *MediaLaws MediaLaws – Rivista del diritto dei media*, 2, 2023; I. Castellucci, F. Coppola, *Il sistema sanzionatorio decentrato del DSA: dinamica dell'apparato istituzionale*, in *Diritto di internet*, 1, 2023.

[95] Corte costituzionale, sentenza n. 349/2007. In quella circostanza l'equivalenza era posta in favore della Costituzione italiana, ma è vero anche l'inverso: laddove le norme convenzionali garantiscono un livello di tutela dei diritti più alto, è alla Convenzione che bisogna affidarsi.

[96] S. Zullo, *Lo spazio sociale della vulnerabilità tra «pretese e di giustizia» e «pretese di diritto». Alcune considerazioni critiche*, in *Politica del diritto*, 3, 2016, 477.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Lascia perplessi lo iato tra le aspettative di tutela espresse nei considerando riservate alle persone vulnerabili e la specificazione della tutela esplicitata nel corpo normativo in ragione delle singole categorie.

Dopotutto, se la scelta di riservare una protezione speciale ai minori origina dalla consapevolezza che laddove la vulnerabilità è maggiore, tanto più adeguato alla situazione deve essere il livello di tutela, non si comprende il motivo dell'omessa estensione soggettiva espressa anche a chi si trova in una condizione di maggiore vulnerabilità psico-fisica in quanto disabile, anziano, donna.

Probabilmente, l'atteggiamento non troppo tipizzante del legislatore europeo è dettato dalla continua mutevolezza del progresso tecnologico che obbliga il diritto a rincorrerlo e rende la regolazione velocemente anacronistica.

Il legislatore europeo ha preferito mantenere un profilo *soft* utilizzando una tecnica normativa più blanda, verosimilmente per non escludere *pro-futuro* ulteriori profili soggettivi che, altrimenti, sarebbero stati estromessi dal modello europeo di garanzia.

L'impostazione *soft*, difensiva e procedurale del modello europeo è probabilmente l'unico modo per governare, al momento, una pluralità di situazioni soggettive che rischierebbero, se tassativamente predeterminate, di ridurre l'area di tutela prevista dalla regolazione che sembra essere pienamente consapevole della circostanza per cui, in questo frastagliato percorso ad ostacoli, non si fa in tempo a disciplinare un aspetto che immediatamente emergono altre questioni da dover regolare.

E però, a questo punto, non ci si può non domandare il perché si è deciso di non estendere la disciplina concernente i minori ad altre *categorie speciali*, il cui livello di vulnerabilità, a prescindere dalla condizione di minorità, è egualmente alto (o più alto), e come tale pretende eguale (o maggiore) attenzione.

Per il futuro, pur con tutte le incognite concernenti l'adattamento costante della regolazione europea al *modus operandi* delle piattaforme digitali, e in attesa della costruzione giurisprudenziale della protezione della persona nel mondo virtuale – soprattutto con particolare riferimento alle categorie vulnerabili – il piano della Commissione UE nel 2023 (*Strategia dell'UE per guidare il web 4.0 e i mondi virtuali: muoversi in anticipo verso la nuova transizione tecnologica*)[97] prova a definire quali saranno i limiti e i pilastri strutturali dello sviluppo di mondi virtuali conformi ai valori fondamentali stabiliti nell'Unione europea[98], assicurando così la sicurezza e la garanzia dei diritti individuali di tutti, certo, ma soprattutto di quei profili di vulnerabilità consustanziali al nucleo duro riconosciuto dalle Carte.

Proprio la Costituzione italiana e la CEDU prescrivono la precedenza della persona rispetto al potere dello Stato.

[97] Il metaverso potrebbe essere la prossima frontiera sulla quale qualcuno ha già messo in luce ulteriori rischi, come R. Bifulco, *Riverberi costituzionali del Metaverso*, in *MediaLaws MediaLaws – Rivista del diritto dei media*, 3, 2023, G. Cerrina Feroni, *Il metaverso tra problemi epistemologici, etici e giuridici*, in *MediaLaws – Rivista del diritto dei media*, 1, 2023, A. Randazzo, *Prime notazioni sulle principali questioni in tema di Metaverso e Costituzione*, in *Dirittifondamentali.it*, 2, 2024, e, se si vuole, L. Di Majo, *L'art. 2 della Costituzione e il Metaverso*, in *MediaLaws – Rivista del diritto dei media*, 1, 2023.

[98] Così anche F. Pizzetti, *Con AI verso la società digitale*. In *Federalismi.it*, 23, 2023, 7 («L'evoluzione delle nostre società ci porterà a capire sempre meglio come reagire e come provvedere a conciliare il nostro passato col nostro futuro, a partire proprio dai diritti fondamentali e dalla loro tutela»).

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Dinanzi al dilagare di forme di sovranismo digitale, il potere privato – nella misura in cui questo assume le vesti di delegatario di fatto di attribuzioni di prerogativa statuale[99] – non può tradire quei principi – espressi a chiare lettere anche nella Convenzione di Oviedo – che sanciscono la prevalenza dell'essere umano sulla tecnica e sul progresso, ossia la centralità della persona umana rispetto al potere come nucleo centrale dell'umanesimo costituzionale.

Il che significa collocare ancora la vulnerabilità dell'utente al centro di una regolazione difensiva, certo, ma non limitata ai dati personali, estesa ai valori democratici (art. 14 *DSA*), con tutte le loro complessità, con il loro carattere evolutivo, tenendo ben salda l'idea che su alcuni valori la negoziazione non è ammessa, in particolare dove il potere (qualunque tipo di potere) fa i conti con una situazione fuori dall'ordinario.

Per aversi una svolta verso un costituzionalismo digitale non solo procedurale, ma anche sostanziale non può che invocarsi l'efficacia non solo verticale, ma anche orizzontale dei diritti fondamentali, il cui rispetto costituisce pertanto limite non soltanto all'esercizio del potere pubblico, ma anche del potere privato.

In quest'ottica andrebbe letto il nuovo *corpus* regolamentare europeo[100] che rintraccia nel rispetto dei principi informatori del *volto costituzionale europeo* non soltanto la condizione di legittimità del progetto stesso di sovranità digitale europea, ma anche il limite all'autonomia statutaria delle piattaforme e degli algoritmi da esse implementati[101].

Pur avvolti nel *velo di oscurità* di un futuro per alcuni inquietante, l'aspetto più importante dovrà, in ogni caso, innalzare l'attenzione sulla tutela dei soggetti fragili, individuando un livello minimo di garanzie sostanziali e procedurali riservate soprattutto alle categorie più vulnerabili che meritano una sorte diversa da quella «apocalittica»[102], non sempre a torto evocata.

[99] Sulla funzione *para-costituzionale* delle piattaforme che gradualmente assorbono le competenze statali, affiancando ad attività tipiche della dimensione privatistica, interventi di natura pubblicistica che incidono sui diritti fondamentali si veda M. Bassini, *Fundamental rights and private enforcement in the digital age*, in *European Law Journal*, 2, 2019, 182-197; si veda anche M. Betzu, *op. cit.*, il quale sembra deporre per una formulazione improntata al riconoscimento di diritti e non solo al divieto di interferenze.

[100] Cfr., da ultimo, S. Calzolaio, A. Iannuzzi, E. Longo, M. Orofino, F. Pizzetti, *La regolazione europea della società digitale*, Torino, 2024.

[101] La posizione in merito all'efficacia orizzontale della Carta Europea dei diritti Fondamentali e per suo tramite della CEDU, con specifico focus sugli artt. 14 e 16 *DSA*, è condivisa da I. de Vivo, *Il potere d'opinione delle piattaforme-online*, cit.

[102] E. De Martino, *La fine del mondo. Contributo all'analisi delle apocalissi culturali*, Torino, 2016.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Cybersicurezza e Intelligenza Artificiale. Un'analisi critica

Raffaella Brighi*

CYBERSECURITY AND ARTIFICIAL INTELLIGENCE. A CRITICAL ANALYSIS

ABSTRACT: The increasing complexity of cyber threats calls for the continuous evolution of cybersecurity strategies. This contribution explores the role of artificial intelligence (AI) in cyber- protection and defence, in particular, taking into account the revolution of cybersecurity practices brought about by advanced techniques such as machine learning and deep learning. The analysis of European and Italian strategies highlights the importance of an integrated approach involving technology, regulation and stakeholder cooperation. The risks related to the use of AI, including new vulnerabilities and potential ethical and social implications, are also discussed, with a view to analysing solutions for a more secure and resilient digital future.

KEYWORDS: Cybersecurity; Resilience; Artificial intelligence; Risk; EU Law.

ABSTRACT: La crescente complessità delle minacce informatiche impone un'evoluzione continua delle strategie di cybersicurezza. Il contributo esplora il ruolo dell'Intelligenza Artificiale (IA) nella protezione e nella difesa cibernetica, evidenziando come tecniche avanzate quali il *machine learning* e il *deep learning* stiano rivoluzionando il campo. Analizzando le strategie europee e italiane, si evidenzia l'importanza di un approccio integrato che coinvolga tecnologia, normativa e cooperazione tra gli stakeholder. Vengono inoltre discussi i rischi legati all'uso dell'IA, incluse le nuove vulnerabilità e le potenziali implicazioni etiche e sociali, analizzando soluzioni per un futuro digitale più sicuro e resiliente.

PAROLE CHIAVE: Cybersicurezza; Resilienza; Intelligenza Artificiale; Rischio; EU Law.

* *Professoressa Associata di informatica giuridica, Università di Bologna. Mail. Raffaella.Brighi@unibo.it.* *La ricerca è stata svolta nell'ambito del progetto PNRR "Partenariato Esteso" SERICS (PE00000014) – EcoCyber, Spoke 8, Finanziato dall'Unione europea – Next Generation EU ed anche nell'ambito del progetto ERC Computable Law ("CompuLaw") - Grant Agreement 833647. Contributo sottoposto a doppio referaggio anonimo.*

Downloaded from www.biodiritto.org.
ISSN 2284-4503

1. Insicurezza informatica. Minacce, vulnerabilità e nuovi rischi

La cybersicurezza rappresenta una sfida di primaria importanza nella nostra società digitale. Il costante aumento delle minacce e degli incidenti informatici, insieme alla sempre maggiore area di esposizione, ha spinto governi e istituzioni a promuovere strategie di resilienza e misure di protezione avanzate che comprendono, oltre alla tecnologia, interventi normativi, politici, economici e sociali[1].

L'Unione Europea ha riconosciuto l'importanza della cybersicurezza vedendola come elemento abilitante alla trasformazione digitale. Questo si è tradotto nell'emanazione di tre diverse *Strategie* che, sin dal 2013, promuovono un approccio di tipo globale, basato sulla cooperazione internazionale, la condivisione di informazioni e la redistribuzione di responsabilità tra settore pubblico e privato[2]. In questo quadro, sono stati adottati o proposti diversi atti giuridici che lungo tre macroaree di intervento – la resilienza, il contrasto al cybercrimine, la cyberdifesa e la cyberdiplomazia – definiscono un nuovo assetto normativo in materia di *cybersecurity*[3]. In Italia, la creazione dell'Agenzia per la Cybersicurezza Nazionale (ACN) nel 2021, all'interno del Piano Nazionale di Ripresa e Resilienza (PNRR), rappresenta un passo significativo verso un sistema di sicurezza più coordinato e robusto[4].

Le minacce informatiche colpiscono un'ampia gamma di soggetti, dai privati cittadini alle grandi aziende, fino agli enti pubblici e le istituzioni governative considerato che ogni aspetto della vita quotidiana – dai servizi pubblici all'istruzione, dal lavoro all'economia e ai processi democratici – dipende da reti, sistemi e tecnologie informatiche in costante evoluzione. Se da un lato l'irreversibile digitalizzazione di attività e servizi rende evidente l'importanza cruciale della sicurezza informatica come ele-

[1] *Ex multis*, T.F. Giupponi, *Il governo nazionale della cybersicurezza*, in *Quaderni Costituzionali*, 2, 2024; R. Ursi (a cura di), *La sicurezza nel cyberspazio*, Milano, 2023; E.C. Raffiotta, *Cybersecurity Regulation in the European Union and the Issues of Constitutional Law*, in *Rivista AIC*, 4, 2022; F. Casarosa, G. Comandé, *Aspettando la NIS2: ovvero il diritto privato della Cybersecurity*, in *Il Diritto dell'informazione e dell'Informatica*, XL, 1, 2024; S. Pietropaoli, *Cybersecurity in Informatica criminale. Diritto e sicurezza nell'era digitale*, Torino, 2023, 99-114.

[2] Commissione europea e alto rappresentante dell'UE per gli affari esteri e la politica di sicurezza JOIN(2013) 1 final; JOIN(2017) 450 final; JOIN(2020) 18 final.

[3] Un primo ambito del nuovo assetto riguarda il rafforzamento della sicurezza delle reti e dei sistemi informativi per incrementare la *cyber resilienza* nei settori essenziali per l'economia e la società, sia pubblici che privati, con l'emanazione della Direttiva NIS (*Network and Information Security* – Direttiva (UE) 2016/1148) e la sua revisione, la Direttiva NIS2 (Direttiva (UE) 2022/2555), la Direttiva CER (*Critical entities resilience directive*, Direttiva (UE) 2022/2557) e il regolamento DORA (*Digital operational resilience act* - Regolamento (UE) 2022/2554). Un secondo ambito riguarda la creazione di un quadro europeo di certificazione della cybersicurezza, tramite il Cybersecurity Act (Regolamento (UE) 2019/881), volto a garantire alti standard di sicurezza per prodotti, servizi e processi ICT con un approccio di sicurezza by design. La promozione della sicurezza sin dalla fase di progettazione è ulteriormente concretizzata nella proposta di nuove norme orizzontali per i prodotti con elementi digitali, attraverso il Cyber Resilience Act, approvato nel marzo 2024. Infine, un terzo ambito prevede la creazione di un 'ciberscudo europeo', attraverso prima la specificazione e il potenziamento del ruolo dell'ENISA, poi con la proposta del c.d. Cyber Solidarity Act (18 aprile 2023), che attraverso quadri di cooperazione operativa già esistenti (EU-CyCLONe e la rete di CSIRTs), intende rafforzare la capacità della UE di prepararsi e gestire gli attacchi su larga scala.

[4] L'Agenzia, che è il cardine della infrastruttura italiana di cybersecurity, è stata istituita con il d.l. 82/2021 e organizzata con il d.p.c.m. 223/2021.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

mento fondamentale per la trasformazione economica e sociale, d'altro canto siamo ancora lontani dal raggiungere un livello di protezione adeguato[5].

In aggiunta a rischi noti quali disuguaglianze, discriminazioni, controllo sociale e concentrazione del potere digitale, le tensioni geopolitiche e i conflitti in corso hanno reso evidente che il rischio cibernetico è globale e di primaria importanza. Attacchi informatici sempre più sofisticati e aggressivi possono veicolare ulteriori gravi pericoli, come il terrorismo internazionale, conflitti tra Stati, attività economiche illecite, campagne di disinformazione. Tali minacce, in grado di colpire «sia gli interessi dello Stato che la fruibilità dei diritti dei soggetti di un ordinamento»[6], hanno ridefinito natura e confini dello stesso concetto di sicurezza pubblica[7]. In questo contesto, gli Stati sempre di più trattano il tema della cybersicurezza come una questione intrinsecamente connessa alla tutela della sicurezza nazionale.

In tal senso, è emblematico il caso dell'Italia che nel 2019 si dota di un autonomo quadro normativo in materia – nelle more del processo di revisione della direttiva UE 2016/1148 (cd. direttiva NIS) – al fine di assicurare un livello elevato di sicurezza delle reti, dei sistemi informativi e dei servizi informatici di soggetti pubblici e privati, da cui dipende l'esercizio di una funzione o la fornitura di un servizio essenziale per lo Stato e dal cui malfunzionamento possa derivare un pregiudizio per la sicurezza nazionale[8]. Peraltro, il nostro legislatore, con il decreto di istituzione di ACN, nel definire il perimetro della cybersicurezza ha richiamato tra gli obiettivi anche la «tutela della sicurezza nazionale e dell'interesse nazionale nello spazio cibernetico»[9]. In definitiva, si tratta di una funzione di sicurezza che interessa complessivamente l'ordinamento statale e le sue componenti, ossia le imprese e i singoli cittadini, dove la complessità della materia e la necessità di coordinamento tra le istituzioni coinvolte e le imprese che operano nel settore, richiamano un intervento ampio, multilivello e trasversale[10].

Il report dell'ENISA (*European Union Agency for Cybersecurity*) – *Foresight Cybersecurity Threats for 2030* – nell'identificare le dieci principali minacce che peseranno sul cyberspazio nel 2030[11], sottolinea l'importanza della implementazione di politiche di mitigazione dei rischi per aumentare la sicu-

[5] Il Rapporto Clusit 2024 sulla sicurezza ICT in Italia evidenzia che il 64% degli incidenti hanno come causa azioni "maldestre", degli utenti o del personale ICT. Malware, Vulnerabilità, Phishing e Account Cracking sono indice di carenze nella cyber igiene degli utenti che restano vulnerabili alle tecniche di più comuni di ingegneria sociale.

[6] G. De Vergottini, *Una rilettura del concetto di sicurezza nell'era digitale e della emergenza normalizzata*, In *Rivista AIC*, 4, 76.

[7] Sul punto si vedano T.F. Giupponi, *Sicurezza e potere*, in *Enciclopedia del diritto*, *I tematici*, V, 1146 ss.; G. De Vergottini, *op cit.*, 65 ss.

[8] Decreto-legge n. 105 del 2019, convertito con modificazioni dalla legge 4 novembre 2019, n. 133.

[9] F. Serini, *La nuova architettura della cybersicurezza nazionale: note a una prima lettura del decreto-legge n.82 del 2021*, in *Federalismi.it*, 12, 241 ss.

[10] In argomento, si vedano T.F. Giupponi, *Il governo nazionale della cybersicurezza,* cit., 180-181; R. Ursi, *La sicurezza cibernetica come funzione pubblica*, in R. Ursi (a cura di), cit., 13 ss.

[11] Tra queste la compromissione della supply chain a causa della vulnerabilità dei molteplici componenti *hardware* e *software* integrati nei nuovi prodotti digitali, la manomissione dei dispositivi cyberfisici per lo sfruttamento di dati comportamentali e sensibili degli individui, l'abuso di sistemi di intelligenza artificiale attraverso la manipolazione intenzionale degli algoritmi e dei dati di addestramento.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

rezza e la resilienza delle infrastrutture su cui si fonderanno le città del futuro. Esiste una asimmetria evidente tra il lato della difesa e quello dell'attacco. Il compito della difesa è notevolmente più complesso e richiede interventi su molteplici livelli. Mentre gli attaccanti possono sfruttare vulnerabilità specifiche e nuove tecniche, i difensori devono proteggere una vasta gamma di sistemi, applicazioni e dati, in modo proattivo, prevedendo potenziali minacce e adottando misure preventive.

Tra le principali innovazioni tecnologiche, l'intelligenza artificiale (IA) sta emergendo come una risorsa chiave nella lotta contro le minacce informatiche, rivoluzionando anche il campo della cybersicurezza. Tecniche avanzate come il *machine learning* (ML) e il *deep learning* (DL) migliorano per rapidità ed efficacia la capacità di analisi dei dati di sicurezza e di decisione autonoma. Queste tecnologie potenzieranno la capacità difensiva con nuovi metodi per prevenire, rilevare e rispondere agli attacchi informatici.

Tuttavia, in ragione della natura *dual use* della IA, comune a molte tecnologie, la stessa è annoverata anche tra i pericoli emergenti per la sicurezza informatica. L'IA infatti amplifica le capacità di attacco, fornendo strumenti agli aggressori che arricchiscono il panorama di nuove minacce in termini sia quantitativi sia qualitativi. Inoltre, anche questi i sistemi non sono esenti da vulnerabilità intrinseche che possono portare ad errori o essere sfruttate in molti modi, tra cui attacchi in grado di avvelenare i dati di addestramento (*data poising*) o raggirare il *prompt* delle IA generative[12].

Questo incremento di complessità crea ulteriori vulnerabilità e potenzia le azioni malevole, rendendo la sicurezza informatica una sfida sempre più articolata.

Per sfruttare appieno l'applicazione dell'intelligenza artificiale nella cybersicurezza, gli attori, governativi o privati, devono padroneggiare gli strumenti e saperne comprendere e affrontare i rischi, per promuovere un utilizzo responsabile e sicuro. Le specificità di natura tecnica dell'AI, quali opacità e *bias* algoritmici, unite alla raccolta e analisi massiva di dati personali e comportamentali, pongono criticità rispetto all'automazione dei processi di difesa informatica che devono essere affrontati sul piano tecnico, normativo e sociale.

In tale contesto, questo contributo intende esplorare l'impatto dell'IA nel miglioramento della cybersicurezza, sia a livello individuale che collettivo. La prossima sezione introduce i paradigmi metodologici della sicurezza informatica, la sezione 3 analizza gli scenari in cui l'IA viene impiegata per automatizzare e rendere più efficaci i controlli di sicurezza, la sezione 4 si soffermerà sul *dual use* della tecnologia in esame e sulle nuove vulnerabilità per avanzare alcune riflessioni sul futuro della AI nella cybersicurezza.

2. Fondamenti tecnico-giuridici della cybersicurezza

L'espressione *cybersecurity* è definita dal Cybersecurity Act (Regolamento (UE) 2019/881) come l'«insieme delle attività necessarie per proteggere la rete e i sistemi informativi, gli utenti di tali sistemi e altre persone interessate dalle minacce informatiche». Ad un alto livello di astrazione essa si estrinseca nello studio, progettazione e implementazione di strategie volte a proteggere la dimensione digitale da un pericolo (o dalla minaccia di un pericolo) di natura volontaria o accidentale[13].

[12] Infra § 4.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Le attività riguardano limitatamente l'adozione di strumenti tecnologici quanto, piuttosto, la definizione di politiche (norme, regole amministrative e procedure organizzative), la predisposizione di meccanismi di controllo e la promozione di comportamenti individuali corretti. Tra queste strategie rientrano procedure di autenticazione, gestione degli accessi, analisi dei rischi, aggiornamento dei sistemi, rilevazione e reazione ad incidenti o attacchi, recupero delle componenti oggetti di attacco, addestramento e formazione del personale.

La progettazione efficace della sicurezza - attraverso procedure, controlli, comportamenti e tecnologie - è guidata dal *controllo del rischio* [14]. La regolazione e la gestione del rischio permea tutta la disciplina della sicurezza informatica e è alla base della più recente legislazione della UE in materia [15]. Non solo rischi per reti e sistemi informativi, ma anche rischi sociali, rischi per l'integrità fisica, rischi per i diritti e le libertà fondamentali: la normativa europea ha in sostanza ampliato il concetto di cybersecurity per includere la governance di un'ampia gamma di rischi, senza concettualizzarlo in modo impropriamente limitato. Il grado di esposizione al rischio, la probabilità che si verifichino incidenti e lo loro gravità sono i parametri in base ai quali determinare l'adozione di misure di sicurezza informatica conformi allo stato dell'arte e agli standard europei (ETSI, CEN) e internazionali (ISO/IEC), gli obblighi di segnalazione degli incidenti e l'adesione a quadri di certificazione della conformità. Le aree in cui si articolano le attività sono sostanzialmente tre: (i) realizzare sistemi robusti in grado di resistere agli attacchi, (ii) progettare metodi per il rilevamento di minacce ed anomalie al fine di garantire la resilienza dei sistemi; (iii) definire le risposte agli attacchi per ripristinare sistemi e servizi [16]. La robustezza dei sistemi è essenziale per mitigare l'impatto degli incidenti, consente alle infrastrutture e ai servizi nazionali critici di funzionare e ai cittadini di fare affidamento su tecnologie sicure. La resilienza e la risposta sono invece il lato attivo della sicurezza informatica che, a questo scopo, si avvale di forme di monitoraggio della rete per identificare attacchi e fonti di attacchi e reagire alle minacce. Ciascuna area comprende una vasta gamma di soluzioni tecniche e misure organizzative.

Alcuni *framework* di riferimento per il settore forniscono un insieme di linee guida standard e *best practice*, richiamate anche dal quadro normativo in materia, che aiutano le organizzazioni a uniformare le pratiche di sicurezza e facilitano comunicazione e cooperazione. Tra tutti, è rilevante per le nostre analisi il modello NIST[17] perché, oltre a essere molto noto nella comunità scientifica, è alla base del *Framework Nazionale per la Cybersecurity e la Protezione dei Dati*[18] e della tassonomia

[13] Per un'analisi del concetto di cybersecurity, V. PAPAKONSTANTINOU, *Cybersecurity as Praxis and as a State: The EU Law Path towards Acknowledgement of a New Right to Cybersecurity*?, *44 Computer Law & Security Review*, 2022; M. TADDEO, *Is Cybersecurity a Public Good?*, in *Minds and Machines*, 29, 3, 2019, 349-354; R. BRIGHI, P.G. CHIARA, *La cybersecurity come bene pubblico: alcune riflessioni normative a partire dai recenti sviluppi nel diritto UE*, in *Federalismi*, 2021, 21, 18-42; G. ZICCARDI, *La Cybersecurity nel quadro tecnologico (e politico) attuale*, in *Tecnologia e Diritto*, III, Milano, 2019, 207-210.

[14] Norme tecniche internazionali (es. ISO 31000) definiscono il rischio come l'effetto dell'incertezza sugli obiettivi di sicurezza del sistema e stabiliscono metodologie e metriche per valutazione, analisi e gestione del rischio.

[15] A. MANTELERO et al., *The Common EU Approach to Personal Data and Cybersecurity Regulation*, in *International Journal of Law and Information Technology* 4,28, 2021, 297-328; P.G. CHIARA, F. GALLI, *Normative Considerations on Impact Assessments in EU Digital Policy*, in *Media Law*, 1, 2024, 86-105.

[16] G. D'ANGELO, G. GIACOMELLO, *Cybersicurezza. Che cos'è e come funziona*, Bologna, 2023.

[17] NIST (National Institute of Standards and Technology) Framework, https://www.nist.gov/cyberframework.

[18] CINI, Cyber Security National Lab, 2019, https://www.cybersecurityframework.it.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

adottata dal decreto attuativo del Perimetro Nazionale di Sicurezza Cibernetica (PSNC), dpcm 14 aprile 2021, n. 81 in materia di notifiche degli incidenti e misure di sicurezza.
Data l'eterogeneità delle soluzioni di cybersicurezza e la molteplicità delle applicazioni di IA che stanno emergendo, introdurre una tassonomia uniforme, accettata e consolidata è utile per tracciare una visione sistematica di opportunità e rischi dell'applicazione della IA alle strategie di cybersicurezza.
La tassonomia del NIST definisce cinque funzioni chiave per il processo di gestione della cybersicurezza nel tempo: identificazione, protezione, rilevamento, risposta e ripristino. La fase di *identificazione* si concentra sull'individuazione delle criticità e dei rischi associati a sistemi, dati, asset e persone, fornendo le basi per le successive fasi di gestione. La *protezione* si occupa di implementare controlli adeguati a prevenire o contenere l'impatto di eventi negativi e attiene alla robustezza del sistema. La *rilevazione* è dedicata all'identificazione tempestiva di incidenti di sicurezza attraverso il monitoraggio continuo e l'analisi delle anomalie. La *risposta* prevede le attività necessarie per intervenire quando un incidente viene rilevato, con l'obiettivo di contenerne l'impatto. Infine, il *ripristino* riguarda la gestione dei piani per recuperare rapidamente la funzionalità dei processi e dei servizi colpiti da un incidente, garantendo la resilienza delle infrastrutture. Per ogni funzione chiave si sono sviluppate e sono riprese metodologie consolidate, strumenti tecnologici, raccomandazioni e strategie organizzative che includono, qualora il contesto lo richieda, anche i vincoli legali.

3. Applicazioni dell'IA a supporto della cybersicurezza

L'aumento del numero, della portata e dell'impatto degli attacchi informatici necessita di una difesa dinamica, proattiva e adattativa, supportata da valutazioni in tempo reale attraverso il monitoraggio continuo e l'analisi dei dati.
La letteratura tecnico-scientifica è concorde nel rilevare che l'intelligenza artificiale viene sempre più integrata nel tessuto della cybersecurity e utilizzata in una varietà di scenari applicativi[19]. Per risolvere i problemi di cybersecurity di oggi sono impiegate diverse tecniche di IA, in particolare l'apprendimento automatico (supervisionato, per rinforzo e non-supervisionato), algoritmi di elaborazione del linguaggio naturale (NLP, *Natural Language Processing*), sistemi di rappresentazione della conoscenza, sistemi per la descrizione e modellazione del ragionamento, sistemi di ragionamento basati sui casi[20].
Esiste un'ampia gamma di tecniche di difesa che possono essere abilitate dall'IA con il potenziale di fornire prestazioni soddisfacenti a basso costo e in tempo reale per la sicurezza delle reti e dei dati, la protezione degli *endpoint*, l'affidabilità degli accessi, il rilevamento, l'identificazione e la mitigazione dei cyberattacchi. In particolare, alcuni studi esplorano le applicazioni della IA e le tendenze di ricerca

[19] In particolare si veda M. Malatji, A. Tolah, *Artificial intelligence (AI) cybersecurity dimensions: a comprehensive framework for understanding adversarial and offensive AI*, in *AI Ethics* (2024); I.H. Sarker, et al., *AI-Driven Cybersecurity: An Overview, Security Intelligence Modeling and Research Directions*, in *SN Computer Science*, 2, 173 (2021); J. Edwards; W. Griffin, *Artificial Intelligence in Cybersecurity* in *The Cybersecurity Guide to Governance, Risk, and Compliance*, Wiley, 2024, 497-510; R. Kaur, D. Gabrijelčič, T. Klobučar, *Artificial intelligence for cybersecurity: Literature review and future research directions*, in *Information Fusion*, 97, 2023,101804.
[20] Per la classificazione dei sistemi di intelligenza artificiale cfr. S. Russel, P. Norvig, *Artificial Intelligence: A Modern Approach*, Prentice Hall Press, 2009; G. Sartor, *L'intelligenza artificiale e il diritto*, Torino, 2022.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

mappandole secondo la tassonomia del NIST, introdotta nella sezione 2, e forniscono una panoramica sistematica dello stato dell'arte in questo campo[21]. Secondo i dati raccolti la maggior parte delle applicazioni di IA che integrano i metodi di sicurezza convenzionali riguardano la fase di rilevamento, un numero minore si occupa dell'identificazione, a seguire protezione e risposta. Poche sono le ricerche e gli strumenti focalizzati sull'applicazione della IA nelle operazioni di ripristino dei sistemi.

A supporto dei processi decisionali nella fase di rilevamento, l'intelligenza artificiale migliora la comprensione delle minacce informatiche tramite l'estrazione automatica, la correlazione e la valutazione di informazioni da molteplici fonti eterogenee, quali database di vulnerabilità, social media, siti di notizie, rapporti sugli incidenti e dark web. Nei sistemi di rilevamento delle intrusioni (IDS), tecniche di ML e DL permettono di analizzare il traffico di rete, rilevare attività sospette, classificare gli eventi e distinguere tra vari tipi di attacchi[22]. Analogamente, l'IA rivela anomalie su sistemi informativi e dispositivi fornendo una visione chiara e dinamica dell'ambiente cyberfisico[23] e, nell'ambiente, dei comportamenti degli utenti interni all'organizzazione.

Con riferimento alla fase di *identificazione*, strumenti di IA gestiscono gli *asset* nelle reti estese, scoprendo e configurando automaticamente dispositivi, applicazioni e utenti del sistema attraverso tecniche di clustering e ML. Sistemi di riconoscimento biometrico fisico e comportamentale sfruttano l'IA per il controllo dell'identità; i modelli di comportamento d'uso, relativi all'interazione dell'utente con il proprio dispositivo e le statistiche delle interazioni con diverse applicazioni, consentono ad esempio di determinare se l'utente corrente sia lo stesso di quello precedentemente autenticato (autenticazione continua).

L'IA previene la perdita dei dati identificando e classificando informazioni in base a caratteristiche condivise, monitora l'attività degli utenti interni per rilevare comportamenti anomali rispetto ai modelli elaborati sulla base del pregresso, aiuta a bloccare le mail di spam e di phishing e con esse i potenziali pericoli, può scoprire *malware* emergenti che generano varianti per eludere gli approcci tradizionali basati su regole e può individuare contenuti digitali alterati (*deep fake*)[24]. La formazione e la sensibilizzazione adattiva offrono contenuti aggiornati e personalizzati, migliorando la consapevolezza e le competenze degli utenti.

La funzione di *risposta* nella cybersecurity è essenziale per gestire e contenere l'impatto degli eventi di sicurezza. L'intelligenza artificiale introduce miglioramenti significativi automatizzando processi complessi e riducendo il carico di lavoro degli analisti[25] con strumenti di gestione dinamica dei casi

[21] R. Kaur, D. Gabrijelčič, T. Klobučar, *op.cit.*

[22] A. Venturi, G. Apruzzese, M. Andreolini, M. Colajanni, M. Marchetti, *DReLAB - Deep REinforcement Learning Adversarial Botnet: A benchmark dataset for adversarial attacks against botnet Intrusion Detection Systems*, in *Data in brief*, 34, 2021, 1-12.

[23] G. Gori, L. Rinieri, A. Melis, A. Al Sadi, F Callegati, M. Prandini, *A Systematic Analysis of Security Metrics for Industrial Cyber–Physical Systems*, in Electronics S, 13(7), 2024, 1-17.

[24] L. Guarnera, O. Giudice, S. Battiato, *Fenomenologia dei Deepfake: aspetti teorici e operativi per la detection di volti umani "artificiali"*, in R. Brighi (a cura di), *Nuove questioni di informatica forense*, Roma, 2022.

[25] M. Ferrazzano, *L'intelligenza artificiale a servizio delle attività di informatica forense*, in *Ordines*, 2, 2023, 132-146; S. Costantini, G. De Gasperi, R. Olivieri, *Digital forensics and investigations meet artificial intelligence*, in *Annals of Mathematics and Artificial Intelligence*, 86/2019, 193-229.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

che registrano scenari di attacco e suggeriscono azioni di risposta appropriate basate sulle lezioni apprese dagli incidenti pregressi.

Nella fase di *ripristino*, l'intelligenza artificiale può automatizzare il recupero dei dati e dei sistemi e può supportare i processi di pianificazione della risposta futura con l'esame delle strategie esistenti e con l'analisi e aggregazione dei dati sugli incidenti e i registri di audit.

L'IA sta, dunque, diventando sempre più essenziale nello sviluppo di strumenti per perseguire gli obiettivi di sicurezza informatica e far fronte alle minacce emergenti. Le soluzioni tecnologiche qui descritte integrano la capacità di apprendimento automatico e profondo per elaborare, in modo più efficace e più veloce rispetto alle persone, grandi flussi di dati e ricavare informazioni che possono essere rilevanti in tutte le fasi della sicurezza informatica, dall'analisi del contesto per la progettazione delle misure di protezione fino al ripristino dei sistemi e servizi. La combinazione di Big Data e IA consente di automatizzare processi di decisione complessi, basati su numerosi fattori e criteri non esattamente predeterminati. Ciò può migliorare la qualità delle decisioni pubbliche e private, ma impone di riflettere sui rischi, collegati alle specificità tecniche dei sistemi di IA (tra cui *bias* algoritmici, opacità, scelta del *dataset*), che sono oggetto del dibattito dottrinale più recente[26].

Raccolta, conservazione e analisi massiva dei dati di sicurezza sono essenziali per lo sviluppo di strumenti di difesa cibernetica basati su IA. Dalla *threat intelligence* per scopi predittivi all'automazione totale dei processi di risposta, questi sistemi utilizzano grandi quantità di dati, tra cui dati personali e dati comportamentali degli utenti, che devono essere dati di qualità, dati recenti e dinamici, e provenire da più fonti per descrivere in modo completo l'ambiente. Se i dati utilizzati per addestrare i modelli di IA sono parziali o incompleti, i modelli stessi possono ereditare questi *bias*, portando a decisioni non accurate e potenzialmente discriminatorie[27]. Modelli di IA addestrati su una determinata area geografia o su uno specifico gruppo di utenti (categorie professionali, gruppi demografici, genere) porteranno a una protezione inadeguata, alla penalizzazione di certi gruppi, a discriminazioni e disuguaglianze[28].

La raccolta e l'analisi massiccia dei dati personali e comportamentali (attività *online*, comunicazioni, comportamenti di utilizzo, dati di localizzazione) rappresentano un rischio significativo per la privacy degli utenti e il diritto alla protezione dei dati personali, creando una tensione tra la necessità di raccogliere informazioni dettagliate per migliorare la sicurezza e i diritti degli individui. Il monitoraggio

[26] Alcuni riferimenti essenziali, F. Pasquale, *Le nuove leggi della robotica. Difendere la competenza umana nell'era dell'intelligenza artificiale*, Roma, Luiss University Press, 2021; U. Ruffolo (a cura di), *Intelligenza artificiale. Il diritto, i diritti, l'etica*, Torino, 2020; A. D'Aloia (a cura di), *Intelligenza artificiale e diritto: come regolare un mondo nuovo*, Milano, 2020; G. Alpa (a cura di), *L'intelligenza artificiale: il contesto giuridico*, Modena, 2021; P. Benanti, *Human in the loop. Decisioni umane e intelligenze artificiali*, Milano, 2022; A.; J. Searle, *Intelligenza artificiale e pensiero umano: filosofia per un tempo nuovo*, trad. it. A. Condello, Roma, 2023; S. Salardi, Intelligenza artificiale e semantica del cambiamento: una lettura critica, Torino, 2023; Th. Casadei, S. Pietropaoli, *Intelligenza artificiale: l'ultima sfida per il diritto?,* in Th. Casadei, S. Pietropaoli (a cura di), *Diritto e tecnologie informatiche*, Milano, 2024.

[27] *Inter alia* J. Kleinberg, J. Ludwig, S. Mullainathan, C. R. Sunstein, *Discrimination in the Age of Algorithms*, in *journal of Legal Analysis*, 10, 2018, 1-62; G. Lasagni, G. Contissa, G. Sartor, *Quando a decidere in materia penale sono (anche) algoritmi*, in Diritto di internet, 4, 619-634; V. Barone, *Le discriminazioni ai tempi dell'intelligenza artificiale: la questione algoritmi*, in *Diritto e tecnologie informatiche*, *op. cit.*

[28] F. De Simone, *Una nuova tipologia di misure di prevenzione: algoritmi, intelligenza artificiale e riconoscimento facciale*, in *Archivio Penale*, 2, 2023.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

continuo e pervasivo dei sistemi e degli utenti, per identificare minacce e anomalie, può trasformare le pratiche di cybersicurezza in strumenti di sorveglianza persistenti e pervasivi. Questo tipo di sorveglianza, che riguarda tanto gli ambiti lavorativi quanto i singoli e la collettività, può erodere le libertà civili e i diritti fondamentali, creando un ambiente in cui gli individui si sentono costantemente osservati e controllati[29].

L'opacità e la complessità dei modelli di IA rappresentano una criticità significativa anche per la governance della cybersicurezza nella misura in cui le soluzioni basate su IA non saranno in grado di giustificare i risultati (dal rilevamento al processo decisionale) e renderli comprensibili all' essere umano. Questo aspetto diventa infatti particolarmente rilevante quando le decisioni automatizzate hanno conseguenze gravi, come la determinazione di minacce, la risposta a incidenti di sicurezza o ancora l'attribuzione di un indice di rischio ai comportamenti degli utenti interni ai sistemi. La comprensione inoltre è strategica per consentire agli operatori, che sono sommersi da decine di migliaia di avvisi di sicurezza al giorno (la maggior parte dei quali falsi positivi), di valutare meglio le potenziali minacce e di ridurre la stanchezza da allarme[30]. Di conseguenza, anche nel settore della cybersicurezza, la sfida di rendere i modelli di IA spiegabili o interpretabili dagli utenti umani è di primaria importanza; conoscibilità e spiegabilità sono il presupposto per garantire che le decisioni automatizzate siano giuste ed equitative[31].

Questi rischi, assieme a vulnerabilità che saranno analizzate nella prossima sezione, sottolineano la necessità di adottare misure di mitigazione come l'uso di tecniche di IA spiegabili, l'implementazione di controlli rigorosi sui dati di addestramento e la combinazione di IA con la supervisione umana per garantire una difesa robusta contro le minacce informatiche.

A tale proposito, è opportuno osservare che i sistemi di IA descritti, indipendentemente dalle diverse funzioni di cybersecurity che implementano, rientrano nella categoria della IA specifica o ristretta[32], come tutte le applicazioni di IA oggi disponibili. Si tratta di strumenti limitati a un singolo compito o gruppo di operazioni, la cui autonomia è ad oggi alquanto ridotta e il controllo umano prevalente. Probabilmente ci saranno progressi nel grado di automazione e nella velocità dei processi, ma difficilmente l'automazione riguarderà l'intero processo di cybersicurezza. Sembra dunque che nessuna so-

[29] A.C. Amato Mangiameli, *Algoritmi e big data*, in *Rivista di Filosofia del Diritto*, 2019, VII, 1, 107-124; F. Lagioia, G. Sartor, *Profilazione e decisione algoritmica: dal mercato alla sfera pubblica*, in *Federalismi.it*, 11/2020, 88 ss., F. Pizzetti (a cura di), *Intelligenza artificiale, protezione dei dati personali e regolazione,* Torino, 2018; F. Faini, Intelligenza artificiale e regolazione giuridica: il ruolo del diritto nel rapporto tra uomo e macchina, in *Federalismi.it*, 2/2023. G. Ziccardi, *Tecnologie per il potere*, Raffaello Cortina, 2019.

[30] F. Charmet et al., *Explainable artificial intelligence for cybersecurity: a literature survey* in *Ann. Telecommun*. 77, 789–812 (2022); R. Guidotti et al., *A Survey Of Methods For Explaining Black Box Models*, in *ACM Computing Surveys*, LI, 93 (2018), 1-42.

[31] In argomento, M. Palmirani, *Interpretabilità, conoscibilità, spiegabilità dei processi decisionali automatizzati*, in *XXVI lezioni di Diritto dell'Intelligenza artificiale*, Giappichelli, Torino, 2020; Pagallo, *Algoritmi e conoscibilità*, in *Rivista di filosofia del diritto*, 1/2020; E. Longo, *I processi decisionali automatizzati e il diritto alla spiegazione*, in A. Pajno, F. Donati, A. Perrucci (a cura di), *Intelligenza artificiale e diritto: una rivoluzione?,* I, Bologna, 2022, 349 ss.; A. Andronico, Th. Casadei (a cura di), *Algoritmi ed esperienza giuridica*, in *Ars Interpretandi*, 1, 2021, 7-164.

[32] Per la differenza tra AI ristretta e generale si veda G. Sartor, *op cit.*, 16.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

luzione di IA sarà in grado di svolgere attività di riposta totalmente non supervisionate sia nella protezione del sistema (correzione automatica delle vulnerabilità) sia nella difesa informatica attiva[33].

4. Vulnerabilità e sicurezza dell'intelligenza artificiale

Come molte altre tecnologie, l'IA è a "doppio uso" il che significa che può essere impiegata per migliorare gli strumenti di contrasto alle minacce per la (*cyber*) sicurezza o anche per scopi dannosi e per ottenere vantaggi competitivi in seguito agli attacchi. Le caratteristiche di efficienza, scalabilità e adattabilità rendono l'IA interessante per diversi tipi di attori (statali e non) e sfruttabile per obiettivi difensivi (*Defensive AI*), offensivi o di altri tipo legati alla sicurezza[34]. Non deve sorprendere dunque che tra i pericoli emergenti per la cybersicurezza venga segnalato proprio l'uso e l'abuso della intelligenza artificiale.

Con l'impiego dell'IA saranno realizzati *cyber* attacchi sempre più sofisticati e complessi[35]. Come per la difesa, gli attacchi colpiscono tutti i livelli del cyberspazio: fisico (dispositivi e apparecchiature), logico (software, protocolli) e semantico (dati, informazioni). Sfruttando le medesime tecniche di cyber intelligence implementate nei sistemi di rilevamento delle minacce, gli aggressori sono in grado di preparare e attuare sofisticati attacchi di ingegneria sociale basati sulle informazioni che le potenziali vittime lasciano nella rete e di creare vettori personalizzati: link, siti web, e-mail persuasive, chat bot convincenti e contenuti manipolati (i *deep fake*) non facilmente identificabili come falsi. Inoltre, grazie all'apprendimento rinforzato le minacce informatiche saranno capaci di eludere il rilevamento, adattarsi ad ambienti mutevoli, scoprire vulnerabilità specifiche, propagarsi e persistere sui sistemi bersaglio. Un *malware* con componenti di IA può essere in grado di offuscare il suo funzionamento nel sistema infettato e rispondere in modo creativo e adattativo ai cambiamenti dell'ambiente e al comportamento degli utenti. La permanenza inosservata sul sistema target consentirà, inoltre, di trovare e classificare contenuti utili per l'esfiltrazione e di individuare nuovi punti di attacco. Gli aggressori potrebbero sfruttare i dati di addestramento per generare una *backdoor* nel sistema software di IA o utilizzare l'IA per determinare quale vulnerabilità vale la pena sfruttare. Scenari in cui l'AI diventa strumento per condurre attacchi cyber si classificano come *Offensive IA*.

I sistemi di IA, inoltre, possono essere vulnerabili a causa di debolezze intrinseche o di meccanismi interdipendenti, ancora non risolti o sconosciuti. Attacchi mirati possono sfruttare le vulnerabilità esistenti nelle librerie software open-source più diffuse (le comunità di IA sono particolarmente aperte in termini di trasferimento della conoscenza) oppure mettere in atto *reverse engineering* del modello addestrato sfruttando interfacce di interrogazione pubblicamente accessibili o ancora sfruttare il *prompt* per riuscire ad ottenere da IA generative informazioni sensibili o illegali. Altri attacchi sono in grado di "avvelenare" i dati di addestramento *(data poisoning)*, in questo caso, si presume che l'attaccante abbia accesso ai dati e sia in grado di alterarli e di introdurre manipolazioni in modo che il sistema, addestrato sui dati avvelenati, esegua elaborazioni o previsioni seguendo gli interessi

[33] M.E. Bonfanti, *Artificial intelligence and the offense–defense balance in cyber security*, in *Cyber Security Politics*, Routledge, 2022, 64-78. Questo tipo di risposta potrebbe essere anche indesiderabile per le conseguenze politiche, legali e strategiche che potrebbe generare.

[34] M. Bonfanti, *op. cit.*, 69 ss.

[35] ENISA, *Artificial Intelligence and Cybersecurity Research*, 2023. *Inter alia*, M. Malatji, A. Tolah, *op. cit.*

Downloaded from www.biodiritto.org.
ISSN 2284-4503

dell'attaccante. Gli *attacchi avversari*, invece, colpiscono le reti neurali profonde con input progettati dall'aggressore per essere classificati in modo errato e alterare di conseguenza la previsione del sistema di intelligenza artificiale. Se da un lato l'AI diventa sempre più indispensabile per la gestione delle minacce informatiche, la manipolazione dei sistemi software di AI può compromettere l'efficacia stessa dei sistemi di sicurezza. Tutti gli ultimi scenari descritti sono classificati come *Adversarial AI*.

È difficile peraltro fare una stima a medio-lungo termine del perimetro e dell'impatto delle vulnerabilità dei componenti di IA e questo vale anche per gli strumenti impiegati nell'ambito della difesa della (cyber)sicurezza. Si tratta di innovazioni emergenti, la ricerca e le applicazioni negli ultimi anni hanno compiuto progressi molto rapidi e quindi è ragionevole supporre che molte vulnerabilità siano sconosciute e i potenziali abusi debbano ancora essere esplorati. Le implicazioni per la cybersicurezza in questo campo sono ancora meno prevedibili che in altri contesti.

Oltre ad interrogarsi sui molti modi in cui l'IA può essere sfruttata per la protezione del cyber spazio o sulla sua capacità offensiva, è opportuno riflettere sulle buone pratiche di sicurezza informatica per garantire che le componenti di IA inserite nei sistemi siano integre, affidabili e disponibili.

La sicurezza informatica deve essere una priorità e un requisito per tutti i sistemi di IA, quindi anche per quelli progettati per il contesto della cybersecurity.

La progettazione della sicurezza per l'IA è più complessa rispetto ai tradizionali sistemi di ingegnerizzazione del software perché entrano in gioco molti fattori e le minacce non sono solo tecniche, legali o ambientali, ma anche sociali, come illustrato nella sezione precedente. Diventa auspicabile indirizzare l'uso delle tecnologie di IA verso gli obiettivi benefici della cybersicurezza che ha un effetto diretto sulla capacità di sostenere le libertà individuali, lo Stato di diritto e la democrazia, e prevenire i possibili esiti negativi con adeguate misure politiche, giuridiche e tecnologiche. Promuovere nel contesto della cybersicurezza pratiche eticamente positive nell'uso dell'IA significa garantire che il suo sviluppo e impiego avvengano in un contesto socio-tecnico inclusivo di tecnologie, capacità umane, strutture organizzative e norme etiche e giuridiche, in cui siano rispettati e promossi gli interessi individuali e i valori sociali[36].

Metodologie e linee guida per la cybersicurezza delle soluzioni di IA sono ancora in fase di studio[37]. I framework esistenti mirano a promuovere la diffusione di un'intelligenza artificiale affidabile grazie a tre fattori essenziali: (1) rispetto dell'articolato quadro normativo in materia; (2) garanzia dei principi e dei valori etici; (3) solidità dei sistemi AI da un punto di vista tecnico e sociale. La robustezza tecnica e la sicurezza del sistema – robustezza nel caso di problemi e resilienza contro i tentativi di alterare l'uso o le prestazioni – costituiscono, insieme, uno dei sette requisiti fondamentali delle linee guida etiche per una IA "degna di fiducia" (*trustworthy* AI) elaborati dal High-Level Expert Group on Artificial Intelligence nel 2019[38].

[36] Per un'ampia introduzione e discussione si veda L. FLORIDI, *Etica dell'artificiale. Sviluppi, opportunità, sfide*, Milano, 2022. Su etica e intelligenza artificiale, tra gli altri: F. FOSSA, V. SCHIAFFONATI, G. TAMBURRINI (a cura di), *Automi e persone. Introduzione all'etica dell'intelligenza artificiale e della robotica*, Roma, 2021; M. ZANICHELLI, *L'intelligenza artificiale e la persona: tra dilemmi etici e necessità di regolazione giuridica*, in *Teoria e Critica della Regolazione Sociale*, 2, 2021, 141-159; F.H. LLANO-ALONSO, *L'etica dell'intelligenza artificiale nel quadro giuridico dell'Unione europea*, in *Ragion pratica*, 2, 2021, 327-348.

[37] ENISA, Mind the Gap in Standardisation of Cybersecurity for Artificial Intelligence, 2023.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Sotto il profilo giuridico, la cybersicurezza dell'IA, assieme alla accuratezza e alla robustezza, è affrontata dalla nuova legge sulla intelligenza artificiale (Regolamento (UE) 1689/2024, c.d. AI Act) all'articolo 15. Ai considerando 75 e 76, il legislatore sottolinea che la robustezza tecnica è un elemento cruciale perché i sistemi di IA ad alto rischio siano in grado di resistere a comportamenti dannosi o indesiderati che possono emergere a causa di errori, guasti, o situazioni impreviste e, parimenti, la sicurezza informatica è fondamentale per proteggere i sistemi di IA dai tentativi di attacco da parte di aggressori che mirano a sfruttare le vulnerabilità del sistema. I requisiti di cybersicurezza stabiliti dall'AI Act contemplano quattro elementi principali[39]: (1) i sistemi di intelligenza artificiale ad alto rischio devono essere garantiti e progettati per essere resilienti ai tentativi di alterarne l'uso, il comportamento e le prestazioni e di comprometterne le proprietà di sicurezza da parte di terzi malintenzionati che ne sfruttano le vulnerabilità; (2) per raggiungere questi obiettivi devono essere implementate misure organizzative e tecniche; (3) per i sistemi di IA ad alto rischio deve essere effettuata una valutazione del rischio di cybersecurity e, infine, (4) le soluzioni tecniche devono essere adeguate alle circostanze e ai rischi pertinenti.

Dopo l'entrata in vigore della legge sull'IA, tutti i sistemi di IA ad alto rischio definiti dalla legislazione dovranno essere sottoposti a una valutazione di conformità e rispettare i requisiti di cybersicurezza prima di poter essere utilizzati o messi in servizio nel mercato dell'UE.

Le disposizioni in materia di cybersicurezza introdotte dalla legge sulla intelligenza artificiale si inseriscono nel complesso quadro legislativo europeo sulla cybersicurezza, introdotto nella prima sezione, rendendo dunque necessario il coordinamento tra i diversi perimetri normativi. In particolare, per i sistemi di IA ad alto rischio che rientrano anche nell'ambito di applicazione del cd. Cyber Resilience Act (CRA), regolamento relativo a requisiti orizzontali di cibersicurezza per i prodotti con elementi digitali – e quindi suscettibile di ricomprendere nel suo ambito di applicazione anche sistemi di IA –, il legislatore europeo ha previsto che tali sistemi debbano essere conformi ai requisiti essenziali di cui all'allegato I del CRA[40].

Per garantire la conformità ai requisiti legislativi, vi è un forte indirizzamento da parte della Commissione europea verso l'adesione a standard tecnici armonizzati. Ciò anticipa la necessità di un continuo processo di standardizzazione sulla cybersicurezza dell'IA nei prossimi anni. La gestione della cybersicurezza può avvalersi di standard tecnici e pratiche consolidate che comprendono procedure sui principi organizzativi, sulla gestione del rischio e misure di sicurezza; tuttavia, gli standard esistenti non sono ancora stati estesi alle specificità degli scenari IA.

Sul punto, il *Multilayer framework for Good cybersecurity practices for AI* di ENISA (2023) suggerisce di progettare la sicurezza in tre livelli: al primo livello (*Cybersecurity Foundations*), l'insieme di conoscenze e pratiche di base della cybersecurity che devono essere applicate a tutti gli ambienti ICT,

[38] I-HLEG, High-Level Expert Group on Artificial Intelligence, *Ethics Guidelines for Trustworthy AI*, European Commission, 2019.

[39] European Commission, Joint Research Centre, H. JUNKLEWITZ, et al., *Cybersecurity of artificial intelligence in the AI Act – Guiding principles to address the cybersecurity requirement for high-risk AI systems*, Publications Office of the European Union, 2023.

[40] P.G. CHIARA, *Il Cyber Resilience Act: la proposta di regolamento della Commissione europea relativa a misure orizzontali di cybersicurezza per prodotti con elementi digitali*, in *Rivista Italiana di Informatica e Diritto*, 1, 2023, 143-153.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

compresi quelli che ospitano i sistemi di IA; al secondo livello (*AI specific),* le pratiche di cybersecurity necessarie per affrontare le specificità dei componenti dell'IA con una visione del loro ciclo di vita, delle loro proprietà, delle minacce e dei controlli di sicurezza che sarebbero applicabili indipendentemente dal settore; al terzo livello (*Sectorial AI*), le diverse *best practices* che possono essere utilizzate dagli stakeholder settoriali per proteggere i loro sistemi di IA. In questo livello dovrebbero collocarsi i sistemi di IA ad alto rischio, come identificati dalla legge sull'IA.

5. Conclusioni

Alla luce di quanto presentato, l'intelligenza artificiale rappresenta un potente strumento per migliorare la protezione contro le minacce informatiche. Guardando al futuro, l'IA consentirà lo sviluppo di sistemi di difesa proattivi e adattativi, capaci di evolversi e adattarsi continuamente in risposta alle nuove minacce, e favorirà una maggiore collaborazione e condivisione delle informazioni tra settore pubblico e privato, migliorando la resilienza complessiva. L'IA introduce anche importanti sfide alla cybersicurezza: dalla maggior capacità offensiva da parte di attori malevoli a problemi specifici – quali il rischio di iniquità e discriminazioni e la profilazione – e a nuove e inedite vulnerabilità legate all'incertezza dell'innovazione. Sullo sfondo, è opportuno ricordare come l'adozione sempre più pervasiva di queste tecnologie di sicurezza solleva – paradossalmente – preoccupazioni circa la normalizzazione della sorveglianza: per essere efficienti ed efficaci, le tecnologie di cybersicurezza analizzate nel presente contributo, soprattutto se combinate con sistemi di intelligenza artificiale, devono trattare una vasta mole di dati, anche personali, potenzialmente compromettendo il diritto alla protezione dei dati e alla privacy.

I sistemi per il rilevamento di anomalie, in particolare, svolgono un monitoraggio continuo del traffico di rete, dei comportamenti degli utenti e dei processi per identificare, con l'aiuto di algoritmi di apprendimento automatico, deviazioni rispetto al comportamento normale del sistema o dell'utente. Questo approccio, essenziale per la cybersicurezza, perché aiuta le organizzazioni a rispondere in tempo reale a minacce esterne e interne, evidenzia come la sicurezza informatica possa diventare una giustificazione per la sorveglianza, in diversi contesti.

La proposta di Regolamento di attuazione della direttiva NIS 2 (UE) 2022/2555 pubblicata dalla Commissione europea il 27 giugno 2024, che stabilisce norme per l'applicazione della direttiva con riguardo ai requisiti tecnici e metodologici delle misure di gestione del rischio di cibersicurezza[41], inserisce le tecnologie di rilevamento delle anomalie nelle pratiche di cybersicurezza obbligatorie per un serie di soggetti e ne incentiva l'uso in tutti gli Stati membri. Non vengono fornite, tuttavia, indicazioni su come queste tecnologie debbano essere impiegate in modo da rispettare i diritti fondamentali alla protezione dei dati e alla privacy della Carta dei Diritti fondamentali dell'Unione europea.

La possibilità di trarre reale beneficio dalle tecnologie emergenti, per garantire un futuro digitale più sicuro, giusto e resiliente, dipenderà non solo dalla capacità delle organizzazioni sia pubbliche che private di fare progressi nelle ricerche su IA e cybersicurezza, nella loro applicazione e nello sviluppo delle competenze, ma anche dallo studio critico continuo degli impatti di queste tecnologie sulle sfe-

[41] Cfr. https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/14241-Cybersecurity-risk-management-reporting-obligations-for-digital-infrastructure-providers-and-ICT-service-managers_en.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

re giuridiche dei cittadini e, pertanto, dalla risposta legislativa che deve essere proporzionata e adeguata. L'approccio tecnico-ingegneristico negli ambienti complessi di oggi non è sufficiente ma occorre sviluppare una visione olistica di controllo del rischio che comporti la protezione in modo coordinato dei molti valori in gioco e che diventi anche un requisito etico e giuridico alla base dello sviluppo dei sistemi informatici.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Il diritto alla città intelligente e la cittadinanza vulnerabile. Spunti per una critica socio-tecnica dell'IA

Paolo Vignola*

THE RIGHT TO THE INTELLIGENT CITY AND THE VULNERABLE CITIZENSHIP. NOTES FOR A SOCIOTECHNICAL CRITIQUE OF AI

ABSTRACT: The "legal underdetermination" () of smart cities leads one to speak of cities with vulnerable citizenship. As noted in the fields of AI ethics, the implementation of predictive algorithms in smart cities exacerbates the vulnerability of the citizen-user, subject not only to privacy restrictions, but to processes of dispossession and forms of epistemic injustice (). Correlatively, the normative foundations of identity, freedom, autonomy and responsibility are undermined. Taking a cue from Stiegler's pharmacological perspective and David Berry's work on digital infrasomatisation and the social right to explainability, the paper proposes a socio-technical reflection on the vulnerability of citizenship.

KEYWORDS: Algorithmic Governmentality; Pharmacology; Smartness; Epistemic injustice; Infrasomatisation.

ABSTRACT: La «sottodeterminazione giuridica» (Izzo) delle smart cities induce a parlare di città a cittadinanza vulnerabile. Come rilevato nei campi dell'etica dell'IA, l'implementazione di algoritmi predittivi nelle smart cities esacerba la vulnerabilità del cittadino-utente, soggetto non solo a restrizioni della privacy, ma a processi di espropriazione e a forme di ingiustizia epistemica (Battaglia). Correlativamente, a essere minati sono i fondamenti normativi dell'identità, della libertà, dell'autonomia e della responsabilità. Prendendo spunto dalla prospettiva farmacologica di Stiegler e dai lavori di David Berry sull'infrasomatizzazione digitale e sul diritto sociale all'esplicabilità, il paper propone una riflessione socio-tecnica della vulnerabilità della cittadinanza.

PAROLE CHIAVE: Governamentalità algoritmica; farmacologia; smartness; ingiustizia epistemica; infrasomatizzazione.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

* *PhD, Pontificia Università Antonianum. Mail: p_vignola@antonianum.eu.* *Contributo sottoposto a doppio refereraggio anonimo.*

1. Introduzione

Come noto, le smart cities si basano sull'uso sistematico dell'automatizzazione, dei Big Data e dell'intelligenza artificiale, delle tecnologie digitali, sull'accesso generalizzato, pubblico e continuo delle rete Internet a tutti i cittadini, sulla fornitura di servizi di alta qualità, sul continuo aggiornamento del design urbano e sulla riconversione ecologica della produzione verso forme di economia sostenibile. In questo senso, attraverso la connettività totale dei cittadini, l'obiettivo esplicito è dunque quello di innovare la sicurezza, l'imprenditorialità, la partecipazione democratica, l'istruzione e la formazione[1]. A ben vedere, l'idea sottostante, in linea con il soluzionismo tecnologico evidenziato criticamente da Evgeny Morozov[2], è che tutti i problemi della città — sicurezza, impiego, mobilità, salute e alimentazione, oltre al fattore ecologico — siano di natura squisitamente tecnica e che possano essere risolti più efficacemente dal calcolo algoritmico, piuttosto che dalle istituzioni e dai cittadini, che si ritrovano così messi in questione nel loro stesso statuto e, perciò, scoprono la loro vulnerabilità di fronte all'innovazione tecnologica. In questo intervento ci si focalizzerà sulla figura del cittadino in quanto vulnerabile, nel senso del rischio di indebolimento giuridico e politico del suo stesso statuto[3].

Sullo sfondo delle analisi che presenteremo si ritrova il «diritto alla città»[4] di Lefebvre, inteso come «diritto alla vita urbana trasformata e rinnovata»[5] e orizzonte critico dal quale segnalare come le promesse di emancipazione e partecipazione integrale alla vita civile attraverso la tecnologia siano state tendenzialmente disattese o addirittura represse[6]. Il diritto alla città, quando quest'ultima diventa smart, è fonte di rivendicazione di fronte a processi di morfogenesi urbana indifferenti nei confronti della necessità di tutela dei soggetti vulnerabili per genere, fascia d'età, reddito, minoranze etniche, linguistiche e culturali[7]. In tal senso, tale diritto è da intendersi quale piattaforma teorica di rivendicazione giuridica nell'ambito urbano. Inoltre, tali casi godono di una ricca letteratura critica, che incrocia un ampio ventaglio di discipline differenti — dalle diverse frange del diritto a quelle della sociologia, dall'etica alla filosofia politica, dalla data science alle scienze cognitive, dall'urbanismo digitale ai

[1] Per una ricognizione generale in lingua italiana, cfr. G. FERRARI (a cura di), *Smart City. L'evoluzione di un'idea*, Milano–Udine, 2020; ID. (a cura di), *Innovazione e sostenibilità per il futuro delle smart cities*, Milano, 2023; F. BRIA, E. MOROZOV, *Ripensare la smart city*, Torino, 2018; S. BOLOGNINI, *Dalla Smart City alla "Human Smart City" e oltre*, Milano, 2017.

[2] E. MOROZOV, *Internet non salverà il mondo: perché non dobbiamo credere a chi pensa che la rete possa risolvere ogni problema*, Milano, 2014.

[3] A tal proposito, come evidenziato da Valerio Nitrato Izzo, la «sottodeterminazione giuridica» delle smart cities induce a parlare di città a cittadinanza vulnerabile. Cfr. V. NITRATO IZZO, *Urbanizzazione intelligente e trasformazioni della cittadinanza: nuove generazioni dei diritti nella città digitale*, in G. FERRARI (a cura di), *Innovazione e sostenibilità per il futuro delle smart cities*, Milano, 2023, 393.

[4] H. LEFEBVRE, *Il diritto alla città*, Verona, 2014.

[5] *Ivi*, 113.

[6] cfr. K. WILLIS, *Whose Right to the Smart City?*, in R. KITCHEN, P. CARDULLO, C. DI FELICIANTONIO (a cura di), *The Right to the Smart City*, Bingley, 2018, 27-42. Cfr. inoltre R. KITCHIN, T.P. LAURIAULT, G. MCARDLE (a cura di), *Data and the City*, Londra, 2017.

[7] Un'eccellente analisi delle discriminazioni algoritmiche in generale è fornita da E. FALLETTI, *Discriminazione algoritmica*. *Una prospettiva comparata*, Torino, 2023. Su città e discriminazione, cfr. l'importante saggio di F. CIARAMELLI, *La città degli esclusi*, Pisa, 2023.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

gender studies —, ma si distinguono per la diversità d'approccio, il primo più empirico e aderente alla materia legislativa, mentre il secondo risulta essere più di carattere speculativo e olistico del diritto alla città.

Il presente lavoro intende indagare questa seconda opzione, indirizzata alla cittadinanza in generale, dal punto di vista di una particolare declinazione della filosofia della tecnica, che definiamo in questa sede "farmacologica", e che è possibile rintracciare nell'incrocio virtuoso tra prospettive distinte, in quanto provenienti da discipline diverse, ma decisamente compatibili a partire da una comune diagnosi socio-tecnica dell'implementazione del digitale e dell'IA negli spazi urbani[8]. In particolare, con filosofia farmacologica della tecnica ci riferiamo alla prospettiva propiziata innanzitutto da Bernard Stiegler, che pensa ogni tecnica e tecnologia come un *pharmakon*, da intendersi quale dispositivo antropogenetico assolutamente non neutrale, bensì ambivalente, ossia nella sua funzione al tempo stesso costituente e destituente delle facoltà umane, così come delle istituzioni sociali[9].

In estrema sintesi, riprendendo la questione del *pharmakon* platonico decostruita da Derrida, per cui la scrittura non è solo un rimedio e un veleno per l'anima e la memoria, bensì in definitiva il supplemento necessario del logos e dell'anamnesi — dunque di ciò che sarebbe il proprio dell'uomo —[10], Stiegler estende tale statuto alla tecnica in generale, intesa come esteriorizzazione delle funzioni organiche e delle facoltà mentali e quindi quale vettore di una memoria artificiale, esteriorizzata, che concorre sia al processo di ominazione, sia alla trasmissione spaziale e temporale dei saperi, sia ancora al susseguirsi delle epoche e delle crisi sociali[11]. Pensare la tecnica come *pharmakon* non significa, come anticipato, considerarla neutrale, a disposizione di un soggetto che autonomamente può farne un buono o un cattivo uso, bensì al contrario la dimensione farmacologica risiede nel fatto che essa determina sempre e sistematicamente degli effetti trasformativi, per cui la stessa autonomia di fronte alla tecnica non è un a priori ma il risultato di processi collettivi di adozione critica. Per Stiegler pensare la tecnica come *pharmakon* significa perciò concepire i rapporti con le tecnologie come terapeutiche della memoria — dunque diagnosi e prognosi — che siano al tempo stesso sociali, cognitive, economiche e politiche.

A tale prospettiva, e relativamente a ciò che concerne il presente lavoro, si intende fare afferire, in particolare, la critica dell'algoritmizzazione della governamentalità e del diritto sviluppata soprattutto da Antoinette Rouvroy e Thomas Berns[12], i lavori di *critical digital humanities* di David Berry[13], le analisi sul «soluzionismo tecnologico» di Morozov, così come gli studi di Robert Mitchell e Orit Halpern sulle smart cities[14]. Proprio a partire da questi due ultimi autori è possibile dotarsi di un concetto critico di

[8] La prospettiva farmacologica è l'orizzonte condiviso dalla rete internazionale Digital Studies, lanciata da Bernard Stiegler nel 2014: https://digital-studies.org/wp/call-for-digital-studies/ (ultima consultazione 22/11/2024).

[9] Cfr. B. STIEGLER, *Prendersi cura. Della gioventù e delle generazioni*, Napoli-Salerno, 2014.

[10] Cfr. J. DERRIDA, *La farmacia di Platone*, Milano, 1978.

[11] Cfr. B. STIEGLER, *La tecnica e il tempo I. La colpa di Epimeteo*, Roma, 2023.

[12] T. BERNS, A. ROUVROY, *Gouvernementalité algorithmique et perspectives d'émancipation. Le disparate comme condition d'individuation par la relation ?*, in *Réseaux*, 177, 2013.

[13] D. BERRY, *Smartness et le tournant de l'explicabilité*, in B. STIEGLER (a cura di), *Le nouveau génie urbain*, Parigi 2020, 31-68.

[14] Cfr. O. HALPERN, R. MITCHELL, B.D. GEOGHAGAN, *The Smartness Mandate: Notes Toward a Critique*, in *Grey Room*, 68, 2017, 106-129; O. HALPERN, R. MITCHELL, *Smartness, populations et infrastructures*, in B. STIEGLER (a cura di), *Le nouveau génie urbain*, Parigi, 2020, 69-86.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

smartness relativo alle smart cities, dove il significato che si dà all'intelligenza ci avvicina al cuore del problema dello statuto stesso della cittadinanza e si pone come sfondo critico su cui tratteggiare l'idea di un diritto alla città intelligente.

Mitchell e Halpern sottolineano innanzitutto la radicale differenza tra l'intelligenza civica, nel senso dell'insieme dei saperi funzionali al prosperare di una città, ma anche nel senso di una volontà collettiva, e la *smartness* in quanto effetto generale di un'infrastruttura informatica finalizzata a ottimizzare ambienti, energie, informazioni e interazioni urbane. La differenza incommensurabile risiede nell'autofinalità della *smartness*, che «è sia un mezzo che un fine», nel senso che «il *telos*» dei dispositivi e delle infrastrutture smart è innanzitutto «"più *smartness*"» [15] — non "più diritti" o "più equità" — il ché significa non solo semplicemente più efficienza, performatività, precisione predittiva, bensì un radicale cambio di prospettive: la *smartness* "considera" il cittadino come un mero insieme di dati in movimento che alimenta le tecniche di apprendimento automatico delle infrastrutture digitali, ossia se stessa.

Nei prossimi paragrafi proveremo ad analizzare i contorni giuridico politici e normativi di quello che Mitchell e Halpern hanno definito l'imperativo della *smartness* (*smartness mandate*), e lo faremo passando per tre filtri concettuali: la governamentalità algoritmica e il Leviatano elettronico; la *smartness* e l'ingiustizia epistemica; l'avvolgimento digitale e l'infrasomatizzazione. L'obiettivo è quello di sviluppare una riflessione socio-tecnica sulla vulnerabilità della cittadinanza, sottolineandone la sua *condizione farmacologica*[16].

2. La città senza cittadini e la fine della res publica

L'AI Act, approvato dall'Unione Europea il 13 marzo 2024, dopo una serie di step preliminari e lavori preparatori tra cui il *Libro bianco sull'IA* a cui faremo riferimento, rappresenta un encomiabile tentativo, peraltro ragionevolmente riuscito, di introdurre un quadro normativo e giuridico comune in merito all'IA, estendendo l'ambito di applicazione a tutti i settori della società (salvo quello militare), e a tutti i tipi di intelligenza artificiale. L'obiettivo generale, di fronte ai sistemi di IA ad alto rischio, è la protezione dei diritti fondamentali dei cittadini dell'Unione, garantendo la democrazia, lo Stato di diritto e la protezione dell'ambiente, e promuovendo al tempo stesso l'innovazione nel conferire all'Europa un ruolo da leader nel settore[17]. Più in particolare, si dichiara che «Lo scopo del presente regolamento è migliorare il funzionamento del mercato interno istituendo un quadro giuridico uniforme in particolare per quanto riguarda lo sviluppo, l'immissione sul mercato, la messa in servizio e l'uso di

[15] *Ivi*, 70.

[16] Sulla condizione farmacologica, cfr. B. STIEGLER, *États de choc. Bêtise et savoir au XXI° siècle*, Parigi, 2012.

[17] Ad essere vietati saranno in particolare determinati sistemi di categorizzazione biometrica, la creazione di banche dati di riconoscimento facciale in base a estrapolazioni indiscriminate; i sistemi di riconoscimento delle emozioni sul luogo di lavoro e nelle scuole, i sistemi di credito sociale, le pratiche di polizia predittiva basate sulla profilazione, i sistemi manipolatori del comportamento umano e delle vulnerabilità. I sistemi di IA identificati ad alto rischio includono le infrastrutture critiche, l'istruzione e la formazione, le componenti di sicurezza dei prodotti, l'occupazione e il mercato del lavoro, i servizi privati e pubblici essenziali, la gestione dei flussi migratori e l'amministrazione della giustizia in tutte le sue forme.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

sistemi di intelligenza artificiale (sistemi di IA) nell'Unione, […] promuovere la diffusione di un'intelligenza artificiale (IA) antropocentrica e affidabile»[18].

Da sottolineare che l'Unione Europea si è contraddistinta per la maggiore volontà di regolamentare by the Law la realtà dell'IA rispetto agli Stati Uniti o ai paesi asiatici, mostrando preoccupazioni sul tema dei diritti molto più chiare ed efficaci, almeno sulla carta. Dal punto di vista che qui intendiamo presentare, quello di filosofia della tecnologia nella sua declinazione farmacologica, vi è però una questione che merita essere posta ad analisi, e riguarda innanzitutto il rapporto tra affidabilità dell'IA e autonomia dei cittadini.

Per osservarla, occorre ritornare al *Libro bianco sull'intelligenza artificiale – Un approccio europeo all'eccellenza e alla fiducia*, del 19 febbraio 2020[19]. Prima ancora di evidenziare le criticità e i rischi connessi all'IA, il documento si apre appunto in modo estremamente fiducioso: «[l]'intelligenza artificiale si sta sviluppando rapidamente. Cambierà le nostre vite migliorando l'assistenza sanitaria […], aumentando l'efficienza dell'agricoltura, contribuendo alla mitigazione dei cambiamenti climatici e all'adattamento ai medesimi, migliorando l'efficienza dei sistemi di produzione mediante la manutenzione predittiva, aumentando la sicurezza dei cittadini europei e in molti altri modi che possiamo solo iniziare a immaginare»[20]. Tale approccio non solo conferisce uno statuto rivoluzionario — e non semplicemente innovativo — o trascendente alla tecnologia in quanto risolutiva di ogni problema che sta affrontando l'umanità, ma si dimostra anche troppo leggero nell'assumere che i cittadini digitali siano una realtà, ossia persone che possiedono le competenze necessarie per comprendere autonomamente la continua innovazione tecnologica e attraverso di essa partecipare alla vita democratica, impegnandosi nel rispetto dei diritti umani.

Nella sua preziosa analisi semantica dell'ordine del discorso istituzionale in merito all'IA, Silvia Salardi mostra da un lato i rischi nel conferire un tipo di fiducia quasi messianica alla tecnologia, dall'altro l'equivoco che può generarsi nel credere a un'autonomia dei cittadini di fronte alle innovazioni tecnologiche. In particolare, l'equivoco si basa «sul presupposto che, data l'autonomia dei destinatari delle tecnologie, bastino le regole del libero mercato a orientare le scelte. In altre parole, in questa visione l'autonomia è qualcosa di innato negli esseri umani. Mentre più realisticamente occorre riconoscere che l'autonomia è un work in progress durante l'esistenza di un individuo, come tale va nutrita e accompagnata nel suo sviluppo»[21]. Come anticipato, la prospettiva farmacologica, non solo ha una visione analoga, dunque processuale, dell'autonomia individuale e concepisce negli stessi termini il rapporto con la tecnologia, ma evidenzia anche il rischio costante della perdita di autonomia in tale

[18] https://digital-strategy.ec.europa.eu/it/policies/regulatory-framework-ai (ultima consultazione 22/11/2024). Per una ricognizione in lingua italiana di tali tematiche, cfr. G. PITRUZZELLA, *La libertà di informazione nell'era di Internet*, in *MediaLaws*, 1, 2018, 30 ss.; F. PIZZETTI (a cura di), *Intelligenza artificiale, protezione dei dati personali e regolazione*, Torino, 2018; A. D'ALOIA, (a cura di), *Intelligenza artificiale e diritto. Come regolare un mondo nuovo*, Milano, 2020; U. RUFFOLO (a cura di), *Intelligenza artificiale. Il diritto, i diritti, l'etica*, Milano, 2020; P. SEVERINO (a cura di), *Intelligenza artificiale. Politica, economia, diritto, tecnologia*, Roma, 2022.

[19] Accessibile al sito https://commission.europa.eu/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en (ultima consultazione 22/11/2024).

[20] *Ibidem*.

[21] S. SALARDI, *Intelligenza artificiale e semantica del cambiamento: una lettura critica*, Torino, 2023, 45.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

rapporto. Vedremo ora il primo ostacolo alla costruzione di tale autonomia mediante il concetto di governamentalità algoritmica e l'immagine del Leviatano elettronico.

All'interno del lavoro collettivo e interdisciplinare sulle città intelligenti condotto dal gruppo di ricerca Internation[22], il cui cuore diagnostico è la progressiva delega della facoltà decisionale, da parte del cittadino, a sistemi predittivi o di decisione autonoma, nel 2020 Bernard Stiegler si domandava provocatoriamente «come si può definire una città "intelligente", se non immaginandola senza i suoi stessi abitanti?»[23]. Se, come sottolinea dal punto di vista del diritto Valerio Izzo, «la città intelligente rischia di diventare l'epifenomeno digitale di un nuovo modello di esclusione in ambito urbano, di città senza cittadinanza»[24], Stiegler constata l'emergere di un'idea di città per così dire senza cittadini. La provocazione è però ben altro che sterile o improvvisata, dal momento che questo evitamento della partecipazione cittadina è un tema ricorrente nelle analisi critiche delle smart cities. A proposito dell'eventualità di una cittadinanza smart, Laura Sartori segnalava già nel 2015 come il cittadino e le sue istanze siano stati assenti nel dibattito sulle smart city, rimuovendo così il contributo di intelligenza civica che esso può apportare[25]. In tal senso, seguendo ancora Izzo, potremmo dire che, nella smart city, "intelligente" fa rima con "indifferente", dal momento che l'innovazione tecnologica urbana rimane spesso distante «nei confronti dell'implementazione di strumenti di diritto internazionale e nazionale a tutela dei diritti umani»[26], così come sembra disinteressarsi delle persone concrete in generale, per non parlare delle minoranze e di tutti i soggetti vulnerabili in termini di reddito, genere, estrazione sociale o fascia d'età.

Ci preme a questo proposito sottolineare come, da spazio pubblico di produzione e condivisione, la città divenga così un dispositivo reticolare di estrazione di dati e colonizzazione attraverso di essi di questo stesso spazio pubblico e del tempo delle coscienze individuali dei cittadini, divenuti meri utenti. A tal proposito, sono illuminanti le analisi di Antoinette Rouvroy e Thomas Berns sulla «governamentalità algoritmica», ossia una forma di governo dei comportamenti attraverso l'estrazione, l'analisi, la profilazione e la correlazione dei Big data, a fini essenzialmente predittivi, in quanto tali processi anticipano, modulano e selezionano azioni e desideri individuali e collettivi, con l'obiettivo di garantire condotte specifiche relativamente all'impiego e ai consumi, o minimizzare espressioni inappropriate dal punto di vista giuridico o politico. In quest'ottica, gli individui, in un processo generalizzato che li vede impegnarsi come utenti delle piattaforme prima che come cittadini, contano solo come profili, disegnati algoritmicamente attraverso flussi di dati quantitativi e metadati: «Frammentato in miriadi di dati, l'individuo diventa infinitamente calcolabile, comparabile, indicizzabile e intercambiabile»[27]. Dal punto di vista giuridico politico, le forme di gestione algoritmica dei dati dei cittadini non generano spazio pubblico, bensì, precisamente, «una colonizzazione dello spazio pubblico da parte di una sfera

[22] https://internation.world/ (ultima consultazione 22/11/2024)

[23] B. Stiegler, Collectif Internation, *L'assoluta necessità. In risposta ad António Guterres e Greta Thunberg*, Roma, 2020, 110.

[24] V. Nitrato Izzo, *op. cit.*, 400-401.

[25] L. Sartori, *Alla ricerca della smart citizenship*, in *Istituzioni del federalismo*, 4, 2015, 927-948.

[26] V. Nitrato Izzo, *op. cit.*, 394.

[27] A. Rouvroy, *The end(s) of critique: data-behaviourism vs. due-process*, in M. Hildebrandt e E. De Vries (a cura di), *Privacy, Due Process and the Computational Turn. Philosophers of Law Meet Philosophers of Technology*, Londra, 2013, 157.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

privata ipertrofica»[28]. Questa è allora la questione dirimente, che trova sulla stessa linea l'ormai celebre teoria di Benjamin Bratton sull'organizzazione geopolitica digitale, dall'autore definita come "la pila" (the Stack), da intendersi come stratificazione gerarchica di sei "spazi" di potere (*Earth, Cloud, City, Address, Interface, User*), che andrebbero a costituire il nuovo *nomos* della Terra, ossia l'immagine di una «geografia politica condensata verticalmente»[29]:

> Cosa dovremmo intendere con "pubblico" se non ciò che è costituito da tali interfacce, e dove altro potrebbe risiedere la "governance" – intesa qui come la necessaria e deliberata composizione esecutiva di soggetti politici durevoli e delle loro mediazioni – se non precisamente in esse? […] nelle immanenti, immediate e perfettamente presenti interfacce che ci ritagliano e ci legano. Dove dovrebbe risiedere la sovranità se non in ciò che è tra noi, derivante non da ognuno di noi individualmente, ma da ciò che disegna il mondo attraverso di noi?[30]

Alla dissoluzione del pubblico, o almeno dell'idea di spazio pubblico, secondo Rouvroy si associa la disintegrazione dell'autonomia del soggetto, che nel parallelo con la dimensione pubblica viene a essere la dissoluzione dell'idea stessa di cittadinanza. Per la filosofa belga, infatti, se le tecniche di profilazione psicografica annunciano «una nuova "trasparenza digitale" della psiche individuale», tale trasparenza è funzionale a «indirizzare il comportamento in una fase preconscia, in modo letteralmente subliminale, rendendo possibili forme inedite di sfruttamento delle "vulnerabilità" psicologiche degli individui, in particolare nel campo del marketing personalizzato, o nuove forme di governo comportamentale»[31]. La vulnerabilità del cittadino risiede dunque nell'essere colto non più come una persona, bensì come un «aggregato di propensioni» funzionali ad alimentare le tecniche di apprendimento automatico delle infrastrutture digitali urbane e, dunque, a condizionare le azioni future dei singoli e della collettività. Come segnalato da Garapon e Lassègue, «il soggetto di diritto diventa contemporaneamente consumatore e prodotto, predatore e preda, sorvegliante e sorvegliato, giornalista e spettatore, agito e agente del proprio mondo»[32]. Questa dis-integrazione parallela del pubblico e del cittadino viene ripresa da Stiegler, che in un testo ancora inedito ne trae una conclusione radicale, ossia il dissolversi della *res publica*:

> Le popolazioni della biosfera, essendo calcolate nella loro totalità, come un insieme, e in modo permanente, attraverso le loro azioni e i loro gesti diventano "servizi". La statistica, scienza dello Stato, è sostituita dalla scienza dei dati, i cui protocolli sono sviluppati nell'opacità funzionale e al servizio esclusivo dell'economia dei dati, come amministrazione delle cose in cui la *res publica* in quanto tale, la cosa pubblica, cioè il diritto, è dissolta[33].

In questo senso, nella stessa misura in cui per Rouvroy i principi tradizionali che definiscono l'*homo juridicus* come soggetto di diritto si stanno dissolvendo negli schemi statistici dell'*homo numericus*

[28] T. Berns, A. Rouvroy, *op. cit.*, 172.
[29] B. Bratton, *The Black Stack*, in *e-flux journal*, 53, 2014, in https://www.e-flux.com/journal/53/59883/the-black-stack/ (ultima consultazione 22/11/2024).
[30] *Ibidem*.
[31] A. Rouvroy, *Homo juridicus est-il soluble dans les données ?*, in E. Degrave et al. (a cura di), *Law, norms and freedoms in cyberspace = Droit, normes et libertés dans le cybermonde: Liber Amicorum Yves Poullet*, Bruxelles 2018, 419-420.
[32] A. Garapon, J. Lassègue, *La giustizia digitale. Determinismo tecnologico e libertà*, Bologna, 2021, 211.
[33] B. Stiegler, *Technics and Time 4* (inedito), 129.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

(Rouvroy, 2018)[34], possiamo osservare, nelle analisi di Stiegler, la tendenza della *res publica* a dissolversi nello stato di fatto del capitalismo delle piattaforme. Fenomeno che chi scrive aveva precedentemente inquadrato come il divenire *res extracta* della *res publica*[35], ossia il mero risultato delle attività di estrattivismo dei dati e del calcolo algoritmico e finanziario che su di essi viene effettuato. Ci pare del resto che Garapon e Lassègue si muovano nello stesso terreno critico quando segnalano che «il diritto si sottomette così a un ordine più vasto, a cui ha cessato di dare un fondamento giuridico»[36]. Tale ordine, propiziato dalla «piattaformizzazione delle istituzioni»[37], ha per Stiegler come correlato concettuale la figura del Leviatano elettronico. Questa forma algoritmica del Leviatano esprime in effetti bene la doppia disintegrazione, del cittadino e della *res publica*. Così come il Leviatano di Hobbes prevedeva il trasferimento, da parte del singolo individuo, del diritto di governarsi allo Stato, nel Leviatano elettronico tale diritto viene trasferito, attraverso i dati digitali, a «un governo automatico puramente computazionale», che disintegra tanto i cittadini quanto le istituzioni: esso infatti «ha bisogno di individui psichici da disintegrare, così come questi hanno bisogno di sistemi sociali che a loro volta disintegrano servendo il Leviatano elettronico e decadente»[38]. Ora, tale trasferimento del *diritto all'autonomia*, che comprende quello del *diritto alla critica* dello stato di fatto, e dunque alla stessa noeticità dei cittadini (ossia il loro pensiero critico, razionale e politico), a differenza del Leviatano tradizionale, è frutto di un atto totalmente involontario e causato dall'effetto di rete[39]. Questa è per il filosofo francese la condizione dell'attuale «società automatica», che per tale ragione dovrebbe essere intesa come una «dis-società automatica», nel senso che il calcolo algoritmico produce un cortocircuito tra gli individui e la dimensione collettiva e pubblica. Ciò poiché la delega al governo algoritmico e alla conseguente sincronizzazione automatica degli utenti significa «abbandonare la propria capacità diacronica e singolare di contribuire per se stesso all'individuazione collettiva» e ciò «non può che condurre alla sterilizzazione di quel che formava la fecondità transindividuale […] produttrice di fatti e di diritti negantropologici – vale a dire di culture, di culti, di cure e di sollecitudini»[40]. Questa fecondità collettiva, oggi a rischio di sterilizzazione, è per noi precisamente quello che dà forma e struttura a una società — ciò che dà dunque cittadinanza.

3. Smartness e ingiustizia epistemica

In un recente saggio sulla questione costituzionale dell'IA, Andrea Simoncini constata che le tecnologie smart cortocircuitano di fatto le categorie di agente e di strumento, così come di mezzo e fine[41]. Abbiamo già anticipato il problema dell'autofinalità della *smartness* all'interno delle smart cities, per cui

[34] Cfr. A. ROUVROY, *Homo juridicus est-il soluble dans les données ?*, cit.
[35] Cfr. S. BARANZONI, P. VIGNOLA, *Para acabar con la imagen extractivista del pensamiento. Una ficción filosófica*, in *Culture Machine*, 21, 2022, 1-24.
[36] A. GARAPON, J. LASSÈGUE, *op. cit.*, 201.
[37] G. CRISTOFARI, *Bratton and the Double Movement of State Platformization and Platform Institutionalization*, in *La Deleuziana*, 13, 2022, 83-101.
[38] B. STIEGLER, *La società automatica I. L'avvenire del lavoro*, Roma, 2019, 243.
[39] *Ivi*, 403.
[40] A. SIMONCINI, *L'algoritmo incostituzionale: intelligenza artificiale il futuro delle libertà*, in A. D'ALOIA, (a cura di), *Intelligenza artificiale e diritto*, Milano, 2020, 173.
[41] B. STIEGLER, *La società automatica*, cit., 403.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

essa sarebbe al tempo stesso un mezzo e un fine, nella misura in cui l'obiettivo delle tecnologie cibernetiche è l'ottimizzazione del calcolo e la sua progressiva estensione a ogni aspetto della realtà, dunque «"più *smartness*"». Seguendo Mitchell e Halpern, la confusione di mezzi e fini si ripercuote nella comprensione del rapporto tra persone e infrastrutture nella smart city:

> Tradizionalmente pensiamo alle infrastrutture come elementi stabili, continuamente accessibili e, dal punto di vista degli utenti, come sistemi […] che forniscono un quadro per le attività di una popolazione: le reti stradali, ad esempio, consentono di convogliare i diversi membri di una popolazione tra le diverse parti del territorio, mentre le infrastrutture elettriche forniscono energia alla maggior parte della popolazione. La smartness, al contrario riconfigura la popolazione umana non solo in termini di utilizzo dell'infrastruttura, ma come infrastruttura stessa[42].

Se l'obiettivo della *smartness* non riguarda i fini della città o i diritti dei cittadini, è perché l'umanità diviene oggetto di calcolo, si fa risorsa tramite la datificazione, e giunge a confondersi con l'infrastruttura. La *smartness* sostituisce la razionalità umana e per i due autori ciò implica che «le popolazioni umane funzionino come infrastrutture per le tecniche di apprendimento automatico»[43]. Affinché ciò diventi possibile è allora necessario che «le attività della popolazione umana urbana siano catturate in modi stabili, coerenti e continuamente disponibili [e che] funzionino come un'infrastruttura per gli algoritmi di apprendimento e i dati sulle popolazioni che mobilitano»[44]. In altre parole, è necessario un regime di estrazione sistematica dei dati, che Matteo Pasquinelli e Vlad Joler hanno definito *knowledge extractivism* e che esprime un nuovo tipo di colonizzazione capace di investire le popolazioni umane nei territori fino a poco tempo fa inesplorati delle emozioni e delle facoltà cognitive, invadendo sempre più in profondità i corpi e le menti degli individui. In tal senso, nel parlare di intelligenza artificiale, occorre comprendere che l'intelligenza e il sapere che emergono dalle macchine non sono creati algoritmicamente, bensì «gli algoritmi estraggono "intelligenza" dalle fonti di dati»[45], così come estraggono il sapere, sottraendoli dunque a chi li ha effettivamente prodotti, ossia i soggetti in carne ed ossa.
A tal proposito, come rilevato nei campi dell'etica dell'IA, l'implementazione di algoritmi predittivi nelle smart cities esacerba la vulnerabilità del cittadino-utente, soggetto non solo a restrizioni della privacy, ma a processi di espropriazione o disumanizzazione e a forme di ingiustizia epistemica[46]. Il concetto di ingiustizia epistemica, sviluppato da Fiorella Battaglia, ha il merito di diagnosticare problematiche più specifiche rispetto al principio di privacy legato alla protezione dei dati personali. Mentre quest'ultimo «si muove appunto nel campo designato dai diritti umani, l'ingiustizia epistemica si volge ai torti perpetrati nei confronti di una persona nel suo carattere di soggetto conoscente»[47]. Più in particolare, Battaglia indica due tendenze correlate all'ingiustizia epistemica: la disumanizzazione del soggetto conoscente, che nella interazione con l'IA si riduce a un utente latore di dati, dunque «alla

[42] O. HALPERN, R. MITCHELL, *op. cit.*, 71.
[43] *Ivi*, p. 72.
[44] *Ivi*, p. 71.
[45] M. PASQUINELLI, V. JOLER, *The Nooscope manifested: AI as instrument of knowledge extractivism*, in *AI & Society*, 36, 2021, 1266.
[46] F. BATTAGLIA, *Algoritmi predittivi e ingiustizia epistemica*, in M. GALLETTI, S. ZIPOLI CAIANI (a cura di), *Filosofia dell'Intelligenza Artificiale*, Bologna, 2024, 63-82.
[47] *Ivi*, 67.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

stregua di una cosa», e l'espropriazione dei contenuti mentali di questo stesso soggetto, la cui conseguenza è la perdita di autorità di quest'ultimo su di essi[48].

Intendiamo allora sottolineare che il soggetto epistemico viene decostruito e destrutturato rispetto alle proprie facoltà cognitive, morali e decisionali, nonché nella sua stessa percezione dell'identità personale, per essere ricostruito algoritmicamente come profilo. Come descrive con chiarezza Simona Tiribelli, «il profilo assegnato algoritmicamente a ciascun utente […] intende descrivere *chi sei* e, in modo predittivo, con una certa probabilità, *chi sarai*, sulla base dei dati personali generati, comparati a quelli di altri, dalla prospettiva *di chi li osserva ed elabora* […] ossia dalla prospettiva degli algoritmi»[49]. Correlativamente, a essere minati, soprattutto attraverso i nudges digitali e i *bias* algoritmici, sono i fondamenti normativi dell'identità, della libertà, dell'autonomia e della responsabilità. È proprio qui che le misure tradizionali di protezione dei dati risultano in qualche modo obsolete. Ancora Tiribelli segnala come sia illusorio pensare che le architetture digitali, mediante l'analisi dei dati degli individui e delle loro decisioni prese precedentemente, conferiscano possibilità di scelta ad personam, che rispecchiano cioè l'identità personale degli utenti: «In realtà, alimentano *bias* di conferma, promuovono ambienti chiusi e propongono una rappresentazione dell'individuo standardizzata […] facendo scomparire l'individualità all'interno di categorie generalizzanti […] oppure […] spingendo l'utente verso scelte in linea con la propria identità, [lo rinchiudono] in una bolla di relazioni e possibilità chiusa all'eterogeneità»[50].

L'automazione dei processi decisionali non può non essere foriera di discriminazioni o di processi che scavalcano l'autonomia dei soggetti, sempre più abbandonati nel limbo della «disintermediazione informativa», che Giuseppe Riva descrive come il processo attraverso cui tendono a scomparire i filtri di mediazione civile e costituzionale tra gli utenti e le imprese o le istituzioni, assorbiti dalle piattaforme[51]. Da considerare correlativamente alla disintermediazione è l'aggiustamento automatizzato, che per Eric Sadin esprime la funzione sociale della piattaformizzazione delle istituzioni[52]. L'aggiustamento va qui inteso, in senso generale, come l'ottimizzazione della relazione tra due o più corpi che vengono a incontrarsi senza una mediazione previa. In quanto supplemento alla disintermediazione, concepiamo dunque l'aggiustamento come il risultato più visibile della *smartness* davanti agli occhi degli utenti, mentre resta da analizzare ciò che non si pone davanti, bensì per così dire all'interno dei soggetti, vale a dire l'infrasomatizzazione digitale.

4. Dalla spiegazione alla comprensione, dall'avvolgimento alla infrasomatizzazione

Nell'ambito della filosofia della tecnologia e dell'etica del digitale, sono note le considerazioni di Luciano Floridi in merito allo statuto cognitivo dell'Intelligenza artificiale, secondo le quali quest'ultima

[48] *Ivi*, 79.

[49] S. TIRIBELLI, *Identità personale e algoritmi: una questione di filosofia morale*, Roma, 2023, 63.

[50] *Ivi*, 141. Per un'analisi più ampia di tali tematiche, cfr. M. GALLETTI, *Quando l'Intelligenza Artificiale incontra le scienze comportamentali. Una riflessione sui valori morali*, in M. GALLETTI, S. ZIPOLI CAIANI (a cura di), *Filosofia dell'Intelligenza Artificiale,* Bologna, 2024, 123-146.

[51] Cfr. G. RIVA, *Fake news*, Bologna, 2018, 85.

[52] Cfr. E. SADIN, *L'humanité augmentée. L'administration numérique du monde*, Montreuil, 2013, 162-163. Sull'aggiustamento, da un punto di vista più strettamente giuridico, cfr. A. GARAPON, J. LASSÈGUE, *op. cit*., 223-240.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

rappresenta una scissione radicale e inedita della capacità di agire dall'intelligenza umana atta a realizzarla, ed è tale forbice a determinare il problema etico fondamentale, che consiste nel comprendere come adottare la potenza non intelligente del calcolo computazionale[53]. L'esplicabilità, sintesi di intelligibilità (dell'azione) e responsabilità (rispetto all'azione), è per Floridi il principio etico fondamentale per un'IA affidabile e rispettosa dei principi democratici e dello stato di diritto. Tale principio afferma la necessità di poter comprendere a ogni livello il modo di operare degli algoritmi, la qualità e la provenienza dei dati usati, così come il metodo di addestramento che conduce a una decisione o a un esito invece di un altro. Per il filosofo italiano l'esplicabilità è anche la *conditio sine qua non* perché vengano effettivamente garantiti anche gli altri principi condivisi a livello globale dalla comunità scientifica e dalle istituzioni di regolazione: beneficienza, non maleficienza, autonomia, giustizia[54]. Affinché l'IA sia *benefica* e non *dannosa* è necessario *comprendere* come e in che misura sta agendo sugli individui così come sulle istituzioni; per promuovere *l'autonomia* dei soggetti è necessario che le nostre decisioni su dove l'IA possa sostituire l'azione umana siano *informate* su come essa agirebbe al posto nostro, così come affinché sia soddisfatto il principio di *giustizia* devono essere *chiare* le responsabilità non solo legali ma anche etiche di fronte a esiti dannosi.

Il principio di esplicabilità (*explainability*) in merito ai sistemi di decisione automatica è diventato parte integrante della risposta legislativa dell'Unione Europea ai rischi legati alle disuguaglianze nonché alla parzialità e all'opacità degli algoritmi. La questione che però rimane sul tavolo riguarda il tipo di spiegazione, poiché se tale principio è evidentemente necessario, non sembra essere sufficiente ad una vera e propria comprensione olistica, dunque anche sociale, normativa e politica. In tal senso, David Berry, che propone «una critica immanente della nozione di esplicabilità», ossia rivolta alle sue effettive condizioni di possibilità, mostra come essa miri sostanzialmente a una trasparenza dell'operazione tecnica ricalcata sulla logica della spiegazione scientifica, per cui mancherebbe una riflessione sugli effetti sociali del calcolo a stretto, a medio e a lungo termine. Ciò, per lo studioso inglese, si rende necessario dal momento che «le logiche infrastrutturali di calcolo decentrano e sovraccaricano deliberatamente i modi di pensare — ad esempio minando la concentrazione, la focalizzazione e l'attenzione». Si rivela perciò una contraddizione cognitiva nella stessa nozione di esplicabilità, per cui essa non può ridursi a «una semplice risposta tecnica al problema contemporaneo dei sistemi decisionali automatizzati, ma richiede un'indagine filosofica per essere adeguatamente posta nel contesto storico e concettuale»[55].

È allora ancora più interessante per il presente articolo il concetto floridiano di "avvolgimento", mediante il quale si intende la progressiva trasformazione degli ambienti umani, disegnati secondo forme sempre più compatibili con i dispositivi di intelligenza artificiale, per rendere più performanti le operazioni algoritmiche[56]. L'IA non è e non sarà dunque intelligente come gli umani o perfino di più, mentre sono e saranno gli ambienti e le infrastrutture ad essere sempre più compatibili con la *smartness*. Praticamente in qualsiasi sfera dell'esistenza umana gli spazi, che siano pubblici o privati, urbani, rurali o domestici, vengono tradotti in superfici da cui estrarre dati di ogni genere, con la progressiva

[53] Cfr. L. FLORIDI, *Etica dell'intelligenza artificiale. Sviluppi, opportunità, sfide*, Milano, 2022, 39-64.
[54] *Ivi*, 93-102.
[55] D. BERRY, *op. cit.*, 56.
[56] Cfr. L. FLORIDI, *op. cit.*, 61-84.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

scomparsa di abitudini, località, relazioni, rapporti con i luoghi e forme dell'agire. Il concetto di avvolgimento risulta però davvero interessante per cogliere la problematicità dell'autonomia dei soggetti di fronte all'IA se si prova ad approfondirne gli effetti non solo sugli spazi, su cui Floridi sembra soffermarsi in larga misura, ma anche e soprattutto sugli individui. È perciò giunto il momento di introdurre l'ultimo concetto farmacologico, ossia l'infrasomatizzazione, così definita da David Berry.

L'infrasomatizzazione è un concetto le cui componenti, a loro volta, sono già state introdotte, in quanto per Berry si tratta del risultato dell'automatizzazione (Stiegler), della piattaformizzazione (Bratton) e dell'algoritmizzazione (Rouvroy) degli ambienti e dei corpi umani. Molto brevemente, con infrasomatizzazione lo studioso inglese intende un processo di strutturazione che inscrive nuove forme del sociale nel corpo e nella mente degli individui, così come nel funzionamento delle istituzioni attraverso gli algoritmi, che vengono considerati non quali semplici mezzi o strumenti, bensì come una nuova forma di infrastruttura cognitiva[57]. L'infrasomatizzazione è da intendersi come complemento simmetrico dell'esosomatizzazione, altro concetto farmacologico, coniato da Stiegler, con cui si intende l'insieme dei processi di esteriorizzazione delle funzioni organiche e noetiche in artefatti tecnici, dai graffiti rupestri ai big data, dalla stampa alle stazioni spaziali per intenderci. In tal senso, l'infrasomatizzazione è il terzo termine tra l'endosomatico, cioè il vivente in generale, e l'esosomatico, ossia la forma propria dell'animale umano che crea organi artificiali esteriorizzando le proprie funzioni e facoltà. Se la tecnologia come *pharmakon* sistematicamente destituisce e costituisce l'umano, l'infrasomatizzazione è perciò comprensibile come l'effetto di ritorno farmacologico, al contempo di destituzione (di saperi, abitudini, facoltà, ecc.) e costituzione (idem), sui corpi e sulle coscienze, dei processi di esosomatizzazione. Volendo ora stringere i nodi dei concetti presentati precedentemente, ci sembra possibile indicare, da un lato, che ad essere infrasomatizzata negli individui è allora nientemeno che la *smartness*, e dall'altro, che l'autonomia del soggetto dipende dal gradiente di infrasomatizzazione digitale e dalle conoscenze che quest'ultimo deve sviluppare per farvi fronte:

> Per l'utente, queste infrasomatizzazioni avvengono attraverso smartphone e tablet che chiudono il circuito dal cervello all'ambiente esterno, in modo che l'apertura del pensiero sia mediata e compressa. Di conseguenza, la capacità del cervello umano di percepire che gli algoritmi stanno organizzando i suoi pensieri, o addirittura di percepire che gli algoritmi sono al lavoro, viene alterata o addirittura distrutta […], sovvertendo direttamente, e in casi estremi sostituendo, alcuni elementi dei processi cognitivi del pensiero e dell'esperienza umana[58].

Una volta spinto alle sue conseguenze infrasomatiche, il concetto di avvolgimento può allora rendere *sociale* l'esplicabilità, che deve concretizzarsi in un «diritto sociale alla spiegazione», in cui anche quest'ultima è appunto chiamata a essere sociale, nel senso di una funzione atta a portare l'insieme dei problemi infrasomatici, i quali si traducono in elementi di vulnerabilità cognitiva, a una forma di visibilità etico normativa. È in tal senso che, con Berry, riteniamo auspicabile una rinascita delle scienze umane per l'elaborazione di categorie e concetti in grado di trasformare l'esplicabilità in comprensibilità, vale a dire in un insieme olistico d'informazioni non solo tecniche o legali, bensì relative anche ai rapporti sociali, alle condizioni materiali di accesso ai dati, ai processi culturali, economici, simbolici

[57] Cfr. D. BERRY, *op. cit.*, 41-42.
[58] *Ivi*, 62.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

che stanno dietro alle operazioni di infrasomatizzazione, sia per ciò che concerne i diritti fondamentali che per i diritti digitali. Il senso del «diritto sociale alla spiegazione» sembra cioè rispondere alla necessità di comprendere in che modo i diritti fondamentali possono essere trasformati dalla digitalizzazione, ma anche i processi normativi in senso ampio che prendono piede dopo ogni innovazione tecnologica. Condizione perché si raggiunga una piena esplicabilità sociale, e dunque per il mantenimento di uno spazio critico, è «creare un "rifugio" per la ragione critica» che renda possibile una continua messa in discussione dello status quo e l'ideazione di nuovi percorsi democratici[59]. Tale rifugio, ossia uno spazio che sfugga all'avvolgimento, sarà dunque la condizione per sviluppare una critica di quella che si sta manifestando come una ragione computazionale, ma anche per una nuova epoca dei diritti e del diritto alla città intelligente.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

[59] *Ivi*, 63.

Data protection and AI compliance in health research: a relevant resource for institutions and companies against algorithmic vulnerability

*Giuseppe Claudio Cicu, Riccardo Michele Colangelo, Luca Saba**

DATA PROTECTION AND AI COMPLIANCE IN HEALTH RESEARCH: A RELEVANT RESOURCE FOR INSTITUTIONS AND COMPANIES AGAINST ALGORITHMIC VULNERABILITY

ABSTRACT: Artificial intelligence (AI) is becoming integral to health research, with applications in diagnosis, prognosis, and imaging segmentation across several medical fields. However, integrating health, biometric, and genetic data into AI systems raises ethical, legal, and practical challenges, particularly concerning discrimination and bias. Studies highlight the presence of bias, hindering AI model development in healthcare. Compliance with current legislation (e.g., GDPR), international frameworks (e.g., ISO), and forthcoming European AI regulation is pivotal. This paper emphasizes integrating these requirements into public entities and private organizations to ensure fair AI development and utilization in the health sector.

KEYWORDS: AI Act; GDPR; Compliance; Health Data; Algorithmic vulnerability.

ABSTRACT: L'intelligenza artificiale (AI) sta diventando parte integrante della ricerca sanitaria, con applicazioni nella diagnosi, nella prognosi e nella segmentazione delle immagini in diversi campi medici. Tuttavia, l'integrazione di dati sanitari, biometrici e genetici nei sistemi di IA solleva sfide etiche e giuridiche, in particolare per quanto riguarda *bias* e discriminazione. Diversi studi evidenziano la presenza di bias, che ostacolano lo sviluppo e l'impiego di modelli di IA nel settore sanitario. La conformità alla legislazione vigente (ad esempio, GDPR), agli standard internazionali (ad esempio, ISO) e alla normativa europea sull'IA è fondamentale. Questo articolo vuole sottolineare la necessità di una corretta implementazione di questi requisiti negli enti pubblici e nelle organizzazioni private per garantire uno sviluppo e un utilizzo corretto dell'IA nel settore sanitario.

PAROLE CHIAVE: AI Act; GDPR; Compliance; Dati sanitari; Vulnerabilità algoritmica.

* *Giuseppe Claudio Cicu: Ph.D. Student, University of Turin. Mail: giuseppeclaudio.cicu@unito.it; Riccardo Michele Colangelo: Ph.D. Student, Universitas Mercatorum, Rome. Mail: riccardomichele.colangelo@studenti.unimercatorum.it; Luca Saba, Full Professor of Diagnostic Imaging and Radiotherapy, University of Cagliari. Mail: lucasaba@unica.it. Giuseppe Claudio Cicu is author of paragraphs 2.1, 3, 3.1, 5, 7; Riccardo Michele Colangelo is author of paragraphs 2.2, 3, 3.2, 5, 7; and Luca Saba is author of paragraphs 1, 3, 4, 5, 6. The article was subject to a double-blind peer review process.*

Downloaded from www.biodiritto.org.
ISSN 2284-4503

SOMMARIO: 1. Introduction – 2. The regulatory background – 3. Voluntary frameworks and technical standards – 4. AI and bias in medical research – 5. Compliance and mitigation strategies – 6. Bridging the gap: from regulation to practice – 7. Conclusions.

1. Introduction

The advent and proliferation of Artificial Intelligence (AI) in the medical sector marks a pivotal transition in healthcare delivery and medical research[1]. AI's unparalleled ability to analyze vast datasets has unlocked innovative avenues for enhancing diagnostic accuracy, tailoring patient care, and streamlining healthcare workflows. These advancements are not only pivotal in managing complex diseases but also in predicting patient outcomes, thereby revolutionizing the landscape of medical care and research.

A quintessential example of AI's impact can be observed in the field of cardiovascular diseases, the leading cause of global morbidity and mortality[2]. Recent integrations of AI technologies in cardiovascular medicine have demonstrated promising results, ranging from improved diagnostic precision to nuanced patient risk assessments. By leveraging complex algorithms and machine learning models, researchers and clinicians are now better equipped to decode the intricate patterns of cardiovascular diseases, facilitating early detection and intervention.

However, as AI systems become more ingrained in healthcare processes, a spectrum of legal and ethical challenges emerges, particularly concerning data protection, privacy, and the potential for algorithmic bias[3]. The integration of health, biometric, and genetic data into AI systems raises substantial concerns about the safeguarding of fundamental rights. These concerns are exacerbated by evidence of varying AI algorithm performances across different racial and ethnic groups, which can lead to discrimination and bias, undermining the equity and fairness of healthcare services.

As AI continues to redefine the horizons of medical research and healthcare delivery, it is imperative to address these challenges head-on. Ensuring compliance with data protection laws, such as the General Data Protection Regulation (GDPR), adhering to international frameworks and technical standards along with the forthcoming European regulation on AI as well as statements by the competent supervisory authorities (e.g. the Decalogue of the Italian Data Protection Authority regarding health services and AI) is crucial. This paper aims to explore the legal and ethical considerations surrounding the use of AI in healthcare, with a particular focus on mitigating algorithmic bias and enhancing data protection. By exploring these dimensions, we strive to pave the way for a more equitable and responsible integration of AI technologies in the health sector, safeguarding the rights and well-being of individuals across diverse racial and ethnic backgrounds.

[1] M. MOOR, O. BANERJEE, Z. S. H. ABAD, H. M. KRUMHOLZ, J. LESKOVEC, E. J. TOPOL, P. RAJPURKAR, *Foundation models for generalist medical artificial intelligence*, in *Nature*, 616, 7956, 2023, 259–265.

[2] R. CAU, F. PISU, A. PINTUS, V. PALMISANO, R. MONTISCI, J. S. SURI, R. SALGADO, L. SABA, *Cine-cardiac magnetic resonance to distinguish between ischemic and non-ischemic cardiomyopathies: a machine learning approach*, in *European Radiology,* 34, 2024, 5691-5704.

[3] R. VANDERSLUIS, J. SAVULESCU, *The selective deployment of AI in healthcare: An ethical algorithm for algorithms*, in *Bioethics*, 38, 5, 2024, 391–400.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

2. The regulatory background

The integration of AI into healthcare not only promises to enhance medical research and patient care but, at the same time, also necessitates a rigorous legal and regulatory framework to address the myriad challenges it presents. This section provides an overview of the key laws, regulations, and standards governing data protection and AI in healthcare, emphasizing the importance of these frameworks in ensuring the ethical and secure use of AI technologies.

2.1. The European Artificial Intelligence Act

The spread of AI systems raises significant legal and ethical concerns, including issues related to privacy, transparency, accountability, discrimination, and bias. Consequently, the development and deployment of AI technologies require the implementation of appropriate legal rules to ensure trustworthy, accountable, and non-discriminatory access and utilization of such systems, especially when sensitive data categories such as health, genetic, and biometric data are involved[4].

These concerns have led the European Union to adopt a uniform legal framework that establishes harmonized rules on AI, aimed at improving the functioning of the internal market and promoting the uptake of human-centric and trustworthy AI, while ensuring a high level of protection for health, safety, and fundamental rights (the "EU AI Act")[5].

The subjective and objective scope of the regulation's content suggest that the European Union has also sought to achieve the so-called Brussels effect in relation to artificial intelligence. This term refers to the EU's ability to establish its legislation as a global standard within the international regulatory framework[6].

The EU AI Act introduces a risk-based classification of AI systems, aiming to balance technological innovation with the safeguarding of fundamental rights. Specifically, it categorizes AI applications into four risk levels: unacceptable risk, high risk, limited risk, and minimal risk.

AI systems that pose a clear threat to safety, livelihoods, or rights fall into the "unacceptable risk" category and are banned outright[7]. Examples of such systems include social scoring by governments and real-time biometric identification in public spaces.

"High risk" systems are subject to strict requirements and include applications in critical infrastructure, education, employment, essential private and public services, law enforcement, migration, and border control. The EU AI Act mandates rigorous testing procedures, documentation, compliance,

[4] *Study on Health Data, Digital Health and Artificial Intelligence in Healthcare*, Directorate-General for Health and Food Safety, European Commission, 16.

[5] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending regulations (EC) no 300/2008, (EU) no 167/2013, (EU) no 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and directives 2014/90/EU, (eu) 2016/797 and (EU) 2020/1828.

[6] B. Anu, *The Brussels Effect: How the European Union Rules the World* (Oxford University Press, 2020).

[7] See Art. 5, EU AI Act.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

and risk and quality management measures to ensure these systems are transparent, secure, and fair[8].

AI applications characterized by "limited risk" require specific transparency obligations. For instance, users must be informed when they are interacting with an AI system, allowing them to make informed decisions[9].

Several AI systems fall into the category of "minimal risk" and are subject to few legal requirements. Examples include AI-enabled video games and spam filters.

Under this risk-based approach, most AI applications adopted in the healthcare field fall into the high-risk category, reflecting the need to ensure high standards of patient safety, transparency, accountability, data privacy, and ethical standards.

Article 6.1 of the EU AI Act states that an AI system is classified as high-risk if it meets both of the following criteria, regardless of whether it is marketed or utilized independently of the products mentioned in points (a) and (b): «(a) The AI system is intended to be used as a safety component of a product, or it is a product itself, as specified by the Union harmonization legislation listed in Annex I. (b) The product, either the one whose safety component is the AI system mentioned in point (a) or the AI system itself as a product, must undergo a third-party conformity assessment before it can be marketed or put into service, in accordance with the Union harmonization legislation listed in Annex I».

With reference to the first condition, Annex I explicitly refer to Regulation (EU) 2017/745 ("MDR") related to medical devices. Regarding the second condition, Annex VIII, Chapter III, Rule 11 of the MDR (labeled "Classification") provides that software intended to provide information used to make decisions for diagnostic or therapeutic purposes is classified as class IIa, except if such decisions have an impact that may cause death or an irreversible deterioration of a person's state of health, in which case it is in class III; or a serious deterioration of a person's state of health or a surgical intervention, in which case it is classified as class IIb.

Software intended to monitor physiological processes is classified as class IIa, except if it is intended for monitoring vital physiological parameters, where the nature of variations in those parameters is such that it could result in immediate danger to the patient, in which case it is classified as class IIb.

Under these premises, such software often requires a third-party conformity assessment before it can be marketed or put into service. Thus, they fall within the application of the EU AI Act as high-risk systems when related to AI systems.

With reference to the health research activities, the EU AI ACT provides specific exclusions and instruments aimed at assuring that scientific research activities on AI systems are not undermined by the Regulation. Such provisions are without prejudice to the obligation to comply with this Regulation where an AI system falling within the scope of this Regulation is placed on the market or put into service as a result of such research and development activity and to the application of provisions on AI regulatory sandboxes and testing in real world conditions[10].

[8] See Art. 5, EU AI Act.

[9] See Art. 50, EU AI Act.

[10] See Whereas n. 25, EU AI Act.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

In this regard, the main exclusion regarding the research activities – therefore applicable to the health research sector - is set out in Art. 2.6. of AI EU ACT, which states that the EU AI ACT does not apply to AI systems or AI models, including their output, specifically developed and put into service for the sole purpose of scientific research[11].
Moreover, the Regulation also provides that its provisions do not apply to any research regarding AI systems or AI models prior to their being placed on the market or put into service, with the exclusion of testing in real world conditions[12].
Finally, in order to facilitate the involvement of relevant actors within the AI ecosystems, such as research and experimentation labs and individual researchers, the EU AI ACT provides the so called "AI regulatory sandboxes": controlled environments where innovative AI systems can be developed, trained, tested and validated for a limited time before their being placed on the market or put into service[13].

2.2 Data Protection and AI in the health research sector

The considerations set out regarding the EU AI Act must be enriched by a synthetic insight of some relevant data protection issues. All this, with the awareness that the rules of this Regulation neither solve specific problems nor fill gaps in the data protection regulatory framework, even though they apply to multiple sectors, including healthcare and health research[14].
With particular regard to the EU Regulation 2016/679 (General Data Protection Regulation), it is first of all necessary to highlight numerous references to the GDPR laid down in the EU AI Act: significantly, the number of such references increased during the AI Regulation approval[15].
This entails the need to consider both disciplines, in cases of any personal data processing carried out by automated means, among which AI systems are included in whole or in part.
This is all the truer with regard to the (AI) processing of special categories of personal data[16]: data, therefore, that can reveal data subject's vulnerabilities and expose him to discriminatory conducts.
This brief introduction underlines the importance of a reasoned and clear identification of the legal basis of the specific processing, since pursuant to art. 9, paragraph 1 GDPR the processing of special categories of personal data «shall be prohibited», unless there is (at least) one of the legal bases indicated in paragraph 2 of the same article.

[11] See Art. 2.6., EU AI Act.

[12] See Art. 2.9., EU AI Act.

[13] See Art. 57, EU AI Act.

[14] Cfr. P. Falletta, A. Marsano, *Intelligenza artificiale e protezione dei dati personali: il rapporto tra Regolamento europeo sull'intelligenza artificiale e GDPR*, in *Rivista italiana di informatica e diritto*, 1, 2024, 123.

[15] 30 references to EU Regulation 2016/679 are included in the final text of the EU AI Act.

[16] According to art. 9, par. 1 GDPR, special categories of personal data are «personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation».

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Here we can find not only the explicit consent given by the data subject (mandatory, for instance, for personal data processing made by healthcare apps[17]), but also the aim to protect the vital interests of the data subject. A processing of special categories of personal data can be considered lawful also when it «is necessary for the purposes of preventive or occupational medicine, for the assessment of the working capacity of the employee, medical diagnosis, the provision of health or social care or treatment or the management of health or social care systems and services» or «for reasons of public interest in the area of public health, such as protecting against serious cross-border threats to health or ensuring high standards of quality and safety of health care and of medicinal products or medical devices».

Regarding the health field, in 2023 the Italian Data Protection Authority (IDPA) set a Decalogue for the implementation of national health services through Artificial Intelligence[18] focusing, for instance, on the processing legal basis, the roles of natural or legal person involved in the specific processing and the importance of the privacy by design and by default principles. In this Decalogue, the IDPA[19] clarifies and sets out the principles of knowledge, not exclusivity and not algorithmic discrimination, considered as «three cardinal principles that must govern the use of algorithms and AI tools in the execution of relevant public interest».

The principle of knowledge regards «the right to know the existence of decision-making processes based on automated processing and, in this way, case, to receive significant information about the logic used, so as to be able to understand», while the principle of non-exclusivity of the algorithmic decision states that is necessary «in decision-making a human intervention capable of check, validate or deny the automatic decision» (c.d. human in the loop). Last but not least, we can find the crucial principle of algorithmic non-discrimination, meaning that «the data controller uses reliable AI systems that reduce opacity, the errors due to technological and/or human action, periodically checking efficiency also in the light of the rapid evolution of the technologies used, the appropriate mathematical or statistical procedures for profiling, setting out appropriate technical and organisational measures. This, including in order to ensure, the factors leading to inaccuracies in the data are corrected and minimised the risk of error, having regard to the potential discriminatory effects that inaccurate health data may determine against people (cf. recital 71 of the Regulation)».

The same Decalogue also states that, by means of interpretation, the GDPR requires that, in these cases of health data processing by AI, the information provided in compliance with the elements referred to in art. 13 and 14 of GDPR are not sufficient: data controllers have to highlight also “whether the processing is carried out in the learning phase of the algorithm (testing and validation) or in the next phase of application of the same, in the field of health services, representing data processing logics and characteristics; whether there are any obligations and responsibilities of health profession-

[17] R.M. COLANGELO, *App mediche e protezione dei dati personali. Alcuni spunti giuridici tra GDPR, codice privacy novellato e chiarimenti del Garante*, in *Autonomie locali e servizi sociali*, 2, 2019, 275-288.

[18] This Decalogue (in italian: *Decalogo per la realizzazione di servizi sanitari nazionali attraverso sistemi di Intelligenza Artificiale)* can be read on the Italian Data Protection Authority official website: https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9938038.

[19] *Ibidem*, par. 4 (author's translation).

Downloaded from www.biodiritto.org.
ISSN 2284-4503

als, to which the data subject is addressed, to use health services based on AI; the diagnostic and therapeutic benefits of using such new technologies"[20].

This insight confirms the primary role of the data protection regulation to prevent any form of discrimination related to AI systems, particularly in the health sector, not only before the full applicability of the EU AI Act, but also when the recent European Regulation wasn't in force. These considerations are particularly relevant in the context of scientific research, especially with reference to health research, highlighting how the transparency requirements for data subjects, based on the GDPR, must now be integrated - both in the public sector and by enterprises - with specific references to the artificial intelligence systems employed, as well as the stage of the lifecycle of such systems in which the personal data processing for health research purposes takes place.

In completion of these arguments, the healthcare sector, and particularly health research, is also subject, from a *de iure condendo* perspective, to the Italian "disegno di legge" n. 1146, titled Provisions and delegation to the Government on Artificial Intelligence.

Article 7 of the disegno di legge n. 1146 addresses the use of artificial intelligence in healthcare and disability, prohibiting discrimination (paragraphs 1 and 2) and establishing the «right of individuals to be informed about the use of artificial intelligence technologies and the benefits, in terms of diagnostics and therapy, resulting from the use of new technologies, as well as to receive information on the decision-making logic employed». It also highlights the supportive role of such systems (paragraph 5) and the necessity for their reliability, requiring that these systems be «periodically verified and updated to minimize the risk of errors» (paragraph 6).

Even more relevant in this context is Article 8, regarding Research and Scientific Experimentation in the Development of Artificial Intelligence Systems in Healthcare. This article establishes that «data processing, including personal data, carried out by public and private entities for non-profit purposes in research and scientific experimentation in the development of artificial intelligence systems» in healthcare (for the purposes of disease prevention, diagnosis, treatment, drug development", and other specific objectives) «are declared of significant public interest in accordance with Article 32 of the Constitution and in compliance with Article 9, paragraph 2, letter g), of Regulation (EU) 2016/679 of the European Parliament and the Council, of April 27, 2016». Consequently, the legal basis for such processing is established by law where processing is necessary for reasons of substantial public interest, on the basis of Union or Member State law. This implies, as a general rule, that, if the law is approved without amendments, consent will not be required in health research conducted by private entities too.

The following paragraph also authorizes, «without further consent from the data subject where initially required by law, the secondary use of personal data devoid of direct identifiers, including data belonging to the categories referred to in Article 9 of Regulation (EU) 2016/679, by the entities referred to in paragraph 1».

The specific data processing, as outlined in paragraphs 1 and 2, must, however, be approved by the competent ethics committees and communicated to the Italian Data Protection Authority with specific formal requirements. Following this communication, «processing may begin thirty days after the

[20] *Ibidem,* par. 8 (author's translation).

Downloaded from www.biodiritto.org.
ISSN 2284-4503

aforementioned communication unless blocked by a decision from the Data Protection Authority» (paragraph 3).
Another fundamental aspect that involves a partial overlap between data protection and AI legislation is related to the exercise of the rights of the data subject enshrined in the GDPR.
In particular, art. 22, par. 1, GDPR[21], regarding automated individual decision-making processes, including profiling, provides that «the data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her». In the following paragraph some exceptions are mentioned but is fundamental to underline that paragraph 3 states the implementation of «suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision». Art. 22, par. 4, GDPR also specifies the instances where automated decisions based on special categories of personal data could be legal: in these circumstances, «suitable measures to safeguard the data subject's rights and freedoms and legitimate interests» should be taken.

3. Voluntary frameworks and technical standards

Beyond formal regulations, international voluntary frameworks play a crucial role in guiding the responsible use of AI in healthcare. Initiatives like OECD's Principles on AI and the G20's AI Guidelines advocate for principles such as inclusivity, transparency, and accountability[22]. Other interesting voluntary frameworks to ensure a lawful and ethical implementation of AI systems are the ISO technical standard and the AI Pact, which will be described in the following paragraphs 3.1 e 3.2.

3.1. Some considerations regarding ISO applicable to the AI field

The International Organization for Standardization (ISO) plays a key role in shaping the use of AI in healthcare through the development of voluntary international standards. These standards cover various aspects of AI, partly aligned with the provisions of the EU AI Act, including data quality, security, and interoperability, providing guidelines for the ethical and effective implementation of AI technologies. Particularly, ISO/IEC 42001 provides technical standards such as (i) logging and record-keeping as one of the optional controls to consider for implementing risk treatment options, (ii) transparency, with a focus on providing information to users, and (iii) quality management systems, as a high-level standard. Other relevant international standards include ISO/IEC 23894 on AI Risk Management, ISO/IEC 5259, which describes a data quality model for data analytics and AI based on machine

[21] The specific right under art. 22 GDPR is considered better suited to protect the rights of natural persons, while there are no similar effective redress mechanisms in the EU AI Act: cf. O. Pollicino, G. De Gregorio, *Intelligenza artificiale, data protection e responsabilità*, in A. Pajno, F. Donati, A. Perrucci (eds.), *Intelligenza artificiale e diritto: una rivoluzione?*, Bologna, 2022, 355.
[22] M. Rotenberg, *Human Rights Alignment: The Challenge Ahead for AI Lawmakers. In Introduction to Digital Humanism*, Cham, 2024, 611–622.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

learning, and the ISO/IEC 24029 series that provides robustness metrics for supervised classification/ regression models using statistical and empirical approaches[23].
Following such standards, corporate organizations and public bodies may anticipate implementing technical and ethical measures for AI adoption also in the health sector and in the medical research field. However, it should be noted that most of the mentioned international standards do not guarantee - or guarantee only partially - compliance with the provisions of the EU AI Act.

3.2. The EU AI Pact

In addition to the aforementioned technical standards, it is necessary to consider how the European Commission intends to promote and anticipate an effective and appropriate compliance with the EU AI Act, which, as has been stated, aims to avoid as much as possible violations of the fundamental rights of the persons involved and any discrimination on the basis of any erroneous bias by AI systems.
In this regard, the EU AI Act regulates the role of the European AI Office, already established from 24 January 2024 within the European Commission, which is responsible for «contributing to the implementation, monitoring and supervision of AI systems and AI models for general purposes, and AI governance»[24]. This Office now promotes the AI Pact[25], which is a recent initiative of the European Union, and more precisely of the European Commission, that intends to stimulate - also in this case at a completely voluntary level - the proactive adherence to the new discipline on AI, especially before it comes into force[26]. In short, this initiative underlines the importance of considering the AI legislation now in force, although not yet applicable and fully binding, encouraging the adherence to the principles established in it even before (and in view) the full applicability of the EU AI Act as a whole.
The AI Pact is directly aimed at organizations, enterprises and companies - which can be involved in the processing of health data both as data controllers and as data processors - and underlines the growing importance of the development and correct use of AI systems also in the context of business activities, in order to protect individuals (and data subjects) in conditions of vulnerability and therefore to prevent any algorithmic discrimination.
The AI Pact is based on two pillars: the collection and exchange of best practices and information on the EU AI Act implementation process in the specific fields of the AI Pact network, as well as facilitating the commitments of enterprises and companies, so as to encourage both providers and deployers to prepare in time, taking the necessary measures and actions towards (future) compliance with the European framework on AI and its requirements and obligations.
This approach, which highlights one of the shortcomings of a regulation that is likely to be old[27]. It is still agreeable, as it helps to consider how it is not possible - as it was not before the final approval of

[23] For a comprehensive examination of the operational areas of ISO with reference to artificial intelligence, see *Analysis of the preliminary AI standardisation work plan in support of the AI Act*, JRC Technical Report, European Commission, 2023.
[24] Art. 3, par. 1, n. 4, EU AI Act.
[25] Available at: https://digital-strategy.ec.europa.eu/en/policies/ai-pact.
[26] The European Commission, through its Office, is «seeking the industry's voluntary commitment to anticipate the AI Act and to start implementing its requirements ahead of the legal deadline» (ibidem).

Downloaded from www.biodiritto.org.
ISSN 2284-4503

the EU AI Act - to say that the development and use of AI systems is completely free from any regulatory constraint. This is particularly true in the context of health-related data processing implemented through AI systems, regardless of the purposes of the processing and the public or private nature of the data processor or data controller.

4. AI and bias in medical research

For the purpose of this paper, we will focus on the cardiovascular field that represents the first cause of death worldwide[28]. The integration of AI in cardiovascular medicine has been marked by both significant achievements and challenges, particularly concerning biases that impact diagnostic accuracy across different racial and ethnic groups.

4.1. Successes in cardiovascular medicine

One of the most notable successes of AI in cardiovascular medicine is its ability to enhance diagnostic precision. For instance, deep learning models have been employed to interpret magnetic resonance, significantly improving the detection of pathologies[29]. These models analyze patterns in Magnetic Resonance images with a level of detail and accuracy that surpasses conventional methods, leading to earlier and more accurate diagnoses. AI algorithms have also been instrumental in developing predictive models for cardiovascular diseases. By analyzing vast datasets, including electronic health records and genetic information, AI models can predict individuals' risk of developing cardiovascular diseases. This predictive capability enables targeted preventive measures and personalized treatment plans, improving patient outcomes[30].

4.2. Limitations and biases

Despite these successes, the application of AI in cardiovascular medicine has been hampered by significant limitations, particularly biases that affect diagnostic accuracy across racial and ethnic groups[31]. Studies have demonstrated that AI models can exhibit biases that lead to discrepancies in

[27] It should be noted that the relationship between law and new technologies is typically defined by the legislator's delay: cf. R. Rolli, *Il Diritto privato nella società 4.0*, Milano, 2018, XVIII-XIX and M. Pietrangelo, *La società dell'informazione tra realtà e norma*, Milano, 2007, 176.

[28] M. Naghavi, K. L. Ong, A. Aali, H. S. Ababneh, Y. H. Abate, C. Abbafati, R. Abbasgholizadeh, M. Abbasian, M. Abbasi-Kangevari, H. Abbastabar, S. Abd ElHafeez, M. Abdelmasseh, S. Abd-Elsalam, A. Abdelwahab, M. Abdollahi, M. Abdollahifar, M. Abdoun, D. M. Abdulah, A. Abdullahi, C. J. L. Murray, *Global burden of 288 causes of death and life expectancy decomposition in 204 countries and territories and 811 subnational locations, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021*, in *The Lancet,* 403, 2024.

[29] Y. R. Wang, K. Yang, Y. Wen, P. Wang, Y. Hu, Y. Lai, Y. Wang, K. Zhao, S. Tang, A. Zhang, H. Zhan, M. Lu, X. Chen, S. Yang, Z. Dong, Y. Wang, H. Liu, L. Zhao, L. Huang, S. Zhao, *Screening and diagnosis of cardiovascular disease using artificial intelligence-enabled cardiac magnetic resonance imaging*, in *Nature Medicine*, 2024.

[30] D. Gala, H. Behl, M. Shah, A. N. Makaryus, *The Role of Artificial Intelligence in Improving Patient Outcomes and Future of Healthcare Delivery in Cardiology: A Narrative Review of the Literature*, in *Healthcare*, 12(4), 2024, 481.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

diagnostic accuracy[32]. For example, an AI system developed for diagnosing heart disease showed higher sensitivity in identifying conditions in White patients compared to Black patients[33]. This discrepancy arises from the model being trained predominantly on data from White individuals, leading to less accurate predictions for other racial groups. The legal and ethical implications of such biases are profound. From a legal perspective, these biases may violate principles of non-discrimination and equity, as enshrined in regulations like the GDPR and forthcoming EU regulations on AI. Ethically, they raise concerns about fairness and the moral obligation to provide equitable healthcare services to all patients, irrespective of their racial or ethnic background.

4.3. Ethical and legal implications

The existence of bias in AI models used in cardiovascular medicine indicates the urgent need for frameworks that ensure the ethical development and deployment of AI. Legally, it necessitates adherence to principles of fairness and equity, requiring that AI models be developed and tested on diverse datasets that accurately reflect the population's heterogeneity. Ethically, it demands a commitment to minimizing harm and ensuring that AI technologies benefit all segments of society equally. To address these challenges, it is crucial to implement bias detection and mitigation strategies throughout the AI development lifecycle. This includes diversifying training datasets, employing fairness-enhancing algorithms, and conducting rigorous testing across diverse population groups. Additionally, transparency in AI development processes and outcomes is essential to build trust and ensure accountability.

5. Compliance and mitigation strategies

Ensuring compliance with the intricate web of legal requirements and ethical guidelines for the use of AI in healthcare is a complex yet crucial task. Central to this effort is adherence to the GDPR for entities operating within or dealing with data from the European Union, which mandates strict data protection and privacy practices. Similarly, international standards such as those developed by the International Organization for Standardization (ISO) offer guidance on maintaining data security, quality, and interoperability in AI systems. These frameworks, alongside various national and international guidelines, establish a foundation for ethical AI use that respects privacy, ensures fairness, and promotes transparency.

[31] Z. Javed, M. Haisum Maqsood, T. Yahya, Z. Amin, I. Acquah, J. Valero-Elizondo, J. Andrieni, P. Dubey, R. K. Jackson, M. A. Daffin, M. Cainzos-Achirica, A. A. Hyder, K. Nasir, *Race, Racism, and Cardiovascular Health: Applying a Social Determinants of Health Framework to Racial/Ethnic Disparities in Cardiovascular Disease*, in*Circulation: Cardiovascular Quality and Outcomes*, 15(1), 2022.

[32] E. Tat, D. L. Bhatt, M. G. Rabbat, *Addressing bias: artificial intelligence in cardiovascular medicine*, in *The Lancet Digital Health*, 2(12), 2020.

[33] D. Kaur, J. W. Hughes, A. J. Rogers, G. Kang, S. M. Narayan, E. A. Ashley, M. V. Perez, *Race, Sex, and Age Disparities in the Performance of ECG Deep Learning Models Predicting Heart Failure*, in *Circulation: Heart Failure*, 17(1), 2024.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

For healthcare institutions and companies aiming to align their AI systems with these requirements, a multifaceted approach to compliance and bias mitigation is essential. This begins with the comprehensive mapping of AI applications against existing legal frameworks to identify specific compliance obligations. Following this, a thorough risk assessment process can highlight potential areas where AI systems might breach data protection norms or introduce bias in healthcare delivery.

Tools that can be used to enhancing protection of personal rights are the so-called privacy enhancing technologies ("PETs"). PETS are a collection of digital solutions aim at collecting, processing, analysis and sharing information while protecting the confidentiality of personal data[34][1]. PETs can be divided into three categories: data obfuscation, encrypted data processing, and federated and distributed analytics. Data obfuscation tools include zero-knowledge proofs (ZKP), differential privacy, synthetic data, anonymisation and pseudonymisation tools. These tools increase privacy protections by altering the data, by adding "noise" or by removing identifying details. Among them, differential privacy has been successfully applied to several large-scale biomedical data sharing initiatives, including the UK Biobank and the National Institutes of Health's All of Us research program. Concurrently, synthetic data has emerged in the healthcare sector as a powerful tool for analysis and technology development[35]. Synthetic data are data created from real datasets with similar statistical properties, enhancing privacy while allowing researchers to conduct meaningful analyses. This approach has been utilized in various contexts, such as simulation and prediction research[36], algorithm testing[37], public health research[38], and so on. Encrypted data processing tools include homomorphic encryption, multi-party computation, and trusted execution environments. Encrypted data processing PETs allow data to remain encrypted while in use (in-use encryption) and thus avoiding the need to decrypt the data before processing. For example, encrypted data processing tools were widely deployed in Covid tracing applications. Federated and distributed analytics, including federated and distributed learning, allows executing analytical tasks upon data that are not visible or accessible to those executing the tasks. In federated learning, for example, a technique gaining increased attention, data are preprocessed at the data source. In this way, only the summary statistics/results are transferred to those executing the tasks.

Another effective strategy for mitigating these risks is the incorporation of privacy by design principles from the outset of AI system development[39]. This approach ensures that data protection measures are not afterthoughts but are integrated into the core architecture of AI applications. Similarly,

[34] OECD, *Emerging privacy enhancing technologies current regulatory and policy approaches*, in *OECD digital economy papers*, 351, March 2023.

[35] A. Gonzales, G. Guruswamy, S. R. Smith, *Synthetic data in health care: A narrative review*, in *PLOS Digit Health*,2, 1, 2023.

[36] P. Davis, R. Lay-Yee, J. Pearson, *Using micro-simulation to create a synthesised data set and test policy options: The case of health service effects under demographic ageing,* in *Health Policy*, 97, 2–3, 2010, 267.

[37] C. Ngufor, H. Van Houten, B. S. Caffo, N. D. Shah, R. G, McCoy R.G., *Mixed effect machine learning: A framework for predicting longitudinal change in hemoglobin A1c*, in *Biomed Inform.*, 89, 2019, 56–67.

[38] W. T. Enanoria, F. Liu, J. Zipprich, K. Harriman, S. Ackley, S. Blumberg, *The Effect of Contact Investigations and Public Health Interventions in the Control and Prevention of Measles Transmission: A Simulation Study*, in *PloS One*, 11, 12, 2016.

[39] S. Reddy, S. Allan, S. Coghlan, P. Cooper, *A governance model for the application of AI in health care,* in *Journal of the American Medical Informatics Association*, 27, 3, 2020, 491–497.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

implementing rigorous data governance policies helps safeguard patient information, ensuring that data collection, storage, and processing activities comply with legal standards.

Bias mitigation requires a proactive stance, starting with the diversification of datasets to reflect the heterogeneity of the population accurately. This involves not only the inclusion of diverse demographic groups in the data but also the careful annotation of data to identify potential sources of bias. Advanced analytical techniques can then be employed to detect and correct for biases, ensuring that AI models perform equitably across different patient groups.

Beyond technical measures includes regular training for staff on the ethical implications of AI and the establishment of clear guidelines for responsible AI research and development. Ethical review boards, similar to those used in medical research, can provide oversight for AI projects, evaluating them for potential ethical concerns and compliance with legal standards.

Public research bodies and corporate entities alike must integrate these legal and ethical considerations into their AI development processes through continuous engagement with stakeholders, including patients, healthcare professionals, and legal experts. Such engagement ensures that AI systems are developed with a clear understanding of the legal landscape and ethical expectations, facilitating compliance and promoting the responsible use of AI in healthcare.

Through these strategies, organizations can navigate the complexities of AI compliance, transforming legal and ethical challenges into opportunities for innovation in healthcare. By prioritizing data protection, bias mitigation, and ethical considerations, healthcare institutions and companies can leverage AI to enhance patient care, improve healthcare outcomes, and uphold the highest standards of fairness and respect for patient privacy.

6. Bridging the gap: from regulation to practice

The task of aligning the regulatory corpus with the practical exigencies of health research and service delivery is a complex yet essential undertaking for entities operating within the health sector. This alignment is necessary not only for ensuring legal compliance but also for harnessing the full potential of AI in advancing healthcare. To bridge this gap effectively, a comprehensive approach that encompasses policy development, stakeholder engagement, and the establishment of robust oversight mechanisms is required.

Entities can begin by conducting a thorough analysis of how existing regulations impact their operations and identifying any areas where AI applications could potentially lead to non-compliance or ethical dilemmas. This initial assessment should serve as the basis for developing tailored AI governance policies that address specific regulatory and ethical concerns while also meeting the operational needs of healthcare delivery. Such policies should outline clear procedures for data handling, consent management, algorithmic transparency, and bias mitigation, ensuring that all aspects of AI use are covered.

Engagement with stakeholders is another fundamental aspect of bridging the regulatory and practical scenario. This includes not only healthcare professionals and patients but also legal experts, ethicists, and regulators. By fostering open dialogues, entities can gain diverse perspectives on the practical challenges of implementing AI in healthcare settings, identifying common concerns and collabora-

Downloaded from www.biodiritto.org.
ISSN 2284-4503

tive solutions. Stakeholder input can also guide the development of AI applications that are not only compliant with legal and ethical standards but are also aligned with patient care priorities and clinical needs.

To sustain compliance and address ongoing legal, ethical, and practical challenges, a dynamic framework for the monitoring and assessment of AI systems in healthcare is indispensable. Such a framework should include:

- Continuous Monitoring: Regular audits of AI systems to ensure they operate as intended and do not deviate from compliance requirements or ethical norms. Monitoring should also include the tracking of data sources and algorithmic decisions to identify any emergent biases or privacy concerns.
- Impact Assessment: Periodic evaluations of the impact of AI applications on patient outcomes, healthcare equity, and operational efficiency. These assessments can help identify areas where AI is delivering value, as well as those where it may be falling short or inadvertently introducing disparities.
- Adaptive Governance: Mechanisms for revising AI policies and practices in response to new regulatory developments, technological advancements, or changes in healthcare delivery models. Adaptive governance ensures that AI applications remain relevant and beneficial in the face of evolving healthcare landscapes.
- Stakeholder Feedback Loops: Regular opportunities for feedback from healthcare professionals, patients, and other stakeholders to inform the ongoing development and refinement of AI applications. This feedback can provide practical insights into how AI is affecting healthcare delivery and patient care, guiding improvements and adjustments.

By implementing such a framework, entities in the health sector can navigate the complexities of applying AI in healthcare, ensuring that their innovations not only comply with legal and ethical standards but also meet the practical needs of health research and service delivery. This approach fosters an environment where AI can be leveraged responsibly and effectively to improve health outcomes, enhance patient care, and advance the frontiers of medical knowledge.

7. Conclusions

The integration of AI into healthcare holds the potential to transform not only patient care, but also medical research profoundly. However, realizing this potential fully requires us to navigate the complex landscape of legal and ethical challenges diligently.

Effective and non-formal compliance is necessary to maximize the potential of AI in healthcare data processing and capitalize on all opportunities for risk management, beyond simply complying with regulatory requirements.

Therefore, it is becoming increasingly necessary not only to operate from a privacy by design perspective, but to integrate this one with the AI regulatory framework, although not yet applicable.

Taking a proactive approach towards the AI Act, considering it already as a reference model despite not being entirely binding yet, represents a forward-looking and future-proof strategy for the integration of AI technologies

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Furthermore, by pushing collaboration among all stakeholders involved and committing to ongoing research and dialogue, we can ensure that AI serves as a force for good in healthcare, enhancing the wellbeing of individuals and communities worldwide. This balanced approach to innovation will pave the way for a future where AI not only revolutionizes healthcare but does so in a manner that is just, equitable, and respectful of the rights and dignity of all individuals.

Addressing the legal and ethical implications of AI in healthcare, specifically in health research, is vital for ensuring that these technologies benefit all patients equally. By implementing robust compliance and bias mitigation strategies, healthcare institutions and companies can leverage AI and research findings achieved through such systems to enhance patient care, improve health outcomes, and uphold the highest standards of fairness and privacy. The ongoing collaboration between stakeholders, including regulators, healthcare professionals, and patients, will be crucial in the complexities of AI integration and fostering an environment where AI can be used responsibly and effectively to advance medical knowledge and healthcare delivery.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Studi clinici, discriminazioni razziali e intelligenza artificiale: *diversity and inclusion* nel contesto statunitense

Vanessa Lando*

HEALTH-DATA POVERTY, RACIAL DISCRIMINATION AND ARTIFICIAL INTELLIGENCE: DIVERSITY AND INCLUSION IN CLINICAL TRIALS

ABSTRACT: The use of biased artificial intelligence systems in the healthcare field involves the risk to crystallize and exacerbate existing health inequalities. The discriminatory functioning of the algorithm arises, in part, because of the use of an inadequate dataset. The study aims to explore the correlation between the low participation of ethno-racial minorities in clinical trials and the inability of these subjects - already in a condition of subalternity - to benefit from data-driven innovation developed, also, with data from such clinical trial. Moving from the analysis of US context, the paper will highlight the importance of a diversity and inclusion approach in the selection of the sample and in the conduction of the clinical trial, in order to promote equal access to healthcare even - and especially - when this is provided through the use of AI systems.

KEYWORDS: Artificial intelligence; health equity; dataset; ethno-racial minorities; USA.

ABSTRACT: L'utilizzo, all'interno dell'ambito medico-sanitario, di sistemi di intelligenza artificiale affetti da *bias* porta con sé il rischio di cristallizzare le già presenti diseguaglianze nell'ambito della salute. Il funzionamento discriminatorio dell'algoritmo può essere ricondotto, seppur non esclusivamente, all'utilizzo di un set di dati non adeguato. Il contributo mira a mettere in luce la correlazione tra la scarsa partecipazione delle minoranze etno-razziali ai trial clinici e l'impossibilità da parte di tali soggetti - già di per sé in una condizione di subalternità - di beneficiare delle innovazioni tecnologiche sviluppate anche con i dati provenienti da tali attività di ricerca. Si evidenzierà, impiegando come ambito di analisi il contesto statunitense, la necessità di un approccio di *diversity and inclusion* nella scelta del campione e nella conduzione del trial clinico, al fine di promuovere l'equo accesso all'assistenza sanitaria anche – e soprattutto – quando questa viene veicolata tramite l'utilizzo di sistemi di IA.

PAROLE CHIAVE: Intelligenza artificiale; health equity; dataset; minoranze etno-razziali; Stati Uniti d'America.

* *Assegnista di ricerca, Università di Trento. Mail: vanessa.lando@unitn.it.* *Contributo sottoposto a doppio refereraggio anonimo.*

Downloaded from www.biodiritto.org.
ISSN 2284-4503

artificiale – 3. Gli USA e i tentativi di *diversity and inclusion* nei trial clinici – 4. Conclusioni.

1. Dataset, bias e discriminazione algoritmica

Si è recentemente assistito ad una rapida implementazione dei sistemi di intelligenza artificiale (IA) per la sanità con innumerevoli benefici nel campo della diagnostica e della cura, della capacità di allocazione delle risorse e della gestione di problematiche correlate alla tutela della salute pubblica[1]. Gli algoritmi di intelligenza artificiale devono il loro prestigio alla connaturata abilità di estrarre informazioni da grandi moli di dati con velocità e precisione esorbitanti le capacità umane. Tale capacità è strettamente correlata all'attività di *data-training* intesa come esposizione dell'algoritmo ad innumerevoli esempi la cui bontà influenza in modo determinante la performance dell'algoritmo[2]. In questo senso l'accuratezza, la validità e la solidità dell'algoritmo dipendono dalla qualità degli *input* forniti allo stesso[3]. Gli algoritmi di intelligenza artificiale sono infatti interessati dal flusso *garbage in – garbage out* per effetto del quale, se i dati forniti all'IA sono incongrui, inesatti o non affidabili, allora anche le decisioni assunte dall'algoritmo sulla base di questi dati saranno incongrue, inesatte e inaffidabili[4]. Questa dinamica viene definita *bias*. Il termine è utilizzato dai *computer scientists* per indicare un qualsiasi malfunzionamento o problematica strettamente connessa con l'operatività algoritmica[5]. Ai fini della presente analisi è tuttavia necessario esaminare il concetto di *bias* attraverso la lente della non discriminazione[6]. In questa ottica il *bias* costituisce un errore di valutazione o di formulazione di giudizio, dovuto ad assunzioni errate nel processo di apprendimento automatico, che comporta l'emissione di un *output* capace di avvantaggiare o svantaggiare un individuo o un gruppo di individui senza alcuna motivazione in grado di giustificare tale differenza di trattamento[7]. La creazione del set di dati di addestramento viene identificato come momento del *bias*, e cioè come uno dei momenti della fase di creazione e realizzazione del modello in cui maggiormente si

[1] Per una ampia disamina degli impieghi dei sistemi di intelligenza artificiale in ambito medico sanitario si vedano P. RAJPURKAR, E. CHEN, O. BANERJEE, E. TOPOL, *AI in health and medicine*, in *Nature Medicine*, 28, 2022, 31-38; E. TOPOL, *High-performance medicine: the convergence of human and artificial intelligence*, in *Nature Medicine*, 25, 2019, 44-56.

[2] Per una disamina tecnica si veda S. RUSSEL, P. NORVING, *Intelligenza artificiale – Un approccio moderno*, trad. it. F. AMIGONI, Milano – Torino, 2021.

[3] J. GERARDS, R. XENIDIS, *Algorithmic Discrimination in Europe – Challenges and opportunities for gender equality and non discirimination law*, Lussemburgo, 2021, 42.

[4] G. RESTA, *Governare l'innovazione tecnologica: decisioni algoritmiche, diritti digitali e principio di uguaglianza*, in *Politica del diritto*, 2, 2019, 208 e P. ZUDDAS, *Intelligenza artificiale e discriminazioni*, in AA.VV. (a cura di), *Liber amicorum per Pasquale Costanzo – Diritto Costituzionale in trasformazione Vol. I – Costituzionalismo, Reti e Intelligenza artificiale*, Genova, 2020, 463.

[5] R.K.E. BELLAMY ET AL., *AI Fairness 360: an extensible toolkit for detecting, understanding, and mitigating unwanted algorithmic bias,* in *IBM Journal of Research and Development,* Settembre 2019, 1*;* B. FRIEDMAN, H. NISSENBAUM, *Bias in Computer Systems,* in *ACT Transactions on Information Systems*, 14, 3, 1996, 330-347.

[6] J. GERARDS, R. XENIDIS, *op. cit*., 47; D. PESSACH, E. SHMUELI, *Algorithmic Fairness*, 2020, disponibile in https://arxiv.org/abs/2001.09784 (ultima consultazione 23/06/2024).

[7] E. STRADELLA, *Stereotipi e discriminazioni: dall'intelligenza umana all'intelligenza artificiale*, in AA.VV. (a cura di), *Liber amicorum per Pasquale Costanzo – Diritto Costituzionale in trasformazione Vol. 1 – Costituzionalismo, Reti e Intelligenza artificiale,* 2020, 392.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

concretizza il rischio di introduzione di un fattore potenzialmente causa di discriminazione[8]. La correlazione tra scarsa qualità del *dataset* utilizzato per l'addestramento e il fenomeno della discriminazione algoritmica nell'equa tutela della salute viene evidenziato in numerosa letteratura scientifica[9]. Ciò che accomuna gli studi è la certezza che l'emissione di un *output* errato (*rectius* discriminatorio) sia, almeno in parte, da ricondurre all'assenza - all'interno del dataset utilizzato per addestrare l'algoritmo - di dati provenienti da determinati segmenti di popolazione. Si tratta di vere e proprie zone d'ombra[10] in quanto, nonostante si ritenga che i dati siano capaci di rappresentare in modo oggettivo la realtà, vi sono divari significativi nell'emissione di dati da parte di alcune comunità rispetto ad altre[11]. Nell'ambito di indagine del presente contributo, tale assenza viene definita *health data disparity,* intesa quale la sistematica differenza nella qualità e/o nella quantità di dati relativi allo stato di salute riconducibili a determinati gruppi[12]. I soggetti tipicamente interessati da tale circostanza sono, evidentemente, quelli che rientrano nelle c.d. minoranze[13]: individui appartenenti al medesimo gruppo in quanto accomunati da una stessa caratteristica socialmente saliente[14], che, a causa delle dinamiche discriminatorie, si trovano in una posizione di soggiogamento, di subalternità e di vulnerabilità tale da trasformare le caratteristiche socialmente salienti in un fattore di protezione considerato, dal diritto

[8] Secondo la letteratura più recente possono essere individuati cinque tempi del bias. Questi sono: (i) individuazione delle *target variables* e delle *class labels*; (ii) il *data training*; (iii) la *feature selection*; (iv) l'individuazione dei possibili *proxy*; (v) il *masking*. Si vedano F. Z. BORGESIOUS, *Discrimination, Artificial Intelligence and algorithmic decision-making*, Strasburgo, 2018, 15-23 e S. BAROCAS, A.D. SELBST, *Big Data's Disparate Impact*, in *California Law Review*, 2016, 677-692.

[9] Si vedano A. S. ADAMSON, A. SMITH, *Machine learning and Health Care Disparities in Dermatology*, in *JAMA Dermatology*, 154, 2018, 1247-1248; D. WEN ET AL., *Characteristics of publicly available skin cancer image datasets: a systematic review*, in *The Lancet Digital Health*, 4, 2022, 64-74; I. STRAW, H. WU, *Investigating for bias in health care algorithms: a sex stratified analysis of supervised machine learning models in liver disease prediction*, in *BMJ Health & Care Informatics*, 29, 1, 2022, 100457.

[10] K. CRAWFORD, *Think Again: Big data – Why the rise of machines isn't all it's cracked up to be*, in *Foreign policy*, 10 maggio 2013, disponibile in https://foreignpolicy.com/2013/05/10/think-again-big-data/ (ultima consultazione 26/11/2024).

[11] K. CRAWFORD, *The hidden biases in big data*, in *Harvard Business Review*, 1 aprile 2013, disponibile in https://hbr.org/2013/04/the-hidden-biases-in-big-data (ultima consultazione 26/11/2024).

[12] H. IBRAHIM ET AL., *Health data poverty: an assailable barrier to equitable digital health care*, in *Lancet Digital Health*, 3, 2021, 260.

[13] Per un approfondimento sul concetto di minoranza si vedano F. PALERMO, J. WOELK, *Diritto costituzionale comparato dei gruppi e delle minoranze*, Milano, 2021; C. NARDOCCI, *Razza e etnia: la discriminazione tra individuo e gruppo nella dimensione costituzionale e sovranazionale*, Napoli, 2016; D. HELLMAN, *What is discrimination wrong?*, Cambridge (MA), 2008.

[14] Secondo Kasper Lippert-Rasmussen con il termine "caratteristica socialmente saliente" si fa riferimento all'elemento esteriore o interiore dell'individuo che lo accomuna ai membri di un gruppo e che è rilevante e significativo in ciascuna interazione sociale. Si veda K. LIPPERT-RASMUSSEN, *The Badness of Discrimination*, in *Ethical Theory and Moral Practice*, 9, 2006, 169; K. LIPPERT-RASMUSSEN, *Born Free and Equal? – A Philosophical Inquiry into the Nature of Discrimination*, Oxford, 2014.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

antidiscriminatorio[15], come la peculiarità del soggetto sulla quali è possibile fondare una richiesta di tutela nel caso in cui al possesso – o all'assenza – siano riconducibili atti discriminatori[16].

L'*health data disparity* e strettamente connessa con l'*Health data poverty*, ovverosia quella condizione per cui i gruppi o gli individui insufficientemente rappresentati all'interno degli *health dataset* presentano minore probabilità di poter godere delle innovazioni tecnologiche intelligenti guidate dai dati e sviluppate utilizzando il *dataset* stesso[17]. Ciò che accade è che questi soggetti – già di per sé in una condizione di subalternità e vulnerabilità – si vedono esclusi da un progresso tecnologico capace di apportare grandi miglioramenti alla salute tanto individuale quanto collettiva. Le disuguaglianze nell'ambito della salute sono pervasive e sbarrano la strada al raggiungimento di un ideale di uguaglianza che si sostanzia anche nell'ottenimento del più alto livello di salute per ciascuno dei consociati[18] [19]. Per scardinare tale circolo vizioso, e avviare un processo di eliminazione delle disparità in materia di salute e assistenza sanitaria, appare necessario rivolgersi a tutti i consociati, con sforzi mirati e continui orientati dal principio di *diversity and inclusion*[20].

Traslando quanto sopra all'interno del campo delle tecnologie guidate dai dati si comprende come, affinché tutti possano equamente beneficiare delle innovazioni apportate dall'introduzione dei sistemi di IA in ambito medico-sanitario sia necessario che il *dataset* utilizzato per l'addestramento dell'algoritmo sia demograficamente inclusivo, cioè rappresentativo della popolazione su cui il sistema di IA

[15] Si veda S. SCARPONI, *I divieti di discriminazione fra diritto europeo e nazionale,* in AA.VV. (a cura di), *Corso di Diritto antidiscriminatorio – materiali per la formazione,* Bologna, 2014; E. CONSIGLIO, *Che cos'è la discriminazione? Un'introduzione teorica al diritto antidiscriminatorio*, Torino, 2020.

[16] Si noti come non sussista un elenco chiuso e definito dei fattori di protezione, vi è piuttosto un elenco aperto e dinamico. L'elenco è dunque destinato ad aggiornarsi fino a ricomprendere nuovi criteri sulla base delle nascenti esigenze sociali e giuridiche. In merito, A. BARBERA, *La Carta dei diritti: per un dialogo tra la Corte italiana e la Corte di Giustizia*, in *Quaderni Costituzionali*, 1, 2018, 149-174; M. MILITELLO, *Principio di uguaglianza e non discriminazione tra Costituzione italiana e carta dei diritti fondamentali dell'Unione Europea*, in *I Working Papers – Centro studi di Diritto del Lavoro Europeo "Massimo D'Antona"*, 77, 2010, 43 ss. Alcune elencazioni, seppure con le limitazioni di cui sopra, sono presenti all'art. 26 del Patto internazionale sui diritti civili e politici, adottato dall'Assemblea Generale delle Nazioni unite il 16 dicembre 1966 il quale recita «[…] la legge deve proibire qualsiasi forma di discriminazione e garantire a tutti gli individui una tutela eguale ed effettiva contro ogni forma di discriminazione, sia essa fondata sulla razza, il colore, il sesso, le lingue, l'opinione politica o qualsiasi altra opinione, l'origine nazionale o sociale, la condizione economica, la nascita o qualsiasi altra condizione» a cui si aggiunge l'art 2, stesso testo, il quale, riferendosi ai medesimi fattori di protezione, impone un obbligo positivo di non discriminazione in capo agli stati firmatari.

[17] A. ARORA ET AL., *The value of standards for health dataset in artificial intelligence-based applications,* in *Nature Medicine*, 29, 2020, 2929-2938.

[18] J. ALVIDREZ ET AL., *The National Institute on Minority Health and Health Disparities Research Framework*, in *American Journal of Public Health*, 1, 2019, 516 ss.

[19] A titolo di esempio, in dottrina italiana si vedano *ex multis* S. ROSSI, *Diritto alla salute tra equità e sostenibilità. Colloquio sulle forme dell'eguaglianza in sanità*, in *BioLaw Journal – Rivista di BioDiritto*, 2, 2019, 7 ss.; R. BIN, G. PITRUZZELLA, *Diritto Costituzionale*, Torino, 2018, 515 ss.; C. PICIOCCHI, *Il principio d'eguaglianza nell'ambito del Biodiritto*, in *BioLaw Journal – Rivista di BioDiritto*, Special Issue 2, 2019, 113 ss.

[20] NATIONAL ACADEMIES OF SCIENCES, ENGINEERING, AND MEDICINE, *Improving Representation in Clinical Trials and Research: Building Research Equity for Women and Underrepresented Groups*, Washington DC, 2022, p. 32; S. NEGRI, *Salute pubblica, sicurezza e diritti umani nel diritto internazionale*, Torino, 2018, 67.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

sarà poi chiamato ad operare, e diversificato, cioè esplicitante informazioni anche - per quanto di nostro interesse nel presente scritto - relative alla razza[21].

Appare necessario chiarire sin da ora come la scienza sia piuttosto netta nel rimarcare l'incapacità degli studi, tanto in campo biologico quanto in campo genetico, di individuare nell'uomo differenze derivanti dall'appartenenza a diverse razze[22]. È tuttavia assodato che la categoria della di razza – seppur scientificamente inconsistente –sia in realtà un concetto socialmente rilevante[23], di conseguenza ancora utilizzato per descrivere la diversità umana. Le odierne disuguaglianze tra i c.d. gruppi razziali sono infatti il prodotto di circostanze (siano esse storiche o contemporanee) di tipo sociale, economico, educativo e politico[24]. Nonostante la scarsa robustezza scientifica del concetto, questo si presenta comunque come un costrutto socio-culturale capace di incidere profondamente sulla vita quotidiana di alcuni gruppi di minoranza, con ampie ricadute sulla condizione psicofisica e sociale degli individui.

La persistenza di questa esigenza di categorizzazione rende necessaria una ulteriore riflessione, incentrata sul concetto di etnia e sull'(eventuale) rapporto con la nozione di razza. Secondo la Corte europea dei diritti dell'uomo l'etnia e la razza sono concetti correlati e sovrapposti[25]. Tuttavia, mentre il concetto di razza è radicato nell'idea che sia possibile, sulla base di caratteristiche morfologiche come il colore della pelle o le caratteristiche somatiche, una classificazione biologica degli esseri umani in sottogruppi, l'etnia ha origine nell'idea di gruppi sociali caratterizzati da medesime origini, bagagli culturali e tradizionali nonché da una affiliazione nazionale, tribale, religiosa e linguistica[26]. Essendo concetti sovrapponibili, la discriminazione fondata sull'appartenenza ad una determinata etnia risulta essere una forma di discriminazione razziale. Identica, dunque, sarà la valutazione di lesività ai fini giuridici[27]. Nel prosieguo, quindi, i due termini saranno impiegati in maniera intercambiabile, facendo sempre riferimento alla dimensione socio-culturale, e non biologica, della categoria[28].

Le riflessioni che si intendono proporre trovano terreno fertile nell'ordinamento statunitense, caratterizzato da dalla presenza di un folto intreccio di minoranze etno-razziali e da una serie di dinamiche socioculturali – attuali e passate – capaci di indirizzare la ricerca medica verso l'esclusione di determinati gruppi sociali[29].

---

[21] A. ARORA ET AL., *op. cit.*, 2933.

[22] Si veda L.L. CAVALLI-SFORZA, P. MENOZZI, A. PIAZZA, *Storia e Geografia dei geni umani*, Milano, 1997.

[23] J.P. CERDENA, M.V. PLAISME, J. TSAI, *From race-based to race-conscious medicine: how anti-racist uprising call us to act*, in *The Lancet*, 396, 2020, 1125.

[24] AMERICAN ANTIHROPOLOGICAL ASSOCIATION EXECUTIVE BOARD, *AAA Statement on Race,* in *American Anthropologist*, 100, 3, 1998, 712-713.

[25] Si vedano *Sedic e Finci c. Bosia Herzegovina*, n. 27996/06 e n. 34836/06, par. 43, *Stoica c. Romania* [Terza sezione], n. 42722/02, par. 117; *D.H. e altri c. Repubblica Ceca*, n. 57325/00 par 176; Nachova and Others v. Bulgaria [GC], nos. 43577/98 e n. 43579/98, par. 145.

[26] Corte EDU, *Timishev c. Russia*, nn. 55762/00 e 55974/00, 12 dic. 2005.

[27] *Ibidem*.

[28] Si veda C. NARDOCCI, *op. cit.*, 71.

[29] Per quanto concerne le dinamiche passate si vedano, a titolo di esempio: H.A. WASHINGTON, *Medical Apartheid: the dark history of medical experimentation on black americans for colonial times to the present*, New York, 2008; L. VILLAROSA, *Under the skin – racism, inequality and the health of a nation*, Melbourne-Londra, 2022. Delle questioni relative all'esclusione di alcune minoranze razziali si occupa, a titolo di esempio, S. GEORGE, N. DURAN, K. NORRIS, *A systematic review of barriers and facilitators to minority research participation among African Americans, Latinos, Asian Americans, and Pacific Islanders,* in *American Journal of Public Health*, 2, 2014, e16-e31.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

2. Trial clinici, minoranze etno-razziali e intelligenza artificiale

Prima di procedere nell'analisi è necessaria una precisazione rispetto a quanto detto. I *dataset* utilizzati per allenare gli algoritmi di intelligenza artificiale applicati in ambito medico sanitario non provengono esclusivamente dai trial clinici, come potrebbe apparire scontato ad una prima lettura. Spesso vi è una commistione tra dati provenienti dalle attività di ricerca e Real World Data (RWD), ossia dati relativi allo status fisico e mentale dell'individuo, prodotti e raccolti nello svolgimento di attività associate al mondo reale come, a titolo esemplificativo, quelli contenuti all'interno della cartella clinica elettronica, quelli inerenti all'assicurazione sanitaria o allo svolgimento di percorsi di cura e quelli raccolti dai c.d. dispositivi intelligenti[30].

È quindi necessario tenere conto del fatto che, anche negli Stati Uniti, il fenomeno della *Data Health Disparity* tra gruppi razziali è anche il risultato di uno sbilanciamento nella raccolta di RWD, principalmente dovuto al divario digitale e alla presenza di un sistema sanitario il cui accesso è spesso condizionato al possesso di una assicurazione sanitaria[31]. I paragrafi che seguono, tuttavia, si concentreranno solo sui *bias* che riguardano la raccolta di dati all'interno dei trials clinici.

A partire dall'ultima metà del secolo scorso gli studi clinici controllati randomizzati (RCT) sono stati considerati dalla comunità scientifica come il gold-standard per la determinazione dell'efficacia e della sicurezza delle nuove terapie e delle nuove pratiche cliniche. Uno dei principali elementi a favore constava nell'assunto della generalizzabilità, intesa come applicabilità dei risultati della ricerca basata su un campione all'intera popolazione. Negli ultimi decenni, tuttavia, un sempre più corposo numero di studi sembra scalfire tale assunto. In particolare, le ricerche dimostrano come, talvolta, individui appartenenti a gruppi etnici differenti tendano a presentare sintomi e decorsi dissimili per la medesima patologia. Se a ciò si somma la consapevolezza che le minoranze razziali presentano un inferiore tasso di coinvolgimento nella ricerca biomedica – e questo anche quando la condizione clinica studiata affligge maggiormente la minoranza stessa[32] – si comprende come spesso i risultati ottenuti dal trial

Secondo lo studio, tra le barriere al reclutamento delle minoranze all'interno dei trial clinici, oltre al timore di subire discriminazioni, si annoverano: i) l'essere in una condizione economico-finanziaria precaria che spesso costringe il soggetto a svolgere più di una attività lavorativa con conseguente assenza di momenti da dedicare alla partecipazione al trial clinico; ii) la presenza di difficoltà linguistiche cui si somma la mancata predisposizione, da parte dei ricercatori, di strumenti comunicativi precipuamente dedicati alle interazioni con le minoranze; iii) la possibilità di subire ripercussioni da parte delle assicurazioni sanitarie a causa della partecipazione al trial clinico in seguito alla scoperta di patologie tramite la partecipazione al trial clinico stesso.

[30] D. ABRAHAMI ET AL., *Use of real-world data to emulate a clinical trial and support regulatory decision making: Assessing the impact of temporality, comparator choice, and method of adjustment*, in *Clinical Pharmacology & Therapeutics,* 2, 2021, 452– 461; D. ADEDINSEWO ET AL., *Health Disparities, Clinical Trials, and Digital Divide*, in *Mayo Clinic* Proceedings, 12, 2023, 1876.

[31] Si veda M. GHASSEMI ET AL., *Practical guidance on artificial intelligence for health-care data*, in *The Lancet Digital Health*, 1, 2019, 157-159; H. IBRAHIM ET AL., *op. cit.,* 260.

[32] Si veda V. VILACANT ET AL., *Inclusion of Under-Represented Racial and Ethnic Groups in Cardiovascular Clinical Trials*, in *Health, Lung and Circulation*, 31, 2022, 1263-1268; M. AWIDI, S.A. HADIDI, *Partecipation of Black Americans in Cancer Clinical Trials: current challenges and proposed solutions*, in *Journal of Oncology Practice*, 17, 2021, 265-272.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

clinico non siano generalizzabili a tutta la popolazione. A risultarne scalfita è l'equa tutela della salute dei consociati[33].

Particolarmente significativo, per comprendere come la razza incida sulle disparità nell'ambito della salute e come questa siano determinate dal modo in cui sono condotti gli studi clinici, è lo studio *Race/ethnicity reporting and representing in US clincal trials: A cohort study*[34]. Qui i ricercatori hanno analizzato i dati aggregati provenienti da ClinicalTrials.gov[35] relativi a 20.692 trials clinici svolti nel territorio statunitense nel periodo 1° marzo 2000 – 9 marzo 2020. Sono stati poi selezionati gli studi che presentavano qualsiasi indicazione circa la razza coinvolta (8817 su 20.692 – circa il 43%). I dati così ottenuti sono stati confrontati con quelli del *2010 US Census databased for US population statistics*. Complessivamente, dal 2007 il 45% dei clinical trials ha riportato un qualsiasi dato sulle razza/etnia coinvolte, mentre solo il 22% ha fornito dati relativi alla partecipazione di tutti i cinque gruppi razziali considerati all'interno dei diversi studi (*White, Black, Hispanic/Latino, Asian, American Indian*). Ai fini della nostra analisi è bene notare come, nel caso di trials riportanti i dati relativi all'arruolamento di tutti i cinque gruppi, la maggior parte dei partecipanti sia riconducibile al *White group*, con un tasso di arruolamento medio del 79,7%, percentuale che supera di molto quella della popolazione bianca risultante dal censimento americano del 2010 (72.4%)[36]. *Hispanic/Latino, Asian, American Indian* risultano invece tutti sottorappresentati rispetto ai dati demografici, mentre l'arruolamento della comunità afroamericana risulta statisticamente allineato alla popolazione censita. Inoltre, il 10% dei clinical trials analizzati ha riportato il 100% di partecipanti bianchi[37].

Se da una parte è vero che l'adeguata rappresentazione all'interno dei clinical trial non riuscirà di per sé ad appianare completamente le disparità nell'accesso alle cure che caratterizzano il panorama statunitense, è altrettanto vero che la rappresentatività dei soggetti coinvolti nella ricerca è strettamente connessa con la generalizzabilità delle scoperte scientifiche[38]. Dalla ricerca generalizzabile discendono

[33] NATIONAL ACADEMIES OF SCIENCES, ENGINEERING, AND MEDICINE, *op. cit.,* 23.

[34] B.E. TURNER ET AL., *Race/ethnicity reporting and representation in US clinical trials: A cohort study*, in *The Lancet Regional Health – Americas*, 11, 2022, 1 ss.

[35] ClinicalTrial.gov è il database del governo statunitense, il quale raccoglie informazioni sui clinical trial, gli studi osservazionali e i loro risultati. Lo scopo del portale è quello di permettere ai ricercatori, ai professionisti della salute e al pubblico di ottenere precise e dettagliate informazioni circa le ricerche cliniche attuali passate. Il portale nasce nel 2000 sulla spinta della legge Food and Drug Administration Modernization Act del 1997 la quale imponeva al U.S. National Institute of Health (NIH) di creare un database capace di raggruppare le informazioni relative ai clinical trial e agli studi osservazionali sui nuovi farmaci. Successivamente, grazie alle modifiche legislative, vi è stato un ampliamento sia degli studi per i quali viene richiesta la menzione all'interno del portale, sia delle informazioni richieste per ciascuno studio. Si vedano in particolare FDA Amendments Act del 2007 (FDAAA); Federal Policy for Portection of Human Subjects (The Common Rule) così come pubblicata nel Federal Register (FR) in 19 gennaio 2017 e modificata dalla revisione 45 CFR 46 del 2018. Per una più approfondita disamina si consulti https://clinicaltrials.gov/about-site/about-ctg (ultima consultazione 01/07/2024).

[36] È opportuno notare come successivamente all'approvazione del FDA Amendments Act del 2007, sia aumentato il coinvolgimento dei soggetti correlati alle diverse minoranze razziali negli studi registrati nel portale ClinicalTrial.gov. Nel decennio 2008 – 2018 infatti il numero di studi riportante un qualsiasi dato generico riferibile al fattore di protezione razziale sono passati dal 26% al 91%. Tuttavia, vi è stato un aumento di solo 30 punti percentili (dall'11% al 41%) degli studi riportanti tutti e 5 i gruppi razziali considerati. B.E. TURNER ET AL., *op. cit.,* 4.

[37] B.E. TURNER ET AL., *op. cit*., 5.

[38] Si veda T. BERMA, C.P. GROSS, J.E. MILLER, *Clinical Trial Diversity – Will we know it when we see it?,* in *JAMA Oncology*, 9, 6, 2023, 765-767.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

migliori protocolli di accesso, diagnosi, cura, trattamento e follow-up, tutti elementi indispensabili per alleviare le disuguaglianze attualmente presenti nel godimento del diritto alla salute in condizioni di uguaglianza[39]. In secondo luogo, poi, la formazione del campione rappresentativo secondo l'approccio di *diversity and inclusion*, insieme soprattutto all'analisi differenziata dei dati e alla trasparenza e leggibilità degli stessi, è fondamentale ai fini della diffusione della c.d. medicina personalizzata[40], che consente di declinare diagnosi e cura in base alle caratteristiche individuali (o di gruppo)[41].

Quanto detto sino a qui risulta ancora più complesso se si considera che i dati estratti dai trials clinici vengono impiegati – seppur assieme ad altre moli di dati provenienti da fonti eterogenee – per l'addestramento dei sistemi di intelligenza artificiale utilizzati in ambito medico sanitario. Infatti, nonostante il *bias* algoritmico sia multifattoriale, le ricerche fin ora condotte dimostrano come il fattore preponderante nel determinarlo consti nel fatto che la maggior parte dei dati di addestramento sono riconducibili a individui appartenenti al *white group*[42], con la conseguenza che un c.d. set di allenamento più rappresentativo condurrebbe certamente a un modello più equo.

Diverse ricerche hanno cercato di approfondire i motivi sottesi alla scarsa rappresentatività razziale interna ai trial clinici[43]. Nonostante ciascun gruppo razziale riferisca – come si è prima analizzato[44] – diversi tipi di impedimenti alla partecipazione, ve ne è uno di trasversale: l'aver subito durante il processo di cura, o il timore di subire durante i trials clinici, atti discriminatori[45].

Queste dinamiche vengono definite *cicli di esclusione*[46]. La ricerca clinica è infatti caratterizzata da cicli di esclusione c.d. primari che si autoalimentano, strutturati essenzialmente in quattro fasi: (i) la cultura anti-minoritaria si manifesta nella percezione, radicata nei ricercatori, per cui i membri delle minoranze sarebbero intrinsecamente diffidenti e dunque meno interessati nella partecipazione alla ricerca; (ii) i ricercatori mettono in atto pratiche di reclutamento discriminatorie che portano alla formazione di campioni demograficamente distorti; (iii) i membri delle comunità minoritaria, percependo il minor

[39] NATIONAL ACADEMIES OF SCIENCES, ENGINEERING, AND MEDICINE, *op. cit.*, 28.

[40] La medicina personalizzata può essere definita come un approccio globale alla prevenzione, alla diagnosi, alla cura e al monitoraggio delle malattie basato sulle caratteristiche genetiche e su altri dati relativi alla salute provenienti da ciascun individuo. Con "dati relativi alla salute" ci si riferisce a informazioni relative alle condizioni di salute (es. allergie, difficoltà visive, uditive), alle cure (es. precedenti interventi chirurgici, farmaci, medicamenti), ai metodi di misurazione tradizionali (es. radiografie, sfigmomanometro, EEC, EEG), allo stile di vita (es. alimentazione, attività fisica, consumo di alcol o sostanze stupefacenti, tabagismo), al contesto di vita (es. qualità dell'acqua ed aria nell'abitazione, condizione lavorativa attuale e pregressa, esposizione a sostanze tossiche). W. DUCH ET AL., *Artificial intelligence approaches for rational drug design and discovery*, in *Current Pharmaceutical Design*, 13, 2007, 1497-1508.

[41] M. TOMASI, *Genetica e Costituzione. Esercizi di eguaglianza, solidarietà e responsabilità,* Napoli, 2019, 198 ss.

[42] M. BARTON, M. HAMZA, B. GUEVEL, *Racial Equity in Healthcare machine Learning: Illustrating Bias in Models with Minimal Bias Mitigation*, in *Cureus*, 15, 2, 2023, 35037.

[43] Si veda *ex multis* M. HUSSAIN-GAMBLES ET AL., *Why ethnic minority groups are under-represented in clinical trials: a review of the literature*, in *Health Soc Care Community,* 5, 2024, 382 ss.; V.I. SHAVERS ET AL., *The state of research on racial/ethnic discrimination in the receipt of health Care*, in *American Journal of Public Health*, 5, 2012, 953 ss; A. SHEIKH, *Why are ethnic minorities under-represented in US research studies?*, in *PLoS Medicine*, 3, 2006, 49 ss.

[44] Si veda *supra* nota 29.

[45] A. DEVLIN ET AL., *The Effect of Discrimination on the Likelihood of Participation in a Clinical Trial*, in *Journal of Racial and Ethnic Health Disparities*, 7, 2020, 1124-1129.

[46] Si veda A. BRACIC, *Breaking the exclusion cycle: how to promote cooperation between majority and minority ethnic groups,* Oxford, 2020.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

interesse da parte delle istituzioni in un loro reclutamento, dimostrano riluttanza nella partecipazione agli studi; (iv) i ricercatori attribuiscono tale ostilità al gruppo minoritario di appartenenza del soggetto, corroborando così i propri pregiudizi iniziali.

Su questo contesto di per sé discriminatorio si innestano le tecnologie intelligenti, le quali vanno a cristallizzare e a esacerbare le dinamiche preesistenti. Si delineano infatti cicli di esclusione c.d. secondari, sovrapposti e interdipendenti a quelli sopra. Le fasi rimangono quattro: (i) il sistema viene addestrato sulla base di dati non rappresentativi e affetti da *bias*; (ii) l'algoritmo emette *output* discriminatori utilizzati poi nel processo di cura; (iii) il destinatario della decisione percepisce l'assenza di *equità* e adotta un atteggiamento respingente; (iv) il ricercatore, percependo l'ostilità, è meno incline ad arruolare soggetti appartenenti alla minoranza razziale all'interno dello studio. Ecco che il sistema di intelligenza artificiale, se allenato con i dati emergenti da trials clinici affetti da scarsa rappresentatività demografica, non potrà che introiettare il razzismo strutturale preesistente[47], emetterà *output* discriminatori e andrà a cristallizzare le disuguaglianze nell'ambito della salute già presenti. Infatti, l'utilizzo di un sistema di intelligenza artificiale affetto da *bias* non potrà che condurre alla erogazione di prestazioni mediche sanitarie che replicano tale retaggio discriminatorio[48].

Al fine di impedire che l'IA diventi veicolo per la cristallizzazione delle disuguaglianze e per evitare l'acuirsi delle condizioni di vulnerabilità che già affliggono le minoranze razziali, appare necessario adottare un approccio antidiscriminatorio preventivo basato sul principio di diversità ed inclusione che necessita, per la sua realizzazione, di un duplice ordine di interventi, rispettivamente in campo sociale e scientifico. Innanzitutto, è necessario adottare, quale approccio ai trials clinici, quello della *race-conscious medicine*[49] dove la razza, che spesso – e come si è visto erroneamente – è considerata come un fattore geneticamente capace di innalzare il rischio di patologie, o il loro decorso, o la reazione ai farmaci, venga piuttosto percepita come veicolo delle disuguaglianze strutturali che caratterizzano il contesto socioculturale, capaci di influire sulla qualità e sulla quantità delle cure[50]. In secondo luogo, appare necessario che i ricercatori integrino il principio di non discriminazione – che è alla base dell'ordinamento[51] – nel modo in cui la ricerca sulla salute è progettata, finanziata, condotta, analizzata e diffusa[52].

3. Gli USA e i tentativi di diversity and inclusion nei trial clinici

Negli ultimi tre decenni l'inclusione delle minoranze razziali nei trials clinici è tornata più volte al centro delle politiche del governo federale statunitense. Appare opportuno indagare le diverse iniziative al

[47] Z.D. Bailey et Al., *How structural racism works – Racist Policies as a Root Couse of U.S. Racial Health Inequities*, in *The New England Journal of Medicine*, 8, 2021, 768-773.

[48] A. Bracic, S.L. Callier, W.N. Prince II, *Exclusion cycles: Reinforcing disparities in medicine*, in *Science,* 377, 2022, 1160.

[49] J.P. Cardena, M.V. Plaisime, J. Tsai, *op. cit.*, 1126.

[50] *Ivi*, 1125.

[51] Si fa riferimento, in relazione al contesto USA e al fattore di protezione razziale, al XV Emendamento della Costituzione degli Stati Uniti e – tra gli altri -al titolo 6 Civil Right Law e alla sezione 1557 dell'Affordable Care Act.

[52] K.H. Chaiyachati et al., *Weaving equity into the fabric of medical research,* in *Journal of General and Internal Medicine*, 8, 2022, 2067.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

fine di comprenderne l'efficacia (o meno) sul piano della *diversity and inclusion*[53]. Queste verranno presentate suddividendo la trattazione in base alla agenzia che le ha emanate.

Nel 1993 il Congresso degli Stati Uniti ha approvato il *NIH Revitalization Act* (PL 103-43) con lo scopo di migliorare il coinvolgimento delle minoranze razziali nei trial clinici. Otto anni più tardi, il National Institute of Health ha provveduto ad aggiornare le linee guida del 1986 intitolate *Inclusion of Women and Minorities as Subject in Clinical research* [54], giungendo ad affermare la necessità, per ciascuno studio, di indicare quali minoranze etniche fossero state ingaggiate nella ricerca[55] con lo scopo di migliorare la rappresentatività del campione coinvolto. Con la modifica del 2001, in linea con le nuove consapevolezze in ambito genetico[56], è stata aggiunta *l'Office of management and Budget Directive's* sulla base della quale si chiarisce come le categorie utilizzate per la classificazione delle razze debbano considerarsi meramente costrutti sociopolitici non scientificamente o antropologicamente fondati. Il nuovo testo fornisce alle agenzie federali occupate nella tutela della salute pubblica linee guida per meglio comprendere le sfumature etniche della società statunitense, permettendo così alle stesse di raccogliere, tabulare e analizzare con maggiore precisione i dati relativi alla eterogeneità demografica (e dunque anche razziale) della partecipazione alla ricerca[57]. L'NIH, al fine di assicurare il rispetto alle linee guida dei trials clinici, ha da un lato proceduto alla creazione dell'*Inclusion Governance Committee*, il quale si occupa di verificare che le ricerche finanziate dall'NIH riportino risultati diversificati per ciascuna minoranza coinvolta; dall'altro richiede ai ricercatori da lui finanziati la compilazione annuale del *Research Performance Progress Report (RPPR)* all'interno del quale è necessario riportare – tra le altre voci – anche i dati di inclusione disaggregati per categorie di ricerca, condizioni socio-economiche e stato di salute[58]. Nonostante le numerose iniziative intraprese, la sottorappresentazione delle minoranze etno-razziali all'interno dei trial clinici che vedono coinvolto l'NIH rimane elevata a causa del rapporto sfavorevole tra costi e benefici risultante dall'introduzione delle modifiche ai protocolli di ricerca previsti dagli strumenti sopra presentati[59].

Il *Revitalization Act* del 1993, pur coinvolgendo direttamente solo il National Institute of Health (NIH), ha indotto altre agenzie federali all'adozione di politiche a favore dell'inclusione. Tra queste vi è il *Centers for Desease Control and Prevention* (CDC). Nel biennio 1995-1996 il CDC ha emesso due policy a favore dell'inclusione delle minoranze etniche nelle attività di ricerca condotte in prima persona e in quelle finanziate ma condotte da soggetti esterni. Queste sono rispettivamente *l'Inclusion of Women and Racial and Ethnic Minorities in Research* e la *Policy on the Inclusion of Women and Racial and Ethnic Minorities in externally Awarded Research* e. Nel 2010 le linee guida hanno subito alcune modifiche, oltre che la fusione all'interno di un unico testo. Il nuovo *Inclusion of Women and Racial and Ethnic*

[53] K. BIBBINS-DOMINGO ET AL., *The imperative for Diversity and Inclusion in Clinical Trials and Health Research Participation*, in *Jama*, 23, 2022, 2283.

[54] https://grants.nih.gov/policy-and-compliance/policy-topics/inclusion/women-and-minorities/guideline (ultima consultazione 03/07/2024).

[55] Per una prospettiva europea si veda M. FASAN, C.M. REALE, *Genere e sperimentazioni cliniche: il Regolamento (UE) n. 536/2014, una occasione mancata?,* in *BioLaw Journal – Rivista di BioDiritto*, 4, 2022, 251 ss.

[56] R. DESALLE ET AL., *Una scomoda scienza – Come la genetica è stata impropriamente usata per definire le razze*, Torino, 2019.

[57] NATIONAL ACADEMIES OF SCIENCES, ENGINEERING, AND MEDICINE, *op. cit.,* 53.

[58] NATIONAL ACADEMIES OF SCIENCES, ENGINEERING, AND MEDICINE, *op. cit.,* 57.

[59] *Ivi*, 52.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Minorities in Research prevede che le minoranze etniche debbano essere adeguatamente rappresentate all'interno delle ricerche finanziate dal CDC, a meno che non vi siano impellenti e gravi ragioni motivanti l'esclusione.

Su un piano più generale, nel 1997 il Congresso americano ha emanato *l'FDMA Section: 115: Clinical Investigations (b) Woman and Minorities Regualtion*, con il quale ha richiesto alla Food and Drug Administration (FDA) di sviluppare linee guida volte all'inclusione e delle donne e delle minoranze nei trial clinici[60]. In risposta l'FDA ha adottato, nel 1998, la *Demographic Rule* in base alla quale è necessario che, all'approvazione di ogni nuovo farmaco, vengano presentati dati sulla sicurezza e sull'efficacia suddivisi per sesso, età e gruppo razziale. Nel 2012 il Congresso ha approvato la *Section 907 dell'FDA Safety and Innovation Act (P.L. 112-144)*. Al fine di soddisfare le richieste del Governo, l'FDA ha prima emesso il report *Collection, Analysis, and Availability of Demographic Subgroup Data for FDA-Approved Medical Products*, dal quale emerge la sottorappresentazione delle minoranze nei trial clinici; poi, nel 2014, ha emanato *l'Action Plan to Enhance the Collection and Availability of Demographic Subgroup Data*, il quale contiene linee guida e raccomandazioni volte a spronare i ricercatori verso l'inclusione delle minoranze all'interno dei trial clinici. Nel 2017 l'FDA pubblica *Evolution and Reporting of Age-, Race-, and Ethnicity- specific data in medical device clinical studies,* una guida con la quale desidera fornire ai ricercatori indicazioni circa le modalità per riportare correttamente i dati demografici all'interno degli studi clinici. È bene notare come la precisa indicazione dei dati demografici dei soggetti coinvolti all'interno del trial clinico non solo appare necessaria ai fini della generalizzazione dei risultati della ricerca, ma costituisce altresì elemento indispensabile per lo sviluppo della medicina personalizzata. Tale approccio, infatti, non può prescindere da una accurata e trasparente indicazione delle caratteristiche – dati sanitari, condizioni socioeconomiche, appartenenza a determinati gruppi etnici – di ciascun soggetto coinvolto. Nel 2020 viene invece pubblicata l'*Enhancing the Diversity to Clinical Trial Populations – Elegibility Criteria, Enrollment Practices, and Trial Designs Guidance for Industry,* con il preciso scopo di riportare l'attenzione sulla necessità di un maggiore/migliore coinvolgimento delle minoranze. Da ultimo è stato pubblicato nel mese di giugno 2024 il report *Enhancing Clinical Study Diversity,* il quale risponde agli imperativi sanciti della *Section 3603 of the Food and Drug Omnibus Reform Act of 2022 (FDORA)*[61] e che presenta nuovamente raccomandazioni, buone pratiche nonché una disamina delle iniziative in atto volte alla maggiore inclusione delle minoranze razziali nei clinical trial.

Nonostante le iniziative e i diversi atti adottati dall'FDA e volti ad aumentare la diversità demografica all'interno dei trials clinici siano molteplici, alcuni gruppi minoritari rimangono sottorappresentati all'interno dei progetti di ricerca. Questo accade per un duplice ordine di motivi. Innanzitutto, vi è l'assenza in capo dall'FDA del potere di far rispettare coercitivamente le linee guida (questo limite viene spesso indicato anche in apertura alle linee guida, dove viene chiarito come tali strumenti rappresentino solamente il pensiero attuale della FDA sull'argomento e come queste non stabiliscano alcun diritto, né alcuna responsabilità giuridicamente vincolante a meno che non sia contrariamente

[60] Food and Drug Administation Modernization Act of 1997, S. 830, 105th Congress, November 21, 1997.

[61] https://www.fda.gov/regulatory-information/selected-amendments-fdc-act/food-and-drug-omnibus-reform-act-fdora-2022 (ultima consultazione 03/07/2024).

Downloaded from www.biodiritto.org.
ISSN 2284-4503

esplicitato)[62]; cui si aggiunge la difficoltà dell'Agenzia di rispondere in tempi brevi alle iniziative Governative, con la conseguenza che, spesso, quando vengono adottati, questi strumenti di soft-law sono oramai sorpassati[63].

4. Conclusioni

I benefici in ambito medico sanitario che l'introduzione dei sistemi di intelligenza artificiale promette di apportare sono innumerevoli. Tali contributi benefici, tuttavia, non giungono scevri da rischi. A causa dei meccanismi intrinsecamente connessi al loro funzionamento, come la dipendenza da un *set di dati* di addestramento e il flusso *garbage in – garbage out*, di fatto, i sistemi di intelligenza artificiale sono capaci di perpetuare e cristallizzare le discriminazioni razziali preesistenti nel godimento del diritto alla salute. Le discriminazioni razziali in ambito medico-sanitario, infatti, non appaiono come un fenomeno nuovo strettamente (ed esclusivamente) correlato con l'implementazione dei sistemi di AI, quanto piuttosto come una dinamica preesistente alle tecnologie intelligenti, che trova tuttavia in queste ultime un mezzo di diffusione e consolidamento. Al fine di invertire tale tendenza appare necessario adottare un approccio di diversità ed inclusione capace di intervenire a monte dei c.d. momenti del *bias*, ovverosia – tra gli altri – nel momento della creazione del *dataset*.

Si è messa dunque in luce l'importanza di estendere il più possibile il pool partecipanti ai trials clinici e di includere, soprattutto, persone facenti parte delle minoranze etno-razziali, così da avere un campione più rappresentativo possibile. Tuttavia, questo approccio è necessario, ma insufficiente al conseguimento di un ideale di eguaglianza nell'ambito della salute. Per raggiungere tale obiettivo appare altrettanto indispensabile una analisi differenziata dei risultati ottenuti. Ciò diviene possibile solo se i risultati dei trials clinici sono trasparenti e dunque leggibili anche attraverso il filtro della razza. Infatti, nonostante la categoria della razza non presenti basi biologiche o genetiche, appare come un costrutto socioculturale fortemente in grado di incidere sullo stato di salute dell'individuo. Quanto sopra appare ancora più impellente se si considera che i dati ottenuti dai trials clinici vengono utilizzati anche per lo sviluppo di sistemi intelligenti di medicina personalizzata[64]. Per fare questo è necessario un cambio di paradigma: bisogna abbandonare la concezione di una *color blind medicine*, la quale considerava l'utilizzo delle categorie razziali come un veicolo per la discriminazione, per abbracciare il metodo della *race conscious medicine*, inteso come un approccio alternativo che pone l'accento sul razzismo piuttosto che sulla razza, come determinante chiave della malattia[65]. Una medicina cosciente richiede anche una ricerca cosciente, nei processi della quale sia garantito un adeguato coinvolgimento di tutte le minoranze etno-razziali, soprattutto là dove la finalità sia quella di produrre sistemi tecnologici di IA,

[62] FOOD AND DRUG ADMINISTRATION, *Enhancing the Diversity of Clinical Trial Populations – eligibility criteria, Enrollment Practices, and Trial Designs – Guidance for Industry*, 2020 https://www.fda.gov/regulatory-information/search-fda-guidance-documents/enhancing-diversity-clinical-trial-populations-eligibility-criteria-enrollment-practices-and-trial (ultima consultazione 26/11/2024).

[63] NATIONAL ACADEMIES OF SCIENCES, ENGINEERING, AND MEDICINE, *op. cit.,* 63.

[64] N.J. SCHORK, *Artificial Intelligence and Personalized Medicine,* in *Cancer Treat Research*, 178, 2019, 265-283.

[65] M.V. PLAISME, J. TSAI, *op. cit.,* 1125.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

che di dati si nutrono e vivono.[66]. L'assenza di dati riferibili anche alle minoranze razziali, la c.d. *data poverty*, non può che porsi come un ostacolo al raggiungimento dell'eguaglianza nell'ambito della salute.

Come si è potuto notare, negli Stati Uniti d'America, nonostante i ripetuti tentativi da parte delle varie agenzie governative, l'inclusione delle minoranze all'interno delle attività di ricerca appare ancora come un obbiettivo mancato. Le linee guida e le raccomandazioni vengono infatti percepite dalla comunità scientifica come suggerimenti, piuttosto che come imprescindibili direttrici di uno sviluppo scientifico capace di lenire le disuguaglianze e realizzare il principio di equità che sta alla base dell'ordinamento.

Potrebbe apparire opportuno per il "nuovo continente" volgere lo sguardo al "vecchio", dove il Regolamento europeo per l'intelligenza artificiale (Reg. UE 2024/1689), consapevole della grande influenza che il fenomeno della *data poverty* ha sulla propagazione delle diseguaglianze, all'art 10, impone, per lo sviluppo e il training di tutti i sistemi ad alto rischio (tra cui rientrano anche i sistemi di IA utilizzati in ambito medico-sanitario) l'utilizzo di *dataset* che presentano le seguenti caratteristiche: rilevanza; rappresentatività, appropriatezza statistica anche – dove applicabile – con riguardo alle persone e ai gruppi di persone sui quali il sistema di AI dovrà performare; nonché la capacità di tenere in considerazione caratteristiche geografiche, sociali e comportamentali del contesto di operatività.

Tutte le ricerche nell'ambito della salute dovrebbero farsi portatrici dei principi di equità e non discriminazione lungo l'intero percorso di ricerca, dalla progettazione alla divulgazione dei risultati.

Vi è la necessità, in conclusione, di sviluppare una maggiore consapevolezza circa l'intreccio tra attività di raccolta dei dati – tra cui rientrano anche i trial clinici – e sviluppo di sistemi di IA. Questa relazione, se non adeguatamente strutturata (anche nella consapevolezza che il raggiungimento della uguaglianza nell'ambito della salute non può prescindere dalla considerazione di un fattore di protezione quale la razza) si rivela capace di propagare le già esistenti dinamiche discriminatorie razziali.

[66] M. TOMASI, *The legacy and the future of the race between science, constitutional lexicon, and political action,* in *BioLaw Journal – Rivista di BioDiritto*, Special Issue 1, 2021, 76.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Vulnerability in the age of artificial intelligence: addressing gender bias in healthcare

Laura Piva*

ABSTRACT: Gender bias represent a source of vulnerability in current clinical practice as they can harm patients, especially those identified as gender minorities. This risk is increased in the age of AI-powered healthcare, which calls for redefining vulnerability and strategies to prevent algorithms from exacerbating already existing inequalities. To this end, we analyse the solutions proposed by the European Union's recent AI Act. Moreover, we consider whether AI itself can be a solution to these issues.

KEYWORDS: Vulnerability; artificial intelligence; gender bias; healthcare; technosolutionism.

SUMMARY: 1. Introduction – 2. Vulnerable patients – 2.1. The role of sex and gender in medicine – 2.2. Gender bias as a source of vulnerability – 3. Vulnerable algorithms – 4. Looking for a cure – 4.1. Can AI be a solution? – 4.2. The AI Act's approach – 5. Conclusions.

1. Introduction

Vulnerability is a multifaceted concept which is of central importance to numerous areas, including medicine. Identifying vulnerable groups within the healthcare context represents the first step towards guaranteeing that everybody enjoys high-quality care and fair access to medical services. Ultimately, this means safeguarding and promoting the fundamental rights to health and equality, which are strongly linked[1].

In this article, we argue that the disruptive advent of artificial intelligence (AI) calls for reshaping the concept of vulnerability and rethinking the ways to deal with it.

On one hand, the digitalisation of our world exposes individuals and groups to new or increased physical and psychological risks and forms of inequality[2]. With AI, this happens partly due to the algorithmic discrimination phenomenon but also due to the uneven distribution of these technologies and the "digital divide".

* *PhD Student, University of Trento. Mail: laura.piva-1@unitn.it.* *The article was subject to a double-blind peer review process.*

[1] L. BUSATTA, *La salute sostenibile. La complessa determinazione del diritto ad accedere alle prestazioni sanitarie*, Torino, 2018, 1-10; M. TOMASI, *Sperimentazioni cliniche e medicina di genere: la ricerca dell'euguaglianza attraverso la valorizzazione delle differenze*, in B. PEZZINI, A. LORENZETTI (eds.), *70 Anni dopo tra uguaglianza e differenza: una riflessione sull'impatto del genere nella Costituzione e nel costituzionalismo*, Torino, 2019, 215-230.

[2] S.S. YILMAZ, Z.Ö. KOTIL, *Final exit before the bridge in AI: Strengthening the right to human oversight of vulnerable subjects with two agents,* in *European Journal of Privacy Law & Technology - Observatory,* 2023, available at: https://universitypress.unisob.na.it/ojs/index.php/ejplt/article/view/1857 (last accessed: 21/11/2024).

Downloaded from www.biodiritto.org.
ISSN 2284-4503

On the other hand, there is the hope that AI will shed light on human bias and mitigate or even overcome them.

The present article assesses these assumptions by taking gender bias in healthcare as a case study. After investigating the origin and role of gender bias in healthcare, we tried to understand how they can be embedded in algorithms and how AI can reproduce or expand them, with a negative impact on individuals' right to health and equality.

This operation implied delving into human and algorithmic bias, for which consulting interdisciplinary literature relating to medicine, bioethics and technology was mandatory.

After having defined this problem, we will review some possible solutions. First, we will draw some reflections on whether AI itself could be used to overcome gender bias in healthcare. Then we will focus on the way the European Union's law deals with this issue with the recently approved Regulation on Artificial Intelligence (AI Act). In the conclusions, we consider how AI can play a role in raising awareness towards existing (gender) bias, thus encouraging developers, policymakers and regulators to seriously address healthcare disparities abandoning the myth of AI solutionism.

2. Vulnerable patients

Although its meaning is not univocal, vulnerability can generally be understood as the propensity to be easily physically or mentally hurt, influenced or attacked[3].

If we adopt an ontological perspective, this is a condition universally experienced by human beings[4] due to their «shared biological fragility»[5] and their relational nature, which makes them dependent upon others and, thus, in need of care[6].

Besides being universal, vulnerability can be understood as context-specific as well, meaning that individuals or groups with certain characteristics are exposed to an increased risk of harm due to social, political, economic, or environmental factors[7].

Both these perspectives are reflected in the healthcare context. On one hand, experiences such as illness, pain and hospitalisation make all patients vulnerable. Moreover, patients depend upon

[3] "Vulnerability", Cambridge Dictionary: https://dictionary.cambridge.org/dictionary/english/vulnerability (last accessed: 01/07/2024).

[4] M.A FINEMAN, *The vulnerable subject: Anchoring equality in the human condition,* in *Yale Journal of Law and Feminism,* 20, 1, 2008, 1-23.

[5] This expression is taken from W. ROGERS, *Vulnerability in Bioethics*, in C. MACKENZIE, W. ROGERS, S. DODDS (eds.), *Vulnerability: New Essays in Ethics and Feminist Philosophy*, Oxford, 2013, 71.

[6] C. BOTTI, *Vulnerabilità, relazioni e cura. Ripensare la bioetica*, in *Etica & Politica / Ethics & Politics*, 18, 3, 2016, 33-57. This is in line with the Oxford Reference's definition of "vulnerability", namely «the position of relative disadvantage, which requires a person to trust and depend upon others», see: https://www.oxfordreference.com/display/10.1093/oi/authority.20110803120303277 (last accessed: 01/07/2024).

[7] B. WISNER, P. BLAIKIE, T. CANNON, I. DAVIS, *At Risk. Natural Hazards, People's Vulnerability and Disasters*, second ed., London, 2004 describe vulnerability as the set of characteristics of an individual or group that, combined with the situation or context in which they live, influence their ability to adapt, resist and anticipate the impact of adverse events. Similarly, ten Have refers to a limited capacity or resilience to absorb, adapt to or recover from that harm: H. TEN HAVE, *Vulnerability: Challenging bioethics*, London, 2016. See also: C. MACKENZIE, W. ROGERS, S. DODDS (eds.), *Vulnerability: New Essays in Ethics and Feminist Philosophy*, Oxford, 2013.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

physicians due to the information asymmetry existing between them[8], even though the shift from medical paternalism to more participatory models of care has partly reduced this gap[9].
At the same time, vulnerability is indisputably shaped by power dynamics[10] and, therefore, some categories of patients historically disadvantaged have been exposed to higher risks or more detrimental consequences than others based on their ethnicity, religion, age, income, sex and gender.
While acknowledging that all these variables – and their intersection[11] – are important to identify vulnerable subjects in healthcare and address the disparities they face and their specific needs, this paper focuses on sex and gender only.
First, we will try to understand which gender biases exist in healthcare and why they can be a source of vulnerability (i.e., of increased risk of harm or discrimination) in this field.

2.1. The role of sex and gender in medicine

Medical literature has proved sex and gender[12] to be relevant in medicine, as they can influence the likelihood of developing a certain condition, its risk factors and development, symptoms, therapeutic

[8] L.A. COYLE, S. ATKINSON, *Vulnerability as practice in diagnosing multiple conditions*, in *Medical Humanities*, 45, 3, 2019, 278-286.
[9] This process, which has led to greater recognition for patients' autonomy, has however not affected everybody equally, as noted by L. BUSATTA, C. CASONATO, S. PENASA, M. TOMASI, *Le "maschere" della vulnerabilità nella cura della persona*, in AA. VV. (eds.), *Liber Amicorum per Paolo Zatti*, vol. 1, Napoli, 2023, pp. 651-663 in relation to minors, elderly, disabled and prisoned patients. See also W. ROGERS, *op. cit*.: «the greater the knowledge and skills imbalance between the practitioner and patient, the more vulnerable the patient is to harm from their medical attendant, and the more important it is that practitioners are bound to protect patient interests». For an overview of different models of care: E.J. EMANUEL, L.L. EMANUEL, *Four Models of the Physician-Patient Relationship,* in *JAMA*, 267, 16, 1992, 2221-2226.
[10] K. TIERNEY, *Disasters: A Sociological Approach*, Cambridge, 2019, 127: «people are not born vulnerable, they are made vulnerable […] different axes of inequality combine and interact to form systems of oppression – systems that relate directly to differential levels of social vulnerability».
[11] C.H.A. KURAN ET AL., *Vulnerability and vulnerable groups from an intersectionality perspective,* in *International Journal of Disaster Risk Reduction*, 50, 101826, 2020, 1-8, https://doi.org/10.1016/j.ijdrr.2020.101826; F. LUNA, *Identifying and evaluating layers of vulnerability – a way forward,* in *Developing World Bioethics*, 19, 2, 2019, 86-95.
[12] Defining "sex" and "gender", and how such terms will be used in the present paper is of paramount importance. "Sex" refers to the individual's biological characteristics (sexual chromosomes XX and YX, genitalia, sexual hormones) at birth, whereas "gender" indicates the behaviours, attitudes, feelings and role perception that a person has of herself (gender identity) or that a culture attributes to an individual (D. CIRILLO ET AL., *Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare*, in *npj Digital Medicine*, 3, 81, 1 June 2020, 1-11, https://doi.org/10.1038/s41746-020-0288-5). Thus, sex is a biological construct while gender a psychological and social one. As for what concerns this article, the examples that will be reported – which were derived from relevant medical literature – mainly address differences between males/men and females/women, however, this does not imply that sex or gender are binary constructs (see: T.E. MADSEN ET AL., *Article Commentary: Sex- and Gender-Based Medicine: The Need for Precise Terminology*, in *Gender and the Genome*, 1, 3, 2017, 122-128; S.S. RICHARDSON, *Sex Contextualism*, in *Philosophy, Theory, and Practice in Biology*, 14, 2, 2022, 1-17, https://doi.org/10.3998/ptpbio.2096). "Gender bias" will be used in relation to prejudices and assumptions concerning both sex and gender.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

needs, drug effectiveness and mortality[13]. Moreover, gender is a social determinant of health which affects aspects like lifestyle, access to healthcare and help-seeking behaviours[14]. Being aware of such differences is pivotal to realise accurate prevention, diagnosis and treatment, as well as healthcare policies directed at achieving equity.

Notwithstanding, medicine has long been an androcentric discipline, ignoring the specificities and needs of all of those deviating from the standard cisgender male model[15].

Even if in the last decades such issues have started being addressed by gender-specific medicine[16], problems persist as clinical trials' samples are still under-representative of women and sex and gender minorities and research towards their specific health conditions and needs gains scarce funding and limited attention[17].

In any case, disparities based on gender can not only be explained by the lack of representative medical data. For instance, healthcare practitioners' attitudes and responses can change depending on whether they perceive the patient as a man or a woman, undermining fair healthcare access and delivery[18].

2.2. Gender bias as a source of vulnerability

What has been described so far highlights two opposite and equally damaging attitudes in medicine: not accounting for sex and gender when such variables have an actual impact on the illness experience and, on the contrary, hyper-focusing on them when they do not[19].

---

[13] S. GARATTINI, R. BANZI, *Una medicina che penalizza le donne*, Cinisello Balsamo, 2022; M.J. LEGATO, P.A. JOHNSON, J.E. MANSON, *Consideration of sex differences in medicine to improve health care and patient outcomes*, in *JAMA*, 316, 18, 2016, 1865-1886; S. GREGO ET AL., *"Sex and gender medicine": il principio della medicina di genere*, in *Giornale Italiano di Cardiologia*, 21, 8, 2020, 602-606.

[14] The WHO individuated gender as one of the core social determinants of health (together with income, education, occupation, social class and ethnicity) in O. SOLAR, A. IRWIN, *A conceptual framework for action on the social determinants of health. Social determinants of health discussion paper 2 (policy and practice),* Geneva, 2010. On the same topic: F. MAUVAIS-JARVIS ET AL., *Sex and gender: modifiers of health, disease, and medicine*, in *Lancet*, 396, 10250, 2020, 565-582; N. BUSLÓN, A. CORTÉS, S. RACIONERO-PLAZA, *Sex and gender inequality in precision medicine: Socioeconomic determinants of health*, in D. CIRILLO, S. CATUARA SOLARZ, E. GUNEY (eds.), *Sex and Gender Bias in Technology and Artificial Intelligence: Biomedicine and Healthcare Applications*, Amsterdam, 2022, 35-54.

[15] For instance, medical research and clinical trials have historically been exclusively or predominantly conducted on men, and the results thereby obtained have been generalised to the whole population. Ironically, this was partly due to the willingness to protect "vulnerable women" from the risks of human experimentation but ended up drawing incorrect and often harmful conclusions towards them and other patients' groups. See: M. FASAN, C.M. REALE, *Genere e sperimentazioni cliniche: il Regolamento (UE) n. 536/2014, un'occasione mancata?* in *BioLaw Journal – Rivista di BioDiritto*, 4, 2022, 251-276; T.E. MADSEN ET AL., *op. cit.*, 122-128.

[16] G. BAGGIO ET AL., *Gender medicine: a task for the third millennium*, in *Clinical Chemistry and Laboratory Medicine*, 51, 4, 2013, 713-727; S. GREGO ET AL., *op. cit.*, 602-606.

[17] M. TOMASI, *op. cit.*, 215-230.

[18] For example, K. HAMBERG, *Gender bias in medicine*, in *Women's Health*, 4, 3, 2008, 237-243 highlighted that physicians tend to interpret men's symptoms as organic and women's as psychosocial, leading to inaccurate or delayed diagnosis for the latter.

[19] M. SUNDAI ET AL., *Law, policy, biology, and sex: Critical issues for researchers*, in *Science*, 376, 6595, 2022, 802-804; I. STRAW, *The automation of bias in medical Artificial Intelligence (AI): Decoding the past to create a better*

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Hence, we can identify two types of gender bias negatively affecting healthcare: those leading to "gender blindness" (i.e., ignoring true biological and social differences)[20] and those creating false or discriminatory assumptions.

When medical decision-making is tainted by such prejudices, there can be serious consequences such as missed or inaccurate diagnoses, ineffective or less effective treatments, adverse effects, and wrong prioritisation during triage or emergency admissions (e.g., due to the downplay of symptoms or underestimation of predictive factors). All of this can put individuals' right to health in danger. This is particularly critical when those most at risk are the same patients who already experience poorer health and/or face barriers to accessing healthcare services due to historical or pre-existing inequities. Thus, gender bias can be qualified as a source of vulnerability for patients.

These issues reach paramount and renovated relevance in the age of AI-powered healthcare, as algorithms are capable of replicating and amplifying such bias and their discriminatory effects.

3. Vulnerable algorithms

Although AI carries the promise of yielding medical decision-making more accurately and fairly by correcting human biases[21] such technology is not neutral[22].

Conversely, AI can incorporate values and prejudices and, thus, be vulnerable to bias «that may disproportionately affect model performance in a certain subgroup»[23].

This means that gender bias illustrated in the previous section can be embedded in algorithms, according to the well-known "garbage in, garbage out" principle[24]. For instance, machine learning (ML) algorithms trained on datasets under-representative of the female population showed sex-related performance disparities in predicting heart failure[25] or the likelihood of developing liver diseases[26]. In particular, the systems produced a significantly higher rate of false negative results when applied to women,

future, in *Artificial Intelligence in Medicine*, 110, 101965, 2020, 1-3, https://doi.org/10.1016/j.artmed.2020.101965.

[20] D. CIRILLO ET AL., *op. cit.*, 1-11.

[21] C.R. SUNSTEIN, *Algorithms, correcting bias*, in *Social Research: An International Quarterly*, 86, 2, 2019, 499-511.

[22] P. TRAVERSO, *Breve introduzione tecnica all'Intelligenza Artificiale*, in *DPCE online*, 1, 2022, 155-167; E. STRADELLA, *Stereotipi e discriminazioni: dall'intelligenza umana all'intelligenza artificiale*, in AA. VV. (eds.) *Liber Amicorum per Pasquale Costanzo Costanzo – Diritto Costituzionale in trasformazione Vol. I – Costituzionalismo, Reti e Intelligenza artificiale*, Genova, 2020, 391-400.

[23] J.K. PAULUS, D.K. KENT, *Predictably unequal: understanding and addressing concerns that algorithmic clinical prediction may increase health disparities*, in *npj Digital Medicine*, 3, 99, 2020, 4, https://doi.org/10.1038/s41746-020-0304-9. See also: A.J. LARRAZABAL ET AL., *Gender imbalance in medical imaging datasets produces biased classifiers for computer aided diagnosis*, in *Proceedings of the National Academy of Sciences of the United States of America*, 117, 23, 2020, 12592-12594.

[24] R. XENIDIS, L. SENDEN, *EU non-discrimination law in the era of artificial intelligence: Mapping the challenges of algorithmic discrimination*, in U. BERNITZ ET AL. (eds.), *General Principles of EU law and the EU Digital Order*, Alphen aan den Rijn, 2020, 151-182.

[25] I. STRAW, G. REES, P. NACHEV, *Sex-based Performance Disparities in Machine Learning Algorithms for Cardiac Disease Prediction: Exploratory Study*, in *Journal of Medical Internet Research*, 26, 26 August 2024, 1-18, https://doi.org/10.2196/46936.

[26] I. STRAW, H. WU, *Investigating for bias in healthcare algorithms: A sex-stratified analysis of supervised machine learning models in liver disease prediction*, in *BMJ Health and Care Informatics*, 29, 1, 2022, 1-8.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

resulting in missed diagnoses and a consequent lack of appropriate and timely care for this patient group.

Even when the data sample is diverse, major errors can arise from how such data were interpreted, selected, cleaned, formatted, and labelled before building the training dataset[27].

Besides data, a second entry point for bias is the algorithm's design which encompasses the definition of the system's objectives and target population, the selection of the model's relevant features and their weight, and the choice of training, testing and validation methodologies[28].

We can conclude that algorithms' "technical" vulnerabilities can make patients vulnerable by replicating existing biases which undermine their right to health and equality. In fact, model performance disparities not only put single individuals' safety at stake but can result in discrimination when patients are negatively affected due to belonging to a certain group[29].

At the same time, I argue that AI generates further forms of vulnerability, both from a quantitative and qualitative point of view.

On one hand, ML, DL and generative AI can not only reproduce biases but also multiply them due to "feedback loops"[30]. On the other, algorithms' spurious correlations could generate inaccurate predictions based on different (and even unexpected) characteristics or combinations of them (e.g., gender, ethnicity, income), leading to intersectional discrimination and reinforcing historical inequities[31].

This shows how algorithms can exponentially enhance inequality and normalise it, as the optimism surrounding AI and the opacity that usually characterises it make it very likely that such phenomena will go undetected[32].

Thus, medical AI calls for looking for ways to target algorithmic discrimination as a source of vulnerability.

[27] For instance, medical data (e.g., parameters and biochemical thresholds) could have been aggregated so that sex and gender indicators do not emerge. Likewise, they could only account for binary definitions of such categories. See: A. GERYBAITE, S. PALMIERI, F. VIGNA, *Equality in Healthcare AI: Did Anyone Mention Data Quality?*, in *BioLaw Journal – Rivista di BioDiritto*, 4, 2022, 385-409.

[28] For instance, an AI model could be built around and item, such as an x-ray scan, which is more informative for the male population than for the female one in detecting a certain medical condition: M. GANZ, S.H. HOLM, A. FERAGEN, *Assessing Bias in Medical AI*, in *Workshop on Interpretable ML in Healthcare at International Connference on Machine Learning (ICML)*, 2021, available at: https://www.cse.cuhk.edu.hk/~qdou/public/IMLH2021_files/64_CameraReady_ICML_2021_Interpretable_Machine_Learning_in_Healthcare_workshop.pdf (last accessed: 21/11/2024).

[29] P.L. LAU, *AI Gender Biases in Women's Healthcare: Perspectives from the United Kingdom and the European Legal Space*, in E. GILL-PEDRO, A. MOBERG (eds.), *YSEC Yearbook of Socio-Economic Constitutions 2023: Law and the Governance of Artificial Intelligence*, Cham, 2023, 247-274; E. STRADELLA, *op. cit.*, 391-400.

[30] "Feedback loops" occur when previous AI's outputs influence the future ones.

[31] L. GOETZ, N. SEEDAT, R. VANDERSLUIS, M. VAN DER SCHAAR, *Generalization – a key challenge for responsible AI in patient-facing clinical applications*, in *npj Ditigal Medicine*, 7, 126, 21 May 2024, 1-4, https://doi.org/10.1038/s41746-024-01127-3.

[32] R. WALKER, J. DILLARD-WRIGHT, *Algorithmic bias in artificial intelligence is a problem – And the root issue is power*, in *Nursing Outlook*, 71, 102023, 2023, 1-4.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

4. Looking for a cure

4.1. Can AI be a solution?

First, we need to consider whether AI itself could be the solution to gender bias and algorithmic discrimination[33]. In fact, one could argue that such issues might be solved by building "better algorithms". Although tempting, the idea of using technology to eliminate bias is utopistic and reductive as it fails to conceive them as a "symptom of power imbalances"[34]. More generally, "technological solutionism"[35] has been criticised as it aims to solve complex social issues – such as healthcare inequities – with technological means only, while they require political and legal action.

This, however, does not imply that algorithms can't play a positive role in overcoming gender bias in medicine. The major aid that AI can give in this respect is to make them evident.

For instance, some authors propose to use algorithms such as linear regression or decision trees to achieve post hoc explanations of opaque AI systems. Turning a black box into a white one or using explainable AI could help identify gender bias, which constitutes the first step to establishing whether they are desirable (i.e., they reflect true biological or social differences) or not[36].

After having established so, a possible strategy might be selective deployment of medical AI tools, meaning that they will be used in relation to the population for which they are able to derive accurate conclusions only[37]. This appears in line with the idea that using AI might not always be desirable, especially when efficiency might be counteracted by discriminatory or dangerous outcomes.

Another possibility might be adopting measures to mitigate bias. Once again, some researchers have proposed to exploit AI to this end, elaborating algorithms suitable not only to detect but even to correct certain biases (e.g., balancing uneven datasets ex-post). However, when problems are not related to datasets' lack of diversity corrections might be difficult to achieve.

Above all, there must be an acknowledgement that technical solutions cannot answer the underlying societal problems that have led to the creation of gender stereotypes or to disregard the importance of conducting medical research adopting a gender perspective.

Moreover, as discussed above, gender bias in healthcare not only originate from social problems but also jeopardise human rights such as the right to health and equality. Therefore, we must now examine how the law could tackle gender bias and vulnerability, focusing on the EU legal landscape.

[33] C.R. SUNSTEIN, *op. cit.*, 499-511.

[34] R. WALKER, J. DILLARD-WRIGHT, *op. cit.*, 1-4.

[35] According to E. MOROZOV, *To Save Everything, Click Here: Technology, Solutionism and the Urge to Fix Problems that Don't Exist*, New York, 2013, 5-16, the term refers to the erroneous belief that every political, social, organizational, administrative, and policy problem can be addressed with technological solutions.

[36] R. CONFALONIERI ET AL., *A unified framework for managing sex and gender bias in AI models for healthcare*, in D. CIRILLO, S. CATUARA SOLARZ, E. GUNEY (eds.), *Sex and Gender Bias in Technology and Artificial Intelligence: Biomedicine and Healthcare Applications,* Amsterdam, 2022, 179-204; P. CHANDAK, N.P. TATONETTI, *Using Machine Learning to Identify Adverse Drug Effects Posing Increased Risk to Women*, in *Patterns – Cell Press*, 1, 7, 2020, 1-15.

[37] L. GOETZ ET AL., *op. cit.*, 1-4.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

4.2. The AI Act approach

The European Union has decided to tackle the risks that AI poses to health, safety and fundamental rights with the horizontal Regulation 2024/1689, known as the AI Act[38].

The way the AI Act strives to protect such rights is *ex-ante*, namely by requiring developers and deployers to incorporate certain requirements into their AI systems[39]. As the Regulation follows a risk-based approach, most of the measures indicated therein are mandatory for high-risk systems only. This category includes some systems to be deployed in the healthcare sector, such as medical and in vitro diagnostic medical devices[40], systems «intended to be used by public authorities or on behalf of public authorities to evaluate the eligibility of natural persons for essential public assistance benefits and services, including healthcare services» and «emergency healthcare patient triage systems»[41].

It is among these requirements, in particular the one related to data and data governance (Art. 10) that we find a possible solution to algorithmic discrimination, as well as a reference to vulnerable subjects.

The main idea expressed by Article 10 is that to produce accurate and non-discriminatory results, AI systems must be trained, validated and tested with high-quality datasets. This means, *inter alia*, that data must have been collected and processed correctly, that they must be relevant to the context and, to the best extent, error-free and complete. Also, it is specified that data shall be «sufficiently representative», and have «the appropriate statistical properties, including, where applicable, as regards the persons or groups of persons in relation to whom the high-risk AI system is intended to be used»[42].

The AI Act recalls the need to balance these objectives with the right to data protection. However, it establishes a new exception for processing special categories of data according to the GDPR – which

[38] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32024R1689&qid=1729690544660#d1e2090-1-1 (last accessed: 21/11/2024). The purpose of the AI Act, which is the world-first horizontal regulation on artificial intelligence, is to «improve the functioning of the internal market and promote the uptake of human-centric and trustworthy artificial intelligence, while ensuring a high level of protection of health, safety, fundamental rights enshrined in the Charter, including democracy, the rule of law and environmental protection, against the harmful effects of AI systems in the Union and supporting innovation» (Art. 1, Reg. EU 2024/1689).

[39] Art. 3(1) defines AI system as «machine-based system designed to operate with varying levels of autonomy, that may exhibit adaptiveness after deployment and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments».

[40] As long as they require a third-party conformity assessment, see: Art. 6(1) and Annex I, Reg. EU 2024/1689.

[41] Art. 6(2) and Annex III, 5(a) and (d), Reg. EU 2024/1689. However, some other AI systems used in the medical context might be classified as "limited" or "minimal risk. In any case, those systems will still need to adhere to specific transparency requirements when interacting with natural persons (Art. 50, Reg. EU 2024/1689). Moreover, chatbots providing medical advice might have to implement the requirements indicated for generative AI. For some considerations related to medical AI and (draft) AI Act's provisions: S. PALMIERI, T. GOFFIN, *A Blanket That Leaves the Feet Cold: Exploring the AI Act Safety Framework for Medical AI*, in *European Journal of Health Law*, 30, 4, 2023, 406-427; H. VAN KOLFSCHOOTEN, *EU regulation of Artificial Intelligence: challenges for patients' rights,* in *Common Market Law Review,* 59, 1, 2022, 81-112.

[42] Art. 10(3), Reg. EU 2024/1689.

Downloaded from www.biodiritto.org.

ISSN 2284-4503

include data regarding sex and gender – «to the extent that it is strictly necessary for the purpose of ensuring bias detection and correction»[43]. In fact, Article 10 also requires putting in place appropriate measures to detect, prevent and mitigate possible biases, acknowledging that AI can «perpetuate and amplify existing discrimination, in particular for persons belonging to certain vulnerable groups»[44].

A second reference to vulnerability is contained in Recital 165, which states that providers and deployers of AI systems of all risk classes should be «encouraged to apply on a voluntary basis additional requirements related, for example, to the elements of the Union's Ethics Guidelines for Trustworthy AI, […] inclusive and diverse design and development of AI systems, including attention to vulnerable persons […] and diversity of the development teams, including gender balance»[45].

These provisions highlight that EU institutions are aware of the fact that datasets and the design of AI systems constitute entry points for bias and their potential discriminatory effects (see §3).

However, regardless of wishing for more representation inside the developers' teams and for stakeholders' participation in the design of AI systems, the AI Act does not make such requirements mandatory, nor it explains how these results should be achieved.

Therefore, criticism has been made that the solutions proposed by the AI Act target technological issues while overlooking the upstream social problems.

One could thus argue that the EU Regulation on artificial intelligence is, to a certain extent, flawed by technological solutionism as well[46].

At the same time, however, it must be noted that the AI Act is a horizontal product compliance Regulation which shapes the rights of EU citizens as "consumer and (product) safety rights"[47]. Then the question is whether such a legal instrument is the most appropriate one to address the problem of (gender) discrimination in healthcare and, above all, to protect the right to health and other human rights[48].

Certainly, the AI Act can constitute a useful starting point but alone it cannot guarantee that we build fair and trustworthy AI systems that make accurate predictions, diagnoses and treatment recommendations for all patients. Which actors shall integrate this source, at which level and by which means remain open questions.

[43] Art. 10(5), Reg. EU 2024/1689.

[44] Recital 67, Reg. EU 2024/1689.

[45] A pivotal role, in this sense, should be played by the AI Office and Member States which are invited to encourage and facilitate the drawing up of codes of conduct (Art. 95, Reg. EU 2024/1689). Specific references to vulnerable persons and to gender balance were not contemplated by the first version of this provision, namely Art. 69 of the AI Act proposal by the EU Commission.

[46] B. PHAM, S.R. DAVIES, *What problems is the AI act solving? Technological solutionism, fundamental rights, and trustworthiness in European AI policy,* in *Critical Policy Studies*, 2 July 2024, 1-19, https://doi.org/10.1080/19460171.2024.2373786.

[47] *Ivi*, 14.

[48] Z. ZŐDI, *The EU AI Act – Can We Protect Human Rights with a Product Compliance Regulation?*, in *IACL-AIDC Blog,* 4 June 2024, available at: https://blog-iacl-aidc.org/2024-posts/2024/6/4/the-eu-ai-act-can-we-protect-human-rights-with-a-product-compliance-regulation (last accessed: 27/06/2024).

Downloaded from www.biodiritto.org.
ISSN 2284-4503

5. Conclusions

In this article, we have conceived vulnerability as a greater propensity to suffer damage or prejudices, focusing on a specific context: healthcare. We have also seen that a context-specific approach to vulnerability can serve to identify and respond to inequalities and different needs.

At the same time, we have argued that vulnerability in healthcare has been shaped by power dynamics. As it is evident that AI is a new form of power, its development and deployment in clinical settings cannot be viewed as a miracle solution but must be carefully addressed.

With regard to gender bias and resulting vulnerabilities, AI is a "double-edged sword"[49]. On the one side, algorithms can amplify and act as an "echo chamber" of existing sex and gender inequalities. Thus, they shall be an object of regulation[50], even if the AI Act alone appears inadequate to target the issues relating to discrimination and inequities.

An option could be integrating this source with sectorial laws, policy instruments or professional guidelines for medical researchers and healthcare professionals. In all cases, including a gender perspective could facilitate a fair development, deployment and use of medical AI.

In fact, just as data quality is affected by the biases of researchers, AI design is heavily influenced by those of developers, which is why the need to reach gender balance and diversity within engineering teams is highlighted by several academic articles, soft law and policy documents[51].

On the other side, algorithms have the potential to reduce these inequalities when properly designed, as they can contribute to bias detection and foster social changes[52].

Indeed, all the attention that AI is gaining is shedding light on bias and their impact[53]. AI could therefore provide an opportunity to reflect on measures to address not only algorithmic biases but also our human biases, such as our prejudices about sex and gender. This way, the advent of AI, when accompanied by a serious political reflection, could allow us to move forward from unfair research and healthcare policies and practices and to leave stereotypes rooted in power imbalances behind.

[49] D. CIRILLO ET AL., *op. cit.*, 1.

[50] S. PENASA, *Verso un diritto "technologically immersive": la sperimentazione normativa in prospettiva comparata,* in *DPCE online*, 1, 2023, 671-696.

[51] Examples of supranational soft law recommending diversity and inclusion in teams developing medical AI systems are, for instance, the ERPS's report "Artificial intelligence in healthcare: Applications, risks, and ethical and societal impacts" (2022), available at: https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU(2022)729512 (last accessed 21/11/2024) and the WHO's "Guidelines on Ethics and Governance of Artificial Intelligence for Health" (2021), available at: https://www.who.int/publications/i/item/9789240029200 (last accessed 21/11/2024).

[52] H.S. SÆTRA, E. SELINGER, *The Siren Song of Technological Remedies for Social Problems: Defining, Demarcating, and Evaluating Techno-Fixes and Techno-Solutionism*, in *SSRN Electronic Journal*, 19 September 2023, 1-18, http://dx.doi.org/10.2139/ssrn.4576687.

[53] S. LINDGREN, V. DIGNUM, *Beyond AI solutionism*, in S. LINDGREN (ed.), *Handbook of Critical Studies of Artificial Intelligence: sociopolitical process and decisions become computationally streamlined*, Cheltenham, 2023, 167-172.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

FEMaLe: la compatibilità di un modello predittivo per l'endometriosi con la tutela dei dati della salute riproduttiva femminile

Vanessa Previti*

FEMaLe: THE COMPATIBILITY OF A PREDICTIVE MODEL FOR ENDOMETRIOSIS WITH THE PROTECTION OF WOMEN'S REPRODUCTIVE HEALTH DATA

ABSTRACT: This contribution analyzes FEMaLe project, which develops a machine learning platform capable of analyzing omics data series and feeding information into a personalized predictive model to improve the women's conditions affected by endometriosis. This study verifies the respect of the protection of female data with reference to Femtech app, which are not being clearly qualified as medical devices. Subsequently, will be analyzed the transfer of such data to third parties, without consent, starting from the American decision regarding the sharing with Facebook and Google, to underline the balance between health and privacy, analyzing the possible replacement with synthetic data.

KEYWORDS: Endometriosis; Femtech; artificial intelligence; synthetic data; inverted privacy.

ABSTRACT: In questo contributo si analizzerà il progetto FEMaLe, che ha come obiettivo l'elaborazione di una piattaforma di *machine learning* che analizza i dati omici e immette informazioni in un modello predittivo personalizzato per le donne affette da endometriosi. Se ne verificherà l'impatto sulla protezione dei dati sanitari gestiti dalle applicazioni. Infine, si analizzerà la cessione degli stessi ai terzi, partendo dal provvedimento della *Federal Trade Commission vs Flo Health Inc.* in merito alla condivisione dei dati sanitari con *Facebook, Google*, tentando infine di risolvere la questione del bilanciamento tra il diritto alla salute e alla riservatezza servendosi dei dati sintetici.

PAROLE CHIAVE: Endometriosi; Femtech; intelligenza artificiale; dati sintetici; dispositivi medici.

* *Dottoranda di Ricerca in Scienze delle Pubbliche Amministrazioni e Cultrice di Diritto Privato e Nuove tecnologie, Università di Messina. Mail: vanessa.previti@studenti.unime.it.* *Contributo sottoposto a doppio referaggio anonimo.*

Downloaded from www.biodiritto.org.
ISSN 2284-4503

ambiti di rischio nelle Femtech – 4. Salute riproduttiva femminile, privacy e intelligenza artificiale: la soluzione conclusiva dei dati sintetici.

1. Endometriosi e Femtech: introduzione, definizioni e punti di connessione

La *Female technology* è l'insieme di *software*, kit diagnostici, prodotti e servizi che nasce con l'intento di supportare la salute delle donne[1]. Il termine *Femtech* è stato coniato nel 2016 da Ida Tin, la co-fondatrice danese di Clue, una app per il monitoraggio del ciclo e dell'ovulazione[2]. Sebbene dall'analisi dei mercati[3] risulti agevole comprendere la redditività del settore[4], gli investitori hanno manifestato un maggiore interesse nello sviluppo di applicazioni che controllano la salute riproduttiva femminile durante il ciclo mestruale[5] o nella ricerca di una gravidanza[6] o nella gestione della gestazione[7], destinando degli investimenti inferiori ai *software* di monitoraggio delle malattie femminili o alla menopausa[8].

L'ambito delle *Femtech* ha costituito da subito oggetto di interesse sia sul versante privatistico – essendo un terreno fertile per i produttori di applicazioni e dei cosiddetti IOT *devices*[9] – che sul versante pubblicistico – come strumento di controllo dei dati relativi alla salute riproduttiva femminile, come è accaduto in Iran[10] a seguito della penalizzazione dell'aborto.

Questa ricerca si focalizzerà sullo studio del progetto FEMaLe avente ad oggetto l'elaborazione di un modello predittivo per l'endometriosi – che è una malattia infiammatoria cronica caratterizzata

[1] K. FOLKENDT, "*So, what is femtech anyways?*", 5 settembre 2019, in https://femtechinsider.com/what-is-femtech/ (ultima consultazione 29/11/2024).

[2] *FemTech Founder: An Interview with Clue CEO, Ida Tin*, 11 febbraio 2021, in https://femtech.live/femtech-founder-an-interview-with-clue-ceo-ida-tin/ (ultima consultazione 29/11/2024); E. JARAMILLO, *Femtech in 2019: 13 Trends and Highlights in Women's Health Technology*, 17 dicembre 2019, in https://www.forbes.com/sites/estrellajaramillo/2019/12/17/femtech-in-2019-trends-investment-in-womens-health-technology/ (ultima consultazione 29/11/2024).

[3] S. CANALI, C. HESSELBEIN, *Using and Interpreting FemTech Data: (Self-)Knowledge, Empowerment, and Sovereignt,* in L. BALFOUR (a cura di) *FemTech: Intersectional Interventions* in *Women's Digital Health*, Londra, 2023.

[4] Secondo C. ROSAS, in *The future is femtech: Privacy and data security issues surrounding femtech applications*, in *Hastings Business Law Journal*, 2, 2019, 319 ss., il valore stimato del mercato delle Femtech era di un bilione di dollari, secondo K. GILDING, *Which femtech apps can you trust?*, 14 febbraio 2020, in https://www.medicalplasticsnews.com/news/which-femtech-apps-can-you-trust/ (ultima consultazione 29/11/2024), il valore del mercato entro il 2025 sarà di 50 bilioni di dollari.

[5] N. FELIZI, J. VARON, *Menstruapps-how to turn your period into money (for others)*, in https://chupadados.codingrights.org/en/menstruapps-como-transformar-sua-menstruacao-em-dinheiro-para-os-outros-2/ (ultima consultazione 29/11/2024).

[6] Principali startup "period tracker": Flo, WomanLog, Glow, NaturalCycles,Maya, Ovia, iOS13, watchOS6, Femometer.

[7] Z. KLEINMAN, *The abortion privacy dangers in period trackers and apps*, 28 giugno 2022, in https://www.bbc.com/news/technology-61952794 (ultima consultazione 29/11/2024).

[8] C. BALTZER, S. BONACINA, *Enhancing Women's Health: An Assessment of Data Privacy and Security of Menopause FemTech Applications* in *Studies in health technology and informatics,* 309, 2023, 155-159.

[9] M. MEHRNEZHAD, T. VAN DER MERWE, M. CATT, *Mind the FemTech gap: regulation failings and exploitative systems* in *Frontiers in The Internet of Things*, 3, 2024, 1-14.

[10] *Iran death penalty threat for abortion unlawful: UN rights experts*, 16 novembre 2021, in https://news.un.org/en/story/2021/11/1105922 (ultima consultazione 29/11/2024).

Downloaded from www.biodiritto.org.
ISSN 2284-4503

dall'anomala crescita di tessuto endometriale al di fuori della cavità uterina[11], comunemente associata al dolore pelvico, infertilità e ad una deteriore qualità della vita – e sulla compatibilità dello stesso con la tutela dei dati relativi alla salute riproduttiva femminile. Sebbene negli ultimi anni si stia attenzionando la malattia, i ritardi nelle diagnosi risultano ancora notevoli – si stimano in un arco temporale che va dai 4 agli 11 anni[12] – e questo ha generato il bisogno di elaborare dei nuovi piani diagnostici che permettano sia di individuare le cause della malattia, sia di abbattere i tempi di diagnosi attraverso l'ausilio dell'IA.

La *Start-up DotLab*, ad esempio, offre uno strumento diagnostico plurifunzionale (app *i-Endometriosis*) per supportare il *management* della malattia: fornisce alle pazienti informazioni generali sulla patologia, consente la quali-quantificazione del dolore attraverso la compilazione di un diario mensile e l'utilizzo di un Misuratore VAS[13] (Visual Analogue Scale) e individua un elenco dei centri specializzati di riferimento. Lo strumento comprende anche funzioni per il ginecologo, che accede ai dati inseriti dalla paziente e al materiale scientifico[14].

Prima di procedere all'analisi del progetto FEMaLe, si è ritenuto opportuno richiamare degli studi clinici che consentono di evidenziare, seppur brevemente, l'approccio multidisciplinare che è necessario adottare per comprendere realmente la malattia e l'impatto che la medesima ha sulla vita delle pazienti.

Il primo studio[15] analizzato da questa ricerca ha ad oggetto l'individuazione di prove genetiche che permettono di trovare un collegamento tra l'endometriosi e altri tipi di dolore cronico: lo stesso ha analizzato il DNA di 60.000 donne affette dalla malattia, individuando 42 regioni nel genoma con varianti che aumentano considerevolmente il rischio di ammalarsi e correlazioni genetiche con 11 tipi di dolore acuto. Da questa analisi è emerso il fondamento genetico della malattia, la possibilità di trattare il dolore con strumenti non ormonali e lo spazio all'elaborazione di una terapia genica[16] che permetta di intervenire allo stato iniziale della patologia con conseguente miglioramento delle condizioni di vita.

Il secondo[17], invece, si è occupato dell'analisi dei bio-marcatori microRNA quale strumento di diagnosi precoce in maniera non invasiva che consente di intervenire sulla progressione della malattia prima dell'insorgenza della sintomatologia.

[11] K.T Zondervan, C.M Becker, S.A Missmer, *Endometriosis* in *The New England journal of medicine*, 382, 13, 2020, 1244-1256.

[12] R. Greene, P. Stratton, S.D. Cleary et al., *Diagnostic experience among 4,334 women reporting surgically diagnosed endometriosis* in *Fertility and sterility,* 91, 1, 2009, 32–39.

[13] M.C. Teodoro, F. Genovese, G. Rubbino, *Endometriosi e dolore pelvico: trattamento laparoscopico*, in Aa. Vv. (a cura di) *Atti della Società Italiana di Ginecologia e Ostetricia*, LXXXVII.

[14] C. Tarabbia, *L'evoluzione tecnologica digitale per la salute della donna: luci ed ombre in medicina*, in *BioLaw Journal – Rivista di BioDiritto*, 3, 2023, 11-26.

[15] N. Rahmioglu, S. Mortlock, M. Ghiasi et al., *The genetic basis of endometriosis and comorbidity with other pain and inflammatory conditions*, in *Nature Genetics*, 55, 2023, 423–436.

[16] P.L. Lollini, *Terapia genica*, in *Le biotecnologie e la qualità della vita*, 2005,129-138.

[17] S. Moustafa, M. Burn, R. Mamillapalli et al., *Accurate diagnosis of endometriosis using serum microRNAs*, in *American journal of obstetrics and gynecology*, 223, 4, 2020.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Il terzo[18] e il quarto[19] studio – che sono il fondamento scientifico del progetto FEMaLe – hanno analizzato l'utilizzo degli algoritmi di *machine learning* non già nella fase di cura ma nella precedente fase di *screening* sia, nel primo caso, tramite i dati di 800 partecipanti francesi che hanno fisicamente preso parte al *trial*, sia, nel secondo, attraverso i dati autonomamente inseriti dalle donne. In entrambi gli studi è stato evidenziato come l'addestramento di algoritmi di intelligenza artificiale alla sintomatologia rappresentata dalle donne abbia permesso di dimezzare i tempi di diagnosi, attraverso un precoce approccio con i medici.

L'inserimento dell'intelligenza artificiale nella relazione terapeutica costituisce un valido coadiuvante dell'interesse primigenio del paziente e cioè quello di ricevere una diagnosi precisa nel più breve tempo possibile, un piano terapeutico personalizzato e, ove possibile, in combinato con le risultanze genetiche, uno strumento predittivo di patologie.

Il limite individuato da svariati studi[20] all'utilizzo degli algoritmi di *machine learning* è il medesimo che ha costituito la *ratio* di genesi delle *Femtech* e cioè la sottorappresentazione di talune minoranze.

Infatti, se da un lato l'inesatto addestramento degli algoritmi può comportare discriminazioni, dall'altro il fenomeno Femtech nasce dal «femminismo liberale che sostiene che le donne sono state storicamente escluse dalle sperimentazioni cliniche, dal processo decisionale e dalla ricerca medica portando alla produzione di dati sanitari prevalentemente incentrati sugli uomini»[21].

Pertanto, un utilizzo davvero intelligente degli algoritmi deve partire proprio da un capillare inserimento dei dati relativi alla salute riproduttiva femminile, per costituirne un ausilio e non un nocumento, come si analizzerà nei paragrafi che seguono.

Il metodo che verrà utilizzato in questa ricerca sarà quello analitico: si partirà dallo studio dei fondamenti del progetto FEMaLe, con *focus* sulle applicazioni inserite nel mercato digitale e sulla disciplina giuridica delle *Health MobileApp*, per poi effettuare una comparazione tra la normativa europea in materia di protezione dei dati personali di carattere sanitario e l'HIPAA statunitense con specifico riferimento all'analisi giurisprudenziale dei casi di cessione dei dati relativi alla salute riproduttiva femminile e concludere con una soluzione assiologicamente orientata per tutelare realmente le donne e cioè quella relativa all'utilizzo dei dati sintetici.

1.1. Progetto FEMaLe: l'inserimento della IA nella diagnosi dell'endometriosi

Il progetto FEMaLe – *Finding Endometriosis using Machine Learning*, si inserisce in uno studio globale finanziato dall'Unione Europea che insieme con ERIN[22] (*Ethically Responsible INnovations in reproductive medicine*) – che si occupa di individuare delle soluzioni eticamente compatibili e finalizzate al

[18] S. Bendifallah, A. Puchar et al., *Machine learning algorithms as new screening approach for patients with endometriosis* in *Scientific Reports*, 12, 2022.

[19] A. Goldstein, S. Cohen, *Self-report symptom-based endometriosis prediction using machine learning* in *Scientific Reports*, 4, 2023.

[20] G.E. Cetera, A.E. Tozzi et al., *Artificial Intelligence in the Management of Women with Endometriosis and Adenomyosis: Can Machines Ever Be Worse Than Humans?* in *Journal of Clinical Medicine*, 16, 2024, 2950.

[21] E. Maestri, *FEMtech e l'avvento della medicina pervasiva: incubo o nobile sogno?*, in *BioLaw Journal – Rivista di BioDiritto*, 3, 2023.

[22] https://erin.ut.ee (ultima consultazione 29/11/2024).

Downloaded from www.biodiritto.org.
ISSN 2284-4503

benessere della paziente – e TRENDO[23] (*Translational Research on Endometriosis*) – che unisce esperti di scienza traslazionale e investitori per individuare bio-marcatori e test non invasivi per abbattere il ritardo nella diagnosi e prevedere un accesso equo all'assistenza sanitaria – mira a indirizzare la ricerca allo studio dell'endometriosi che, essendo una patologia invalidante, incide notevolmente sulla qualità della vita delle donne che ne sono affette.

L'obiettivo di FEMaLe è quello di utilizzare l'intelligenza artificiale nella multiomica che è «l'analisi integrale e completa ottenuta dall'insieme dei risultati di tecniche quali genomica, trascrittomica, proteomica, glicomica, metabolomica, epigenomica»[24], attraverso l'elaborazione di algoritmi che tramite la processazione dei dati omici e delle informazioni fornite dalle pazienti possano elaborare un modello predittivo personalizzato che consenta una diagnosi precoce della malattia – che ad oggi può avvenire tramite supporto radio-diagnostico ovvero con esame istologico del tessuto endometriale acquisito tramite laparoscopia – con metodi non invasivi.

Inoltre, si auspica di trattare la sintomatologia con tecniche innovative in sostituzione della terapia ormonale combinata soppressiva per fornire alla ricerca scientifica dati per elaborare delle alternative farmaceutiche che possano migliorare la vita delle pazienti.

L'iniziativa opera su tre piani: il primo fa riferimento all'elaborazione di un'applicazione che permetta alle pazienti di inserire la propria sintomatologia, il trattamento farmacologico a cui sono sottoposte, le abitudini alimentari e quanto può costituire la base di partenza per l'addestramento degli algoritmi; il secondo attiene all'elaborazione di tre strumenti di supporto alle decisioni cliniche (CDS)[25] per operatori sanitari che consentono di integrare i dati forniti dal paziente con i *Big Data* per incrementare l'accuratezza della valutazione prognostica della malattia; il terzo riguarda un *software* di realtà aumentata che consente la fenotipizzazione chirurgica utilizzando l'apprendimento automatico[26] con l'obiettivo di sviluppare un algoritmo di *deep learning* per valutare in tempo reale la stadiazione dell'endometriosi.

1.2. L'applicazione Lucy: un valido ausilio per la diagnosi precoce ma un problema di qualificazione giuridica

Come affermato nel paragrafo precedente, il primo passo del progetto FEMaLe è stato quello di sviluppare un'applicazione che costituisse un ausilio per la diagnosi dell'endometriosi.

Lucy viene qualificata come una *mobile health application* elaborata con l'intento di suggerire alle donne che ne effettuano il *download* di rivolgersi ad un professionista sanitario nell'eventualità in cui, attraverso l'inserimento di dati quali la durata del ciclo mestruale, lo stile di vita, la presenza di una sintomatologia dolorosa e invalidante, emerga che sia necessario approfondire la condizione clinica.

[23] https://cordis.europa.eu/project/id/101008193/it (ultima consultazione 29/11/2024).

[24] https://www.aboutpharma.com/scienza-ricerca/multi-omica-diagnosi-terapia-diabete/ (ultima consultazione 29/11/2024).

[25] A.E. Tozzi, *Verso una leadership clinica dell'intelligenza artificiale per la salute*, in *L'Endocrinologo*, 24, 2023, 219–223.

[26] Per approfondimenti, https://findingendometriosis.eu/it/di/pacchetti-di-lavoro/ (ultima consultazione 29/11/2024).

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Tramite tale *software*, è stato effettuato uno studio multicentrico[27] attraverso la somministrazione di questionari a 5.000 donne con endometriosi diagnosticata e 5.000 a cui non era stata fornita alcuna diagnosi, per la durata di un anno al fine di raccogliere dati riguardanti i sintomi, la salute mentale, fisica e le condizioni socio-demografiche, fattori economici, informazioni inerenti la dieta e in generale lo stile di vita al fine di elaborare un unico *database*.

La combinazione di questi dati acquisiti tramite l'applicazione *Lucy* e le informazioni sanitarie contenute nelle cartelle cliniche delle pazienti attraverso l'utilizzo delle tecniche di *machine learning* ha come principale obiettivo quello di intervenire precocemente sulla diagnosi, specie negli stadi iniziali di endometriosi[28] o nell'ipotesi di valutazione ecografica negativa[29].

In subordine si è inteso vagliare eventuali peggioramenti o miglioramenti nella sintomatologia a seguito di variazione di taluni alimenti dalla dieta.

Infine, si è approfondito il collegamento tra dolore e talune tecniche di relax o di *mindfulness*[30].

L'obiettivo principale è stato quello di elaborare una descrizione fenotipica della paziente affetta da endometriosi, con l'ausilio degli strumenti di intelligenza artificiale, che permettesse all'utente che inserisce nell'applicazione *Lucy* determinati parametri di ricevere il "suggerimento" di rivolgersi ad un medico per approfondire la sintomatologia.

Tale "suggerimento" consente di affrontare una questione giuridica di rilevante importanza e cioè la qualificazione delle applicazioni relative alla salute in generale e nel particolare quella riproduttiva femminile.

Le applicazioni *Femtech*, così come gli oggetti e i dispositivi dotati di sensori che permettono di trasmettere e ricevere dati, da o verso altre cose e sistemi[31] facenti parte dell'IOT (*Internet of things*)[32] che ineriscono a tale settore, rientrano nell'ambito più ampio rinominato *e-Health*.

Con tale termine si fa riferimento a quel campo della medicina in cui si intersecano «informatica medica, sanità pubblica e attività economica, ricomprendente tutti quei servizi e quelle informazioni sanitarie forniti o condivisi attraverso l'uso di tecnologie informatiche e di telecomunicazione»[33].

Costituendo oggetto di grande interesse da parte dell'Unione europea, la Commissione europea ha evidenziato la necessità di procedere alla digitalizzazione delle informazioni mediche dei cittadini europei per garantire l'abbattimento delle barriere linguistiche e materiali che di fatto ostacolano la fruizione dei servizi sanitari e delle innovazioni terapeutiche.

---

[27] D.B. Balogh, G. Hudelist et al., *FEMaLe: The use of machine learning for early diagnosis of endometriosis based on patient self- reported data—Study protocol of a multicenter trial*, in *PLoS ONE*, 2024.

[28] J. Keckstein, E. Saridogan et al., *The #Enzian classification: A comprehensive non-invasive and surgical description system for endometriosis*, in *Acta Obstet Gynecol Scand*, 100, 7, 2021, 1165-1175.

[29] Becker et al., *op. cit.*

[30] Nell'ambito del progetto FEMaLe si segnala Myendo (Mind Your ENDOmetriosis) che includerà metodi innovativi per la gestione psicologica del dolore per ridurre il potenziale stigma associato alle ripercussioni sulla salute mentale.

[31] Tra gli altri *Elvie smart pump, Daysy cicle computer, Lady comp fertility tracker.*

[32] M. Mehrnezhad et al, *op. cit.*

[33] C. Irti, *L'uso delle "tecnologie mobili" applicate alla salute: riflessioni al confine tra la forza del progresso e la vulnerabilità del soggetto anziano*, in *Persona e Mercato*, 1, 2023, 34.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Fin dal 2004, con il primo piano d'azione per la realizzazione della "sanità digitale": l'*e- Health Action Plan*[34] e, di recente, con la Comunicazione *A European Health Data Space: harnessing the power of health data for people, patients and innovation*[35] di accompagnamento alla Proposta di Regolamento sulla creazione dello Spazio europeo dei dati sanitari[36].

Infatti, anche se ogni secondo viene generata una grande quantità di dati sanitari che può fornire ai ricercatori informazioni rilevanti e che costituisce "l'oro" del settore medico[37], la mancanza di uniformità di normative crea degli ostacoli che si riverberano sull'assistenza dei pazienti.

La pandemia da Covid-19 ha evidenziato un sistema sanitario vetusto, fortemente ancorato alla materialità del dato e ha accelerato il processo di digitalizzazione del medesimo, favorendo anche la diffusione dei dispositivi medici portatili per il controllo a distanza di parametri vitali o sanitari dei pazienti e «la fruizione on-line dei servizi sanitari di quel vasto e complesso orizzonte medico-sociale che comprende al suo interno la telemedicina e la telechirurgia»[38].

Accanto alla diffusione di strumenti che permettono di esercitare la professione sanitaria a distanza, si è assistito all'ampia espansione di applicazioni *software* che sono finalizzate al controllo e alla raccolta di dati sanitari, parametri vitali e abitudini comportamentali degli utenti, programmate per essere scaricate su dispositivi mobili e monitorare tramite biosensori il proprio stato di salute al fine di fornire dei suggerimenti, con l'ausilio dell'intelligenza artificiale, per il miglioramento della propria condizione clinica, in assenza di un controllo medico.

La mancanza di un monitoraggio da parte di un sanitario costituisce il nodo del problema relativo alla qualificazione giuridica delle applicazioni.

La normativa eurounitaria in materia[39] definisce "dispositivo medico" anche «il software progettato per funzionare su apparecchiature portatili di uso comune, purché sia esplicitamente destinato dal fabbricante ad essere impiegato per una o più destinazioni ad uso medico (quali ad esempio diagnosi, prevenzione, monitoraggio, previsione, prognosi, trattamento o attenuazione di malattie)», pertanto il *discrimen* tra *Medical MobileApp* e *Health MobileApp* risiede nella mera destinazione d'uso del fabbricante[40].

[34] Comunicazione 2004/356 rinnovata con le Comunicazioni 2012/736 e 2018/233.

[35] COM (2022) 196 del 3.5.2022.

[36] COM (2022) 197 del 3.5.2022.

[37] Si stima che l'utilizzo secondario dei dati sanitari abbia un valore di circa 25-30 miliardi di euro all'anno e si prevede che questa cifra possa raggiungere 50 miliardi di euro entro dieci anni.

[38] A. MARCHESE, *Profili civilistici dell'information technology in ambito sanitario*, in *Quaderni della Rassegna di diritto civile*, Napoli, 2021.

[39] I Regolamenti UE 2017/745 e 2017/746 disciplinano in modo uniforme in tutti gli Stati membri il settore dei dispositivi medici.

[40] Sul punto, la Corte di Giustizia dell'Unione Europea nella sentenza relativa alla Causa C-329/16, Snitem e Philips France, la quale - nel vigore della dir. 93/42/CEE – aveva affermato che per ricadere nell'ambito di applicazione della direttiva, non è sufficiente che un software sia utilizzato in un contesto medico, ma occorre anche che la sua finalità, definita dal fabbricante, debba essere specificamente medica. Contestualmente ha ritenuto irrilevante, ai fini della qualificazione come dispositivo medico, il fatto che il software agisca o non agisca direttamente sul corpo umano, rinvenendo quale unica condizione fondamentale quella legata alla sua finalità: pertanto un software che tra le sue funzionalità, consenta l'utilizzo dei dati personali di un paziente, allo scopo di rilevare le controindicazioni, le interazioni tra medicinali e le posologie eccessive, costituisce, quanto a tale funzionalità, un dispositivo medico, indipendentemente dal suo agire o meno direttamente nel o sul corpo umano.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

La qualificazione di un'applicazione quale dispositivo medico o *software* del benessere ha notevoli implicazioni dal punto di vista giuridico, clinico ed etico.
Dal punto di vista giuridico, la collocazione nel novero dei "dispositivi medici" implica l'assoggettamento ad una sequela di valutazioni, controlli e verifiche scientifiche da parte di Organismi Notificati individuati dalle Autorità competenti dei singoli Stati membri[41], mentre categorizzare un *software* come attinente al benessere dell'individuo, consente al produttore di arginare le normative europee previste in materia, mantenendo comunque un ruolo dominante nel mercato della salute.
Dal punto di vista clinico, i suggerimenti forniti possono avere conseguenze rilevanti sia sul versante psicologico che medico, essendosi registrati plurimi casi, ad esempio, di gravidanze indesiderate a seguito di segnalazioni da parte di applicazioni della c.d. finestra fertile[42].
Dal punto di vista etico, la questione evidenzia l'algocrazia[43] e la conseguente algoretica[44] della realtà circostante, sottolineando la rinnovata centralità della macchina a sfavore della persona che diventa soltanto un involucro di dati.
Il problema più rilevante diventa, infatti, individuare il regime giuridico di trattamento applicabile all'uso dei dati nella *mobile health* in generale e nel particolare alle informazioni relative alla salute riproduttiva femminile.

1.3. Il regime giuridico del trattamento dei dati nella mobile health e l'assenza di una disciplina specifica riguardante le informazioni sanitarie dei soggetti vulnerabili

Nel novero dell'analisi del regime giuridico applicabile ai dati sanitari che albergano nelle applicazioni, è necessario distinguere l'eventualità in cui questi circolino nelle *Medical MobileApp* ovvero nelle *Health Mobile App*, come Lucy, il *software* analizzato nel paragrafo precedente.
È d'uopo effettuare una premessa in merito al trattamento dei dati sanitari femminili, tenuto conto della mancanza nel Regolamento Generale sulla Protezione dei Dati (GDPR) di una disciplina specifica, gran parte della letteratura scientifica[45] applica l'art. 9, essendo un sottoinsieme delle "categorie particolari di dati" che include informazioni sulla salute, la vita sessuale e l'orientamento sessuale.
A ben vedere, però, sussumere le informazioni femminili nella categoria di cui sopra non sarebbe corretto in quanto talvolta possono disvelare altri aspetti della vita umana più complessi quale ad esempio le opinioni politiche o religiose.

[41] G. CAPILLI, *Diritto privato sanitario*, Pisa, 2022, 54.
[42] J. TRAN, *Natural Cycles: When an Algorithm Digitally Mandates Your Sexual Health*, in *SMU Science and Technology Law Review*, 22, 1, 2019.
[43] F. ZAMBONELLI, *Algocrazia, il governo degli algoritmi e dell'intelligenza artificiale*, Trieste, 2020, l'algocrazia viene definita come «il crescente utilizzo degli algoritmi informatici e dell'intelligenza artificiale al fine di esercitare il controllo di qualsiasi aspetto della vita quotidiana degli individui».
[44] P. BENANTI, *Oracoli. Tra algoretica e algocrazia*, Roma, 2018, l'algoretica invece è «lo studio dei problemi e dei risvolti etici connessi all'applicazione degli algoritmi».
[45] Tra le altre, A. THIENE, *Protezione dei dati sensibili e uso di App per il benessere delle donne. Una questione di consapevolezza*, in *BioLaw Journal – Rivista di BioDiritto*, 3, 2023.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Ad esempio, il Corano proibisce i rapporti durante le mestruazioni per evitare di causare disagio a una donna[46], quindi, il mancato inserimento dei dati relativi al sesso nel periodo mestruale potrebbe consentire di svelare l'appartenenza religiosa dell'utente.

Pertanto, si ritiene necessario prevedere una disciplina specifica che possa tutelare realmente la riservatezza delle donne.

Inoltre, il trattamento dei dati "particolari" nell'ambito delle *Medical Mobile app* non incontra grandi differenze rispetto alle informazioni che circolano nelle tradizionali relazioni di cura, essendo previsto un divieto generale di trattamento a cui fanno seguito due sottoinsiemi di eccezioni: il primo riguarda le ipotesi in cui vi siano degli interessi superindividuali o una finalità terapeutica[47], il secondo attiene alla prestazione del consenso dell'avente diritto, base giuridica per eccellenza.

La questione diventa più complessa quando i dati dei pazienti vengono trattati da applicazioni che implicano l'utilizzo di sistemi di apprendimento automatico o di intelligenza artificiale «in sostituzione del medico o in sua assenza, in grado di indicare al paziente una terapia o un protocollo»[48], dal momento che è un trattamento che «fornisce decisioni automatizzate destinate ad impattare in maniera diretta sulla loro stessa salute e sul loro benessere»[49].

Di fatti il GDPR lo vieta, a meno che non vi sia il consenso dell'interessato o altra idonea base giuridica, anche se sussistono dei dubbi in merito alla spontaneità della prestazione del consenso nell'ambito di un'applicazione medicale, tenuto anche conto della mancanza di una previa adeguata «comunicazione simmetrica e reciproca tra medico e paziente»[50].

Per quel che occupa questo studio, è necessario focalizzare l'attenzione sulle applicazioni di salute e benessere che costituiscono la maggioranza dei *software Femtech*.

La base giuridica richiesta è il consenso espresso e si tratta, nella maggior parte dei casi, di applicazioni scaricabili "gratuitamente" o, per lo meno, apparentemente, in quanto, sebbene sembri che si acconsenta al mero *download* del *software* il "guadagno" del fornitore è rinvenibile nell'autorizzazione a profilare l'utente e a trattare i suoi dati sensibili in un mercato in cui il confine tra ricerca scientifica e commerciale è particolarmente labile.

Inoltre, essendo la destinazione d'uso individuata dal produttore l'unico *discrimen* tra applicazioni medicali e invece quelle afferenti alla salute e al benessere, spesso per arginare le normative sulla privacy più stringenti, *software* che controllano la fertilità o patologie uterine ovvero la gravidanza, rientrano nel novero di queste ultime.

---

[46] Versetto 222 del Corano che recita «Ti chiederanno dei [rapporti durante i] mestrui. Di': «Sono un danno. Non accostatevi alle vostre spose durante i mestrui e non avvicinatele prima che si siano purificate. Quando poi si saranno purificate, avvicinatele nel modo che Allah vi ha comandato». In verità Allah ama coloro che si pentono e coloro che si purificano».

[47] M. CIANCIMINO, *Protezione e controllo dei dati in ambito sanitario e intelligenza artificiale*, Napoli, 2020, 36 ss.

[48] C. IRTI, *op. cit.*

[49] C. BOTRUGNO, *Tecnologie dell'informazione e della comunicazione e tutela della salute: le sfide aperte tra protezione, circolazione e riutilizzo dei dati*, in *Diritto e questioni pubbliche*, 2, 2020, 137 ss.

[50] Secondo il Comitato italiano di bioetica, nello scritto "Mobile Health e applicazioni per la salute: aspetti bioetici" del 28 maggio 2015, «la comunicazione è simmetrica quando i singoli sono parimenti forti nell'interazione e reciproca quando le posizioni fra chi dà l'informazione e chi la riceve si realizzi nel riconoscimento delle rispettive autonomie».

Downloaded from www.biodiritto.org.
ISSN 2284-4503

La mancata assimilazione ai dispositivi medici e il meccanismo di *opting-out*[51] sovente utilizzato dalle applicazioni per la fertilità, implica che l'utente effettua il *download* senza avere reale contezza dell'utilizzo dei propri dati, in quanto la mera scelta binaria tra accettare e rifiutare il trattamento crea, nelle donne che si collocano in una posizione di vulnerabilità e vogliono monitorare il proprio ciclo mestruale o la propria malattia, un meccanismo di analisi privo di una valutazione assiologica che le porta ad acconsentire, spinte dalla volontà di acquisire padronanza del proprio corpo, al trattamento dei dati.

2. La cessione dei dati femminili a soggetti terzi: GDPR e HIPAA

Il punto focale del problema dei dati che vengono immessi nelle applicazioni *Femtech* è che se le tecniche di automonitoraggio stanno rivoluzionando la gestione del proprio corpo, lo stesso finisce col diventare un involucro di dati digitali visualizzabili tramite tabelle e grafici[52] che contribuiscono ad alimentare i Big data, utilizzati da aziende farmaceutiche, governi e centri di ricerca[53] per analizzare e tracciare le scelte e gli stili di vita[54].

Per garantire una maggiore tutela sarebbe dunque necessario assimilare tutte le applicazioni *Femtech* a dispositivi medici, per applicarne la disciplina.

L'Autorità Garante per la *privacy* italiana è intervenuta fornendo dei chiarimenti sul trattamento dei dati personali in ambito sanitario[55], sottolineando che laddove si ravvisi la finalità di cura non è richiesta la base giuridica del consenso al trattamento mentre per «trattamenti connessi all'utilizzo di App mediche, attraverso le quali autonomi titolari raccolgono dati, anche sanitari dell'interessato, per finalità diverse dalla telemedicina oppure quando, indipendentemente dalla finalità dell'applicazione, ai dati dell'interessato possano avere accesso soggetti diversi dai professionisti sanitari o altri soggetti tenuti al segreto professionale», è necessario il consenso esplicito dell'interessato.

Negli Stati Uniti la normativa di riferimento è l'HIPAA (*Health Insurance Portability and Accountability Act*): una legge federale che prevede la «responsabilità degli operatori sanitari, dei piani sanitari, delle camere di compensazione per la sanità, e dei loro associati che trasmettono telematicamente i dati sulla salute e le informazioni correlate (PHI)»[56].

Le regole HIPAA non si applicano alle applicazioni mobili *Femtech* e pertanto anche negli Stati Uniti d'America così come in Europa, i dati sanitari utilizzati non soggiacciono alle regole più rigide previste dalla normativa statunitense o al GDPR.

Dunque, *a contrario*, l'unica base giuridica legittimante riconosciuta in materia è il consenso.

---

[51] I.A. Caggiano, *Il consenso al trattamento dei dati personali tra Nuovo Regolamento Europeo e analisi comportamentale* in *Annali-Università degli Studi Suor Orsola Benincasa*, Annali 2016-2018, 11.1.

[52] S. Sumartojo, S. Pink, D. Lupton, C. Labond, *The affective intensities of datafied space*, in *Emotion, Space and Society*, 21, 2016, 33-40.

[53] E. Maestri, *FEMtech e l'avvento della medicina pervasiva: incubo o nobile sogno?* in *BioLaw Journal – Rivista di BioDiritto*, 3, 2023.

[54] K.D. Haggerty, R.V. Ericson, *The Surveillant Assemblage, in The British Journal of Sociology*, 51, 2000, 605-622.

[55] Provvedimento n. 55 del 7 marzo 2019, doc. web. 9091942, https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9091942 (ultima consultazione 29/11/2024).

[56] Health Insurance Portability and Accountability Act (HIPAA), https://www.hhs.gov/hipaa/for-professionals/index.html (ultima consultazione 29/11/2024).

Downloaded from www.biodiritto.org.
ISSN 2284-4503

E' rilevante interrogarsi sui confini e sui limiti del consenso: spesso l'utente manifesta la propria volontà al *download* dell'applicazione, all'inserimento dei propri dati, al monitoraggio degli stessi ma non è a conoscenza del fatto che gli stessi verranno venduti a soggetti terzi con finalità di profilazione e controllo, dunque l'unica scriminante che avrebbe escluso l'antigiuridicità del fatto non può essere considerata dal momento che manca un'informativa reale sull'utilizzo secondario dei dati o sulla cessione dei medesimi per finalità di lucro.

Negli Stati Uniti d'America, il *Federal Trade Commission* ha imposto[57] alla società Flo Health Inc. – titolare dell'Applicazione Flo, *leader* nel monitoraggio della fertilità – di chiedere e ottenere il consenso delle utenti prima di condividere i propri dati sensibili con soggetti terzi, a seguito di un procedimento in cui la società era stata condannata perché, sebbene promettesse di mantenere i dati sanitari privati, aveva condiviso tali informazioni sensibili riconducibili a milioni di utenti con società di *marketing* e di profilazione quali Facebook e Google[58], sotto forma di *app events* cioè un trasferimento di dati senza un fine specifico, non prevedendo limiti all'utilizzo degli stessi.

Nel 2020, il Procuratore generale della California ha condannato *Glow*[59] – altra azienda *leader* nel settore delle *Femtech* – ad una multa e ad implementare le misure di sicurezza in merito alla cessione dei dati ai soggetti terzi per aver violato la *Confidentiality of Medical Information Act* (CMIA) per la facilità con cui era possibile ottenere il cambio *password* e dunque inserirsi nel registro sanitario riproduttivo dell'utente e per aver introdotto uno strumento che le consente di condividere i propri dati con un secondo soggetto, in assenza di adeguata identificazione.

3. La privacy inversa della sorveglianza dell'intimità: ambiti di rischio nelle Femtech.

I dati «sensibilissimi»[60] negli anni sono stati considerati un elemento da proteggere sia nel GDPR sia nell'HIPAA, sia in tutte le normative che si sono occupate di tutela dei dati a vari livelli.

Nei paragrafi precedenti si è affermato che le applicazioni *Femtech* non sono attualmente considerati dispositivi medici; pertanto, per l'utilizzo delle medesime è sempre richiesto il consenso informato.

Ma è la stessa base giuridica che diventa automaticamente viziata quando, pensando di introdurre delle informazioni nei *software* al fine di avere un controllo sul proprio corpo, lo stesso viene meno in quanto quel dato smette di appartenere alla persona e viene ceduto a soggetti terzi, verificandosi un fenomeno che è possibile definire *privacy inversa.*

Nel prosieguo si analizzeranno diverse ipotesi da cui emergono i rischi della sorveglianza dell'intimità[61].

Il primo caso riguarda l'utilizzo dei dati *Femtech* per ottenere una pubblicità mirata e concerne l'eventualità in cui a fronte del mancato inserimento delle mestruazioni in qualsiasi applicazione di controllo

[57] https://www.ftc.gov/news-events/news/press-releases/2021/01/developer-popular-womens-fertility-tracking-app-settles-ftc-allegations-it-misled-consumers-about (ultima consultazione 29/11/2024).

[58] Federal Trade Commission in the matter of Flo Health inc, file no. 1923133.

[59] https://oag.ca.gov/news/press-releases/attorney-general-becerra-announces-landmark-settlement-against-glow-inc-– (ultima consultazione 29/11/2024).

[60] S. ALLEGREZZA, *Gli archivi di persona tra consultabilità, privacy e diritto all'oblio*, in R. PALLUCCHINI (a cura di) *Storie, archivi, prospettive critiche*, 2019, 417-427.

[61] J. ERICKSON, J. Y. YUZON, T. BONACI, *What you do not expect when you are expecting: privacy analysis of femtech*, in *Transactions on Technology and Society*, 3, 2, 2022.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

del ciclo, l'utente sarà raggiunta da pubblicità su Amazon o su Google riguardanti i test di gravidanza, pur non avendo acconsentito in alcun modo a questo genere di profilazione.
Secondariamente, i dati inerenti alla salute riproduttiva femminile possono essere utilizzati anche nel luogo di lavoro[62], come è accaduto negli Stati uniti con l'applicazione Ovia.
Come emerso da un'inchiesta del Washington Post[63], i dati inseriti da alcune donne in Ovia – sia di monitoraggio del ciclo mestruale, sia riguardanti i rapporti sessuali, l'assunzione di contraccettivi, la ricerca di una gravidanza o il tracciamento della medesima – sono stati venduti ai datori di lavoro. Gli stessi non soltanto effettuavano un monitoraggio delle vite delle dipendenti, prevedendo i giorni di congedo mestruale o le richieste di maternità, attuando delle vere e proprie discriminazioni in favore dei colleghi uomini, ma ne incentivavano l'utilizzo attraverso la dazione di un dollaro al giorno in carte regalo per chi inserisse i propri dati nell'app.
Le informazioni inserite nei *software Femtech* vengono, inoltre, utilizzate per contrarre assicurazioni sanitarie; infatti se in un primo momento gli assicuratori, cercando di minimizzare i costi per gli utenti, richiedono l'accesso ai dati relativi alla salute riproduttiva femminile, successivamente, il sacrificio della riservatezza di pochi soggetti produce effetti distorsivi sul mercato, dal momento che non sarà più possibile fruire di tariffe vantaggiose per la stipula senza acconsentire all'utilizzo delle stesse informazioni creando un innalzamento dei prezzi per le donne che non sono intenzionate a cedere i propri dati, inducendole a fornire un consenso viziato per non sostenere costi elevati[64].
Un punto di vista altresì interessante riguarda l'utilizzo delle informazioni sulla fertilità all'interno di relazioni abusivanti[65], nel paragrafo precedente si è citato il caso dell'applicazione *Glow* che concedeva la condivisione dei dati con un secondo soggetto senza procedere ad un'adeguata identificazione e, di recente, anche l'applicazione Flo, anch'essa al centro di un procedimento per inadeguata cessione dei dati, ha introdotto la possibilità di scegliere l'opzione "Flo a due", che permette la totale condivisione delle informazioni riguardanti il ciclo mestruale o la ricerca di una gravidanza o la gestazione, con un *partner* (di sesso maschile, generando una discriminazione a discapito delle coppie LGTBQIA+) a seguito di un semplice abbinamento che avviene con l'erogazione di un codice facilmente individuabile.
La questione incontra dei problemi sia etici che giuridici: innanzitutto, si attua anche in questa circostanza un sistema di *privacy inversa* dal momento che la gestione del corpo di una donna viene affidata, anche al di fuori delle relazioni patologiche, al proprio *partner*, secondariamente nell'ambito di una condizione abusivante il controllo sul ciclo mestruale implica la possibilità di utilizzare la c.d. finestra fertile per avere rapporti sessuali finalizzati alla procreazione, indipendentemente dalla volontà della

[62] E.A. Brown, *The Femtech Paradox: How Workplace Monitoring threatens women's equity*, in *Jurimetrics*, 61, 2021, 289.
[63] D. Harwell, *Is your pregnancy app sharing your intimate data with your boss?* in *The Washington Post*, aprile 2019, https://www.washingtonpost.com/technology/2019/04/10/tracking-your-pregnancy-an-app-may-be-more-public-than-you-think/ (ultima consultazione 29/11/2024).
[64] M. Crossley, *Discrimination Against the Unhealthy in Health Insurance*, in *University of Kansas Law* Review, 73, 2005.
[65] K.F. Clevenger, *Spousal abuse through spyware: The inadequacy of legal protection in the modern age*, in *J. Amer. Acad. Matrimonial Lawyers*, 21, 2008, 653.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

donna ovvero condizionare la propria compagna nella scelta della continuazione della gestazione[66], anche perpetrando violenza psicologica[67].
Un ulteriore problema riguarda l'eventualità in cui i dati inseriti nelle applicazioni possano costituire oggetto di un *data breach*[68], dal momento che le regole che si applicano ai *software femtech*, non essendo qualificati come dispositivi medici, sono meno stringenti e facilmente arginabili, la diretta conseguenza di una tale violazione è che tali dati possano costituire anche oggetto di ricatto, *revenge porn* e di circolazione non autorizzata.
Infine, in considerazione del fatto che nelle applicazioni esistenti sul mercato è prevista la possibilità di servirsi di forum, l'inserimento – sempre su base volontaria – delle proprie informazioni sanitarie, delle proprie patologie, dei farmaci che si assumono potrebbe generare da un lato una cattiva informazione, dall'altro un improprio utilizzo degli stessi da parte degli altri utenti ovvero da parte di soggetti iscritti regolarmente, interessati ad acquisire i dati e ad addestrare algoritmi di *machine learning,* senza che le donne coinvolte ne siano in alcun modo messe a conoscenza.
Per impedire il verificarsi di queste ipotesi, tutte le compagnie *Femtech* dovrebbero irrobustire i propri sistemi di *privacy* e adottare gli standard FAIR[69] per implementare la qualità delle relazioni di cura, come affermato recentemente nella Dichiarazione di Porto[70].

4. Salute riproduttiva femminile, privacy e intelligenza artificiale: la soluzione conclusiva dei dati sintetici

Analizzare il progetto FEMaLe senza approfondire in maniera analitica il contesto delle *Femtech* non sarebbe stato possibile o, probabilmente, non sarebbe stato utile.
Sebbene lo studio di un modello predittivo con l'ausilio dell'intelligenza artificiale per intervenire precocemente sulla diagnosi dell'endometriosi sia innovativo e avanguardista, la ricerca a tutela delle donne deve andare di pari passo con l'elaborazione di una disciplina giuridica che le qualifichi come soggetti di diritto titolari di situazioni giuridiche soggettive specifiche.
Come detto precedentemente, la nascita di questo settore è stata conseguenziale al superamento di alcuni *bias* che escludevano le donne dagli studi clinici o dalle sperimentazioni dei medicinali ma,

[66] S. De Vido, *Under His Eye: riflessioni sul ruolo della tecnologia sul corpo delle donne a seguito della sentenza Dobbs della Corte Suprema degli Stati Uniti* in *BioLaw Journal – Rivista di BioDiritto*, Sp 1, 2023, 343.
[67] E. Galperin, *The State of the Stalkerware*, 29 gennaio 2020, in https://www.usenix.org/conference/enigma2020/presentation/galperin (ultima consultazione 29/11/2024).
[68] C. Rosas, *The future is femtech: Privacy and data security issues surrounding femtech applications* in *Hastings Bus. Law J*., 15, 2, 2019, 319.
[69] M. Boeckhout, G.A. Zielhuis, A.L. Bredenoord, *The FAIR guiding principles for data stewardship: fair enough?* in *Eur J Hum Genet*, 26, 2018, 931–936.
[70] La dichiarazione prevede: «Lo sviluppo e la promozione di standard di qualità dei dati e di etichettatura; lo sviluppo e la certificazione di prodotti sanitari digitali per avere dati di qualità per la progettazione; la richiesta di etichette di qualità dei dati all'interno di tutti i prodotti digitali per la salute e la ricerca; investire nel cambiamento educativo e organizzativo per migliorare la qualità dei dati; avere dati di alta qualità» https://www.i-hd.eu/health-data-forum-2022/ihd-porto-declaration-on-health-data-quality-2022/ (ultima consultazione 29/11/2024).

Downloaded from www.biodiritto.org.
ISSN 2284-4503

tenuto conto dell'innovativo utilizzo degli algoritmi di *machine learning,* è necessario individuare una soluzione che non riporti il problema della discriminazione.

Se, infatti, l'intelligenza artificiale deve intervenire sull'elaborazione di una forma di medicina di precisione che sia il più personalizzata possibile, allo stesso modo non si può temere la compromissione dell'inclusività[71].

La genesi della discriminazione algoritmica è individuabile nell'enorme mole di dati di carattere personale con cui avviene l'addestramento dei *software* che costituisce un rischio per il soggetto a cui queste informazioni appartengono, essendo stato ipotizzato in letteratura il fallimento dell'anonimizzazione[72].

Infatti, sebbene dovesse, in linea teorica, «massimizzare la protezione dei dati personali e, allo stesso tempo, minimizzarne la perdita»[73], le tecniche ad oggi utilizzate forniscono un'utilità inversamente proporzione alla tutela[74], pertanto, se il fine è quello di proteggere le informazioni e comunque trarne giovamento, una soluzione potrebbe essere l'utilizzo dei dati c.d. sintetici, creati tramite una specifica tecnica di anonimizzazione basata su modelli di *machine learning* di tipo generativo[75], con lo scopo di mantenere le caratteristiche originali dei dati ma rimuovendo ogni corrispondenza tra quelli reali e quelli artificiali[76].

Sebbene non ci sia ancora una normativa in materia, l'*European Data Protection Supervisor*, ne ha sottolineato[77] l'importanza, individuandone vantaggi e svantaggi. Se infatti, l'EDPS individua l'equità come conseguenza positiva dell'utilizzo – in quanto un *data set* equo impedisce il verificarsi di discriminazioni – in dottrina[78],invece, ciò viene considerato come un rischio, poiché si teme che i dati generati artificialmente possano riflettere gli stessi pregiudizi esistenti in società, perpetrando comportamenti discriminatori.

Pertanto, nonostante i dati sintetici costituiscano un'innovativa e valida alternativa per tutelare la *privacy* delle donne, è necessario predisporre *ab origine* un addestramento dell'algoritmo che tenga conto delle differenze ma che non eviterebbe del tutto il problema della *privacy inversa*.

Sarebbe, invece, necessario leggere il tema della salute riproduttiva femminile sotto una lente assiologicamente orientata che metta al centro la persona, attraverso una riqualificazione giuridica delle applicazioni di salute e benessere come dispositivi medici, un consenso in merito all'inserimento dei dati e alla conseguente circolazione dei medesimi reale al pari di una manifestazione di autodeterminazione terapeutica maturata all'interno di una sinallagmatica relazione medicale.

---

[71] R. CONFALONIERI ET AL. (a cura di), *Sex and Gender Bias in Technology and Artificial Intelligence: Biomedicine and Healthcare Applications*, Cambridge (MA), 2022, 179-204.

[72] C.A. TROVATO, C. RAUCCIO, *L'anonimizzazione è morta? Un'analisi dei dati sintetici come proposta per superare la dicotomia dato personale-dato non personale*, in *Ciberspazio e Diritto*, 2, 2022.

[73] *Ibidem*.

[74] F. LIU, *A statistical overview on data privacy*, in *Notre dame journal of law, ethics e public policy*, 34, 2010, 477.

[75] K. EL EMAM, L. MOSQUERA, R. HOPTROFF, *Practical Synthetic Data Generation*, 2020.

[76] C. A. TROVATO, C. RAUCCIO, *op. cit.*

[77] https://www.edps.europa.eu/press-publications/publications/techsonar/synthetic-data_en?etrans=it (ultima consultazione 29/11/2024).

[78] C.A. TROVATO, C. RAUCCIO, *op. cit.*; A. GUPTA, D.L. BHATT, A. PANDEY, *Transitioning from Real to Synthetic data: Quantifying the bias in model*, in *Synthetic Data Generation Workshop at ICLR*, 2021.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Special issue

Intelligenza artificiale, sovranità alimentare e *data governance*

Maria Francesca De Tullio*

ARTIFICIAL INTELLIGENCE, FOOD SOVEREIGNTY, AND DATA GOVERNANCE

ABSTRACT: The research concerns artificial intelligence in the agricultural sector, with a focus on 'food sovereignty', as a synthesis of different constitutional values (right to health, duty of rational land use, cultural rights, etc.). Above all, food poverty and environmental risks will be considered as vulnerabilities which intersect with other forms of social disadvantage. In particular, central to the analysis is the collection and processing of environmental and land data by agricultural technology providers. The research aims to recommend measures to ensure equitable data sharing regimes, based on farmers' and peasants' self-determination, but also general interests involved in agricultural data sharing.

KEYWORDS: Food sovereignty; data sovereignty; agroecology; climate smart agriculture; data sharing.

ABSTRACT: La ricerca riguarda l'intelligenza artificiale nel settore agricolo, con particolare attenzione alla "sovranità alimentare", come sintesi di diversi valori costituzionali (salute, uso razionale della terra, diritti culturali, ecc.). Soprattutto, la povertà alimentare e i rischi ambientali saranno considerati come vulnerabilità che si intersecano con altre forme di svantaggio sociale. Al centro dell'analisi vi è la raccolta e l'elaborazione dei dati ambientali e fondiari da parte dei fornitori di tecnologie agricole. Si raccomandano misure per garantire regimi equi di condivisione dei dati, basati sull'autodeterminazione degli agricoltori e dei contadini, ma anche sugli interessi generali coinvolti nella condivisione dei dati agricoli.

PAROLE CHIAVE: Sovranità alimentare; sovranità dei dati; agroecologia; *climate smart agriculture*; condivisione dei dati.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

* *Ricercatrice in Diritto Costituzionale, Università di Napoli Federico II. Mail: mfdetullio@gmail.com.* *Contributo sottoposto a doppio referaggio anonimo.*

1. Introduzione

Il presente lavoro analizza le tecniche di agricoltura intelligente nel contesto eurounitario, con l'obiettivo di contribuire agli studi giuridici sulle nuove vulnerabilità determinate dallo sviluppo tecnologico e in particolare dall'Intelligenza Artificiale (IA). Tale focus tematico consente di affrontare, al contempo, almeno due snodi essenziali: da un lato, quale disciplina dell'agricoltura possa garantire la sovranità alimentare nell'epoca delle nuove sfide ecologiche e tecnologiche; dall'altro, come la regolazione possa rendere l'IA sempre più alleata dei diritti fondamentali. Le due tematiche si intrecciano al crocevia della *data governance*, presentata – a partire dalla *European Data Strategy*[1] – come volano di uno sviluppo tecnologico ed economico che massimizza i benefici per il pluralismo e l'autodeterminazione degli attori in gioco.

Per rispondere alla domanda di ricerca presentata, si esamineranno le complessità giuridiche della tecnica sotto la lente della vulnerabilità. Quest'ultima nozione appare utile in quanto consente di osservare una visione pluridimensionale dei diritti, inclusiva delle loro intersezionalità e interdipendenze, nonché attenta alle necessità redistributive e all'autonomia delle soggettività interessate. In questo modo, l'*antropocentricity* dell'IA può essere interpretata con uno sguardo ecologico, che mette al centro le interazioni tra umani e non umani in un orizzonte di autodeterminazione individuale e collettiva. Nella specifica tematica del cibo, il concetto si presta altresì a dare centralità al lavoro di cura, che è altresì componente essenziale del lavoro agricolo in quanto direttamente implicato nel sostentamento dell'essere umano.

La tecnica si inserisce in questo quadro giacché – come è noto – essa può essere considerata un fattore di vulnerabilità[2]. L'IA, in particolare, minaccia la progressiva perdita di controllo delle persone sulle decisioni che vengono prese sul loro conto[3]. Infatti, tali sistemi nel tempo sono stati costruiti come *black box* – scatole opache che nascondono il proprio funzionamento interno, anche quando vengono utilizzate per prendere decisioni che risultano nella limitazione dei diritti fondamentali.

[1] Comunicazione della commissione al parlamento europeo, al consiglio, al comitato economico e sociale europeo e al comitato delle regioni Una strategia europea per i dati, COM(2020) 66 final, Bruxelles, 19.2.2020.

[2] Per questo motivo, lo sviluppo tecnico è espressamente menzionato come fattore possibilmente foriero di disuguaglianze: Dichiarazione Universale sulla bioetica e i diritti umani (2005), art. 8.

[3] Tali decisioni possono essere prese anche dal soggetto pubblico. Sul punto, cfr. A. CARDONE, *«Decisione algoritmica» vs decisione politica? A.I. legge democrazia*, Napoli, 2021; J.-B. AUBY, G. DE MINICO, G. ORSONI (a cura di), *L'amministrazione digitale. Quotidiana efficienza e intelligenza delle scelte*, Napoli, 2023.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Nel tempo la regolazione ha garantito l'intervento umano nelle procedure automatizzate (art. 22, GDPR)[4] e successivamente una disciplina generale della intelligenza artificiale – mediante il cd. *AI Act*[5] – mirante a garantire la tutela dell'essere umano rispetto a diverse categorie di rischio, con il divieto di alcune pratiche e la regolazione dell'uso in alcuni contesti ad alto rischio. In questi ultimi ambiti, sono stati previsti dunque obblighi di *risk assessment*[6], monitoraggio, trasparenza e controllo umano. Dall'attuale quadro normativo è chiara l'aspirazione a un'IA *constitutional by design*[7], progettata con l'obiettivo di mettere al centro la persona rispetto allo sviluppo tecnologico[8]. Tale regolazione è messa alla prova da un contesto in continua evoluzione, che richiede talvolta interventi settoriali volti a specificare e rafforzare l'implementazione dell'*AI Act*: ad esempio, gli attacchi alla cybersicurezza, la manipolazione dell'informazione e dei comportamenti elettorali o la sorveglianza generalizzata pubblica e privata[9], ivi inclusa la 'crisi del consenso', cioè la difficoltà di negare quest'ultimo quando è necessario per servizi essenziali.

Un aspetto centrale dell'elaborazione dottrinale è come la tecnica – allo stato attuale dei mercati tecnologici – tenda a consolidare poteri privati che arrivano talvolta a interferire con quelli pubblici e

[4] Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (regolamento generale sulla protezione dei dati), GU L 119, 4/5/2016. M.E. KAMINSKI, *Understanding Transparency in Algorithmic Accountability*, in W. BARFIELD (a cura di), *Cambridge Handbook of the Law of Algorithms*, Cambridge, 2020, 121-138; G. MALGIERI, G. COMANDÉ, *Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation*, in *International Data Privacy Law*, 7, 3, 2017, 5 ss.; T. TZIMAS, *Algorithmic Transparency and Explainability under EU Law*, in *European Public Law*, 4, 2023, 386 ss. In senso critico, cfr. S. WACHTER, B. MITTELSTADT, L. FLORIDI, *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*, in *International Data Privacy Law*, 2, 2017, 84.

[5] Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio, del 13 giugno 2024, che stabilisce regole armonizzate sull'intelligenza artificiale e modifica i regolamenti (CE) n, 300/2008, (UE) n, 167/2013, (UE) n, 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 (regolamento sull'intelligenza artificiale), PE/24/2024/REV/1, GU L 2024/1689, 12/7/2024.

[6] M.U. SCHERER, *Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies*, in *Harvard Journal of Law & Technology*, 29, 2, 2016, 393 ss.

[7] G. DE MINICO, *Too many rules or zero rules for the ChatGPT?*, in *BioLaw Journal – Rivista di BioDiritto*, 2, 2023, 492.

[8] P. NEMITZ, *Constitutional democracy and technology in the age of artificial intelligence*, in *Royal Society*, 15 October 2018; F. PASQUALE, *A Rule of Persons, Not Machines: The Limits of Legal Automation*, in *The George Washington Law Review*, 87, 1, 2019, 44 ss.; C. CASONATO, *Intelligenza artificiale e diritto costituzionale: prime considerazioni*, in *Diritto pubblico comparato ed europeo*, fasc. spec. 2019, 106 ss.; A. D'ALOIA, *Il diritto verso "il mondo nuovo". Le sfide dell'Intelligenza Artificiale*, in *BioLaw Journal – Rivista di BioDiritto*, 1, 2019, 9 ss.; T. GROPPI, *Alle frontiere dello Stato costituzionale: innovazione tecnologica e intelligenza artificiale*, in *Consulta online*, 3, 2020, 8 ss.; E. LONGO, *Rivoluzione digitale e sviluppi della partecipazione democratica nell'Unione europea*, in *Osservatorio sulle fonti*, 3, 2021, 1315 ss.; M. LUCIANI, *Libertà di ricerca e intelligenza artificiale. La sfida dell'intelligenza artificiale*, in *AssociazioneAIC.it*, 12, 2023; G. MOBILIO, *L'intelligenza artificiale e i rischi di una "disruption" della regolamentazione giuridica*, in *BioLaw Journal – Rivista di BioDiritto*, 2, 2020, 406 ss.; A. SIMONCINI, *Verso la regolamentazione della Intelligenza Artificiale. Dimensioni e governo*, in *BioLaw Journal – Rivista di BioDiritto*, 1, 2022, 76 ss.

[9] S. CALZOLAIO, *Vulnerabilità della società digitale e ordinamento costituzionale dei dati*, in *Rivista italiana di informatica e diritto*, 2, 2023, 14 ss.; E. LONGO, *La ricerca di un'antropologia costituzionale della società digitale*, in *Rivista italiana di informatica e diritto*, 2, 2023, 152-153.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

determinano una minaccia per l'autodeterminazione individuale e collettiva[10]. L'esistenza di tali dominanze è riconosciuta e affrontata nei più recenti atti normativi dell'Unione Europea, che con il *Digital Services Act*[11] e il *Digital Markets Act*[12] ha tentato di porre argini all'influenza di tali attori, rispettivamente sulla libertà di informazione e sulla competizione nei mercati digitali, con possibili effetti 'a cascata' su altri diritti fondamentali connessi al mezzo informatico.

Come ampiamente studiato in dottrina, la fonte di tali poteri risiede ampiamente nei dati che sono alla base dei processi di costruzione, 'allenamento' e applicazione dell'algoritmo[13]. Questi ultimi sono altresì centrali nella gestione dei rischi discriminatori insiti in tale tecnologia[14]. Se è vero che le discriminazioni algoritmiche sono aggravate dalla mancanza di trasparenza sull'algoritmo, che rende poco note le ragioni delle decisioni potenzialmente discriminatorie, è altresì possibile affermare che il *data mining* – alla base del funzionamento dell'AI – è intrinsecamente discriminatorio, in quanto si basa sulla clusterizzazione di dati per categorie e sulla decisione per ciascuna categoria[15]. Peraltro, tale osservazione dei dati utilizza tecniche probabilistiche che non guardano necessariamente al nesso causale – come richiederebbe il principio di uguaglianza – ma si basano su mere correlazioni, non necessariamente coerenti con la *ratio* della norma applicata. Infine, può rivelarsi problematico il modo in cui le aziende *high tech* ottengono i dati: la raccolta massiva di informazioni alimenta sistemi di dominanza che creano pressioni sulla persona o azienda interessata che la spingono cedere i propri dati anche a condizioni più svantaggiose di quelle che avrebbe preferito. Da ciò si deduce che le vulnerabilità determinate dall'IA non colpiscono tutte le soggettività allo stesso modo: i profitti dei dati vanno alle *big tech*, mentre le perdite si profilano soprattutto per persone e gruppi che erano già oggetto di sorveglianza e marginalizzazione. Per tali ragioni, un'IA è rispettosa dei diritti fondamentali soltanto se l'intervento regolatorio del soggetto pubblico prevede una *governance* dei dati coerente con l'autodeterminazione informativa, cioè con il diritto delle persone a riprendere il controllo sui propri dati, quale obiettivo ultimo della *privacy*[16]. Tale diritto deve essere garantito a livello tanto individuale quanto collettivo, in quanto proprio le tecniche di analisi in massa dei dati rendono le singole posizioni soggettive interdipendenti le une con le altre. L'attuale contesto mostra che il consenso informato alla cessione dei dati – pur espresso con le regole previste dal GDPR – è comunque suscettibile di rendere vulnerabile altri soggetti, in una società *online* dove i dati raccolti dalle piattaforme sono

[10] F. PARUZZO, *I sovrani della rete piattaforme digitali e limiti costituzionali al potere privato*, Napoli, 2022, 105 ss.

[11] Regolamento (UE) 2022/2065 del Parlamento europeo e del Consiglio del 19 ottobre 2022 relativo a un mercato unico dei servizi digitali e che modifica la direttiva 2000/31/CE, L 277/1, 27/10/2022, in particolare all'art. 33, che definisce le «Piattaforme online di dimensioni molto grandi e motori di ricerca online di dimensioni molto grandi».

[12] Regolamento (UE) 2022/1925 del Parlamento europeo e del Consiglio del 14 settembre 2022 relativo a mercati equi e contendibili nel settore digitale e che modifica le direttive (UE) 2019/1937 e (UE) 2020/1828 (regolamento sui mercati digitali), L 265/1, 12/10/2022, in particolare all'art. 3, che definisce i «gatekeeper».

[13] V. MAYER-SCHÖNBERGER, K. CUKIER, *Big Data*, Boston, 2013, trad. it. *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, Milano, 2013, 56.

[14] C. NARDOCCI, *Intelligenza artificiale e Discriminazioni*, in *Rivista del Gruppo di Pisa*, 3, 2021, 14 ss.

[15] B. GOODMAN, S. FLAXMAN, *EU regulations on algorithmic decision-making and a «right to* explanation, 28 giugno 2016, 27, in http://arxiv.org/pdf/1606.08813v1.pdf (ultima consultazione 02/12/2024).

[16] Sul punto, sia consentito rimandare – anche per le indicazioni bibliografiche – a: M.F. DE TULLIO, Big data *e* privacy *in una dimensione costituzionale collettiva*, in *Politica del Diritto*, 4, 2016, 637-696.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

continuamente connessi tra loro a opera delle piattaforme stesse. Peraltro, la recente normativa eurounitaria, salvo per alcune aperture[17], non ha messo in questione la reale fonte di potere delle piattaforme: la disponibilità di ampie masse di dati delle persone naviganti[18]. Tali lacune nell'attuale tutela dei dati fanno emergere la necessità di un intervento pubblico, capace di restituire alle persone naviganti *agency* sulle proprie scelte in materia di *privacy*, ma anche di prevenire la formazione di dominanza *data-based* capaci di influire sui diritti fondamentali.

Tale riflessione comporta l'individuazione del soggetto pubblico come primo attore obbligato a intervenire per combattere le nuove vulnerabilità in rete, sebbene per lungo tempo il digitale è stato considerato come l'*heaven of self-regulation*[19], con il conseguente dovere per il soggetto pubblico di astenersi da interferenze. Oggi tale tesi è stata pressoché universalmente superata[20], pur restando evidenti i rischi della tesi diametralmente opposta, quella di un'ingerenza pubblica totalizzante. Per evitare gli opposti esiti, il soggetto pubblico dovrebbe sicuramente astenersi da interferenze con i diritti fondamentali, ma al tempo stesso assumere una postura interventista nel ripristino dell'*agency* di chi utilizza le tecnologie.

Tali questioni – che riguardano l'IA in genere – toccano altresì l'agricoltura intelligente (*smart agriculture*), intesa come insieme di sistemi che utilizzano il trattamento automatizzato dei dati per supportare le attività agricole in diverse scelte circa l'uso di pesticidi e fertilizzanti, la semina, l'irrigazione o altre. Tali tecnologie possono incrociare informazioni relative a diversi fattori, tenuto conto anche della variabilità spaziale delle terre. Ad esempio, un trattore intelligente può raccogliere dati sul proprio percorso e salvarli in *cloud* per suggerire indicazioni sulla semina, adeguare la propria guida automatica per ottenere la quantità giusta di sostanze o ridurre gli errori umani. Oppure, i sensori possono identificare la presenza di tossine nelle colture, in modo più efficace e meno costoso di altri metodi, o ancora contribuire al controllo sull'igiene alimentare realizzando una tracciabilità dei prodotti lungo tutta la filiera[21].

Le potenzialità della *smart agriculture* per i diritti sono evidenziabili nella risposta a tre sfide epocali che coinvolgono il settore agricolo: l'innovazione digitale; l'incremento della produttività per sfamare

[17] Si vedano alcuni obblighi di condivisione previsti nel Digital Markets Act, art. 6, co. 11.

[18] I dati sono stati considerati *essential facility* da una parte della dottrina: M. OREFICE, Big data. *Regole e concorrenza*, in *Politica del Diritto*, fasc. 4/2016. pp. 727 ss. Anche diverse decisioni delle Autorità competenti sulla concorrenza hanno definito alcuni illeciti anticoncorrenziali sulla base dei dati: AUTORITÉ DE LA CONCURRENCE, *Décision n° 13-D-20 du 17 décembre 2013*, § 438; EAD., *Avis n° 11-A-02 du 20 janvier 2011*, § 183. Cfr. AUTORITE DE LA CONCURRENCE, BUNDESKARTELLAMT, *Competition Law and Data*, 10 maggio 2016, in https://www.autoritedelaconcurrence.fr/fr (ultima consultazione 02/12/2024), 17-18 e 31-32.

[19] J.P. BARLOW, *A Declaration of the Independence of Cyberspace*, in *EFF.org*, 8 febbraio 1996, in https://www.eff.org/cyberspace-independence (ultima consultazione 02/12/2024).

[20] Per una riflessione, ad esempio, sull'*Internet Governance*, si veda: G. DE MINICO, *Towards an Internet Bill of Rights*, in *Loyola Los Angeles International and Comparative Law Review*, 2015, 5-11, in *http://digitalcommons.lmu.edu/ilr/vol37/iss1/1* (ultima consultazione 02/12/2024); L.B. SOLUM, *Models of Internet Governance*, in A.L. BYGRAVE, J. BING (a cura di), *Internet Governance. Infrastructure and Institutions*, Oxford, 2009, 55; E. BROUSSEAU, M. MARZOUKI, C. MÉADEL, *Governance, Networks and Digital Technologies: Societal, Political and Organizational Innovations*, in ID. (a cura di), *Governance, Regulation and Powers on the Internet*, Cambridge, 2012, 3.

[21] P.J. ZARCO-TEJADA, N. HUBBARD, P. LOUDJANI, *Precision agriculture: An opportunity for EU farmers - Potential support with the CAP 2014-2020*, Directorate-General for Internal Policies – Policy Department B: Structural and Cohesion Policies Agriculture and Rural Development Study, 2014, 40-41.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

la popolazione globale in aumento; la tutela dell'ambiente mediante l'ottimizzazione delle risorse impiegate. Il tema ha un chiaro aspetto ridistributivo, se è vero che sia i cambiamenti climatici sia l'insicurezza alimentare – specie nelle loro forme più estreme – riguardano in modo precipuo i Paesi del Sud globale, anche a causa dei processi di colonizzazione e sfruttamento su cui si fondano le attuali ricchezze del Nord. Anche nello stesso Nord globale, la crisi del 2008 ha messo in questione l'assunto che le preoccupazioni degli Stati dovessero riferirsi solo alla *food safety*, o igiene alimentare, considerando risolto il problema della *food security*[22].

D'altro canto, gli studi sull'IA mostrano che tale tecnologia può anche accrescere le vulnerabilità esistenti o generarne di nuove, specie quando il suo impiego comporta dei *bias* discriminatori e delle opacità che impattano negativamente sui diritti fondamentali e sulla libera concorrenza. Questa seconda, in particolare, è una preoccupazione che riguarda immediatamente le libertà economiche, ma che nello specifico ambito agricolo – come in altri, quali l'informazione – richiede una disciplina che va oltre quella concorrenziale, in quanto tocca altresì i diritti fondamentali. Infatti, il diritto al cibo nelle sue declinazioni contemporanee è innanzitutto un diritto alla sovranità alimentare, dunque a una diversità di colture e modalità produttive. Di conseguenza – come si osserverà più approfonditamente in prosieguo – le dominanze, proprio come nel caso dell'informazione, può essere considerata come un 'male in sé', anche quando è ottenuta ed esercitata con mezzi che il diritto della concorrenza considera leciti.

Da queste ambivalenze dell'agricoltura intelligente muove il percorso argomentativo di questa ricerca. In primo luogo, si ricostruirà il diritto al cibo, utilizzando la nozione di vulnerabilità come chiave di lettura per sviscerarne le diverse sfumature teoriche. Alla luce di tale definizione si identificheranno le principali questioni regolative aperte dell'agricoltura intelligente, concentrandosi in modo particolare sulla *governance* dei dati agricoli, che appare un fattore determinante nell'equilibrio tra i diritti coinvolti nei mercati della *smart agriculture*. Infine, si affronterà l'argomento de iure condendo, identificando alcune direttive regolative capaci di amplificare il godimento della sovranità alimentare attraverso la tecnica.

2. Vulnerabilità e sovranità alimentare

Come riferito in Introduzione, la nozione di vulnerabilità è centrale per interpretare la sovranità alimentare e le possibili influenze dell'IA sul relativo diritto. Le ragioni sono da rinvenirsi segnatamente in quattro caratteristiche della nozione, che arricchiscono il contenuto dell'uguaglianza: la dialettica tra il riconoscimento della vulnerabilità universale, nell'attuale 'società del rischio', e quella di specifiche categorie; l'intersezionalità tra i diritti fondamentali; le responsabilità dello Stato e degli attori

[22] A. IANNARELLI, *Il mercato agro-alimentare europeo*, in L. SCAFFARDI, V. ZENO-ZENCOVICH (a cura di), *Cibo e diritto. Una prospettiva comparata*, Roma, 2020, 264. Nella definizione del *World Food Summit Plan of Action* del 1996, la *food security* è identificata come «condizione che si verifica quando tutte le persone, in qualsiasi momento, hanno un accesso fisico, sociale ed economico ad un cibo sufficiente, sicuro e nutrizionalmente adeguato, in grado di rispondere ai loro bisogni e preferenze alimentari e di garantire una vita attiva e in salute». Sull'evoluzione del concetto, cfr.: FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS (FAO), *Policy brief on food security*, 2, 2016, 1, in https://www.fao.org/fileadmin/templates/faoitaly/documents/pdf/pdf_Food_Security_Cocept_Note.pdf (ultima consultazione 02/12/2024).

Downloaded from www.biodiritto.org.
ISSN 2284-4503

privati di fronte al rischio di 'cattura' da parte dei poteri privati; gli obblighi di garanzia in capo ai medesimi attori privati rispetto ai rischi e ai danni generati dalle loro stesse posizioni di potere. Di seguito si illustreranno tali elementi, senza pretesa di esaustività, ma solo in quanto necessari a illuminare la nozione di sovranità alimentare presa in considerazione ai fini di questa ricerca.

Quanto alla prima caratteristica, essa è evidente nella definizione stessa del concetto – polisemico e variegato – di vulnerabilità. Quest'ultima richiama innanzitutto un aspetto ontologico comune a tutte le persone[23] – la precarietà e la finitezza della vita – che rilegge i diritti fondamentali alla luce del lavoro di cura, corporea e non, necessario quotidianamente per costruire e mantenere una sfera personale in un contesto di interdipendenza[24]. Nella *Dichiarazione Universale sulla bioetica e i diritti umani* (2005), tale concetto è esplicitamente considerato come limite da considerare nell'avanzamento scientifico e tecnologico[25]. La medesima nozione, però, può essere letta anche in chiave egalitaria[26], con attenzione alle diverse fonti della vulnerabilità – talvolta di natura patologica – e alle posizioni di particolare svantaggio di alcune categorie colpite[27].

Di qui la seconda caratteristica della vulnerabilità – l'approccio intersezionale – che richiama come la vulnerabilità 'ontologica' sopra individuata è di fatto storicamente e socialmente condizionata. Ad esempio, la salute personale dipende non solo da fattori naturali e genetici, ma anche da elementi economici, sociali e ambientali, che influenzano la possibilità di prevenire e curare le patologie[28]. Oppure, l'idea che esistano territori 'sfruttabili' per le materie prime può portare a uno sviluppo industriale incontrollato che alimenta l'esposizione a problemi di salute attraverso la povertà e i condizionamenti ambientali.

La terza e quarta caratteristica della vulnerabilità, qui identificate, sono legate alla necessaria dimensione prescrittiva di tale circostanza[29], la cui presenza attiva obblighi miranti a ripristinare lo *status*

---

[23] J. BUTLER, *Precarious Life: The Powers of Mourning and Violence*, Londra – New York, 2004.

[24] M.A. FINEMAN, *The Vulnerable Subject: Anchoring Equality in the Human Condition*, in *Yale Journal of Law & Feminism*, 20, 1, 2008, 10-11, in https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1131407 (ultima consultazione 02/12/2024).

[25] Art. 8.

[26] Così è evidenziato nelle definizioni normative della vulnerabilità. In tal senso va la *Dichiarazione di Barcellona*, 1998, redatta dopo un percorso di tre anni promosso dalla Commissione Europea e dal Centre for Ethics and Law di Copenaghen. Tale concetto consente di rifocalizzare altresì la nozione liberale di democrazia: A. MACINTYRE, *Dependent Rational Animals: Why Human Beings Need the Virtues*, Chicago, 1999, 5. Infatti – pur senza negare la responsabilità individuale – l'idea di una precarietà costitutiva dell'essere umano esclude che gli svantaggi individuali siano riconducibili unicamente alla potenziale incapacità della singola persona di cogliere le opportunità esistenti: L. CORSO, *Vulnerabilità, giudizio di costituzionalità e sentimentalismo*, in *Ars interpretandi*, 2, 2018, 60.

[27] A partire da queste diverse occorrenze del termine, nei suoi diversi significati, la dottrina ha elaborato alcune tassonomie, quali: C. MACKENZIE, W. ROGERS, S. DODDS, *Vulnerability. New Essays in Ethics and Feminist Phylosophy*, Oxford, 2014, 8; E. DICIOTTI, *La vulnerabilità nelle sentenze della Corte europea dei diritti dell'uomo*, in *Ars interpretandi*, 2, 2018, 14 ss.; R. CHENAL, *La definizione della nozione di vulnerabilità e la tutela dei diritti fondamentali*, in *Ars interpretandi*, 2, 2018, 39-43.

[28] C. MACKENZIE, W. ROGERS, S. DODDS, *op. cit.*, p. 8.

[29] «Alla variegatezza di tali contorni descrittivi della nozione corrisponde una certa quale univocità circa la funzionalità precettiva della nozione, utilizzata infatti per imporre sugli Stati l'obbligo di una tutela 'particolare', capace di tener conto delle loro specificità»: F. IPPOLITO, *La vulnerabilità quale principio emergente nel diritto internazionale dei diritti umani?*, in *Ars interpretandi*, 2, 2019, 64. Tale costruzione è stata utilizzata altresì per

Downloaded from www.biodiritto.org.
ISSN 2284-4503

formale[30] delle soggettività interessate, ma anche a creare le condizioni affinché queste ultime possano esercitare in concreto un potere decisionale su se stesse[31]. Se così non fosse, il rimedio diventerebbe esso stesso causa di vulnerabilità[32], in una visione paternalistica che alimenta l'impotenza e dipendenza di persone e gruppi coinvolti, mediata dalla loro cristallizzazione stereotipante nel ruolo di 'vittima'[33].

Quanto ai soggetti passivi, obbligati da tale responsabilità, essi possono essere pubblici o privati. Quanto ai primi, il tema si collega a una dimensione della vulnerabilità definita come 'precarietà istituzionale' (*institutional precariousness*)[34]: quando è il diritto stesso a essere incerto e debole nella tutela, esso può diventare agente patologico di vulnerabilità[35]. Quanto al secondo, esso si basa sull'assunto che i soggetti privati sono ugualmente responsabili di non commettere abusi[36], sicché il dovere di protezione spetta innanzitutto al soggetto perpetratore[37], specie quando quest'ultimo si trova in una posizione di dominio[38].

contestare il modello contrattualista e volontarista dell'obbligazione, con l'affermazione per cui la vulnerabilità sarebbe la fonte primaria dell'obbligo: C. MACKENZIE, W. ROGERS, S. DODDS, *op. cit.*, 10.

[30] L'autonomia, infatti, è uno status formale, oltre che una capacità, e il primo aspetto tende a influenzare il secondo: C. MACKENZIE, *The Importance of Relational Autonomy and Capabilities for an Ethics of Vulnerability*, in C. MACKENZIE, W. ROGERS, S. DODDS, *op. cit.*, 44. «Ciò è possibile in quanto il mancato riconoscimento produce un'offesa allo status che ha luogo nei rapporti sociali, invece che nella psicologia individuale; questo significa vedersi negare lo status di «partecipante» a pieno titolo all'interazione sociale dove il soggetto è squalificato in quanto non meritevole di rispetto e di stima»: S. ZULLO, *Lo spazio sociale della vulnerabilità tra pretese di giustizia e pretese di diritto. Alcune considerazioni critiche*, in *Politica del diritto*, 3, 2016, 495.

[31] F. IPPOLITO, *op. cit.*, 72.

[32] C. MACKENZIE, W. ROGERS, S. DODDS, *op. cit.*, 9.

[33] Come affermato dalla Corte EDU nel caso di disabilità: «such groups were historically subject to prejudice with lasting consequences, resulting in their social exclusion. Such prejudice may entail legislative stereotyping which prohibits the individualised evaluation of their capacities and needs»: Corte Europea dei Diritti dell'Uomo, *Alojas Kiss v. Hungary*, Application No. 38832/06, 20/5/2010, in *https://www.refworld.org/jurisprudence/caselaw/echr/2010/en/73000* (ultima consultazione 02/12/2024). Inoltre, è stato osservato che gli organismi delle Nazioni Unite – proprio per evitare lo stigma su determinati gruppi – ha iniziato a riferire l'aggettivo 'vulnerabili' alle situazioni, più che alle persone: F. IPPOLITO, *op. cit.*, 70.

[34] B.S. TURNER, *Vulnerability and Human Rights*, University Park (Pennsylvania), 32. Con accenti diversi, si veda: M.A. FINEMAN, *op. cit.*, 5-7.

[35] A. ABIGNENTE, *Quando il diritto vulnera*, in G. BLANDO, G. CONZA (a cura di), *La vulnerabilità alla prova dell'argomentazione giuridica*, Napoli, 2020, 13; G. BLANDO, *Brevi note metodologiche sul rapporto tra Diritto, argomentazione e vulnerabilità*, in G. BLANDO, G. CONZA (a cura di), *La vulnerabilità alla prova dell'argomentazione giuridica*, Napoli, 2020, 26 e 30; F. CIARAMELLI, *La vulnerabilità: da caratteristica dei soggetti a carattere del diritto*, in O. GIOLO, B. PASTORE (a cura di), *Vulnerabilità. Analisi multidisciplinare di un concetto*, Roma, 2018, 179.

[36] Nell'Unione Europea, ad esempio, è ormai riconosciuta l'efficacia orizzontale dei diritti: Corte di giustizia dell'Unione europea, *Walrave*, C-36/74, 12 dicembre 1974, §§ 16-19. Questo vale per le norme chiare, precise e incondizionate, ma la Corte di Giustizia ha dato un'interpretazione ampia di tale nozione: G. STROZZI, R. MASTROIANNI, *Diritto dell'Unione Europea. Parte istituzionale* (VI edizione), Torino, 2013, 208-209; G. TESAURO, *Diritto comunitario* (V edizione), Padova, 2008, 94.

[37] C. MACKENZIE, W. ROGERS, S. DODDS, *op. cit.*, 9.

[38] R.E. GOODIN, *Protecting the Vulnerable: A Reanalysis of Our Social Responsibilities*, Chicago, 1985, 194.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Le caratteristiche presentate pongono l'elaborazione sulla vulnerabilità come un arricchimento dei diritti fondamentali e dell'uguaglianza[39]. In primo luogo, mettere al centro la condizione di rischio o danno significa poter tenere conto dell'intersezionalità tra fattori differenziati di disuguaglianza, invece che riservare la tutela a gruppi rigidamente individuati[40]. In secondo luogo, la tutela viene chiaramente identificata come restituzione alle persone e ai gruppi sociali di un'*agency* effettiva sulla propria posizione[41].

La nozione qui discussa – in base alle caratteristiche presentate – rispecchia almeno tre nodi essenziali del diritto al cibo: la valorizzazione della produzione alimentare come fondamentale lavoro di cura, rivolto alle necessità vitali e corporali di tutte le persone (vulnerabilità esistenziali), ma anche connesso a situazioni di particolare insicurezza o povertà alimentare (vulnerabilità situazionale); la protezione attraverso la salvaguardia dell'autonomia individuale e collettiva, nella forma della sovranità alimentare; la possibile 'cattura' del regolatore, che indebolisce il diritto nei suoi doveri di tutela.

Per quanto riguarda il primo nodo, il cibo è riconosciuto nella Dichiarazione universale dei diritti Umani[42] e nel Patto internazionale sui diritti economici sociali e culturali[43] (1966) come diritto che riguarda tutte le persone, ma che contempla situazioni di particolare svantaggio. Il medesimo assetto valoriale è osservabile nella ricostruzione del diritto al cibo nell'ordinamento italiano[44]. Qui sono centrali gli artt. 2 e 32 Cost., sotto il profilo del diritto alla vita, ma anche della tutela ambientale (art. 9 Cost.): le pratiche e le sostanze impiegate nell'agricoltura possono danneggiare i suoli e l'inquinamento

[39] «La vulnerabilità, a differenza di altre nozioni, come quelle dei diritti, tra i quali il diritto all'autodeterminazione tutelato dall'art. 8 CEDU, o della proporzionalità, non ha una sua autonomia concettuale e definitoria, ma svolge solo una funzione servente alla tutela effettiva dei diritti fondamentali. Quando fa ricorso alla nozione di vulnerabilità la Corte constata che il soggetto in questione si trova in una situazione particolare che lo legittima a richiedere un esame individualizzato della sua posizione e una protezione rafforzata dei suoi diritti proprio in virtù e alla luce della sua particolarità. La nozione di vulnerabilità deve essere quindi letta in relazione al suo essere funzionale e strumentale a garantire la massima effettività dei diritti fondamentali e alla giustificazione dell'esercizio o dell'assenza dell'esercizio della forza pubblica, ossia delle ingerenze o delle omissioni delle autorità»: R. CHENAL, *op. cit.*, p. 51. Cfr. M.G. BERNARDINI, *Vulnerabilità e disabilita a Strasburgo: il «vulnerable groups approach in pratica»*, in *Ars interpretandi*, 2, 2018, 82.

[40] F. IPPOLITO, *op. cit.*, 69-70; T. CASADEI, Diritti umani in contesto: forme della vulnerabilità e "diritto diseguale", in *Ragion pratica*, 2, 2008, 295-296. Ad esempio, è stato notato, gli effetti concreti della stessa condizione di disoccupazione – indubitabilmente una situazione di svantaggio – sono diversi se ci si trova in un Paese del Nord Europa o dell'Africa: S. ZULLO, *op. cit.*, 489-490.

[41] L'*agency* è definita come «autonoma capacità di azione, intesa tuttavia come azione compiuta non in una condizione di astratta libertà, ma piuttosto nel contesto di vincoli sociali e culturali che strutturano l'identità soggettiva quali il genere, la razza e la classe»: M.G. GIAMMARINARO, *L'influenza trasformativa delle prospettive femministe. Vulnerabilità e agency*, in *Rivista di filosofia del diritto*, 2, 2022, 341.

[42] Dichiarazione Universale dei Diritti Umani, 10 dicembre 1948, art. 25, par. 1.

[43] Patto internazionale di New York relativo ai diritti economici, sociali e culturali, 16 dicembre 1966, art. 11, par. 1.

[44] In tale ordinamento, il diritto in questione non è esplicitato, come in altri (ad es., l'art. 7 della Costituzione brasiliana). Cfr. L. CHIEFFI, *Scelte alimentari e diritti della persona: tra autodeterminazione del consumatore e sicurezza sulla qualità del cibo*, in *Diritto Pubblico Europeo Rassegna online*, 1, 2015, 235-236. Tuttavia, esso può essere ricostruito attraverso un insieme di valori costituzionali: M. BOTTIGLIERI, *The protection of the Right to adequate food in the Italian Constitution*, in *ForumCostituzionale.it*, 23 novembre 2015, 2-5.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

può ridurre nel tempo la quantità e qualità delle derrate alimentari[45]. Come è noto, tali valori sottendono la salvaguardia dei diritti fondamentali, ma anche esigenze redistributive (art. 3, primo e secondo comma) intra generazionali e intergenerazionali, a tutela delle generazioni future.

Rispetto a tale scenario, le riflessioni svolte sulla vulnerabilità aiutano a cogliere alcune sfumature e contraddizioni che caratterizzano la tutela del cibo e dell'agricoltura[46], rilevanti anche per il discorso sulla *smart agriculture*. In particolare, la nozione costituisce una bussola per conciliare tra loro due fondamentali imperativi, che non sempre coesistono in concreto: la disponibilità di cibo accessibile e la qualità del cibo.

Possibili contraddizioni tra detti valori possono verificarsi allorché l'agricoltura intensiva e industriale mantiene alta la produttività, e bassi i costi al dettaglio, basandosi su esternalità negative quali: l'intensa meccanizzazione, la massiccia immissione di immissioni di elementi di sintesi (fertilizzanti, pesticidi, erbicidi, etc.) negli ecosistemi[47], la standardizzazione dei processi produttivi, la rigida selezione delle razze animali e varietà vegetali[48], lo sfruttamento del lavoro lungo tutta la filiera (inclusa la logistica) e il caporalato[49]. Particolarmente rilevante in questa sede è come questi modelli economici tendano alla concentrazione del potere di mercato[50] che si ripercuotano anche sulla proprietà terriera attraverso il cd. *land grabbing*: la concorrenza dell'agroindustria rende vulnerabile la piccola proprietà terriera (art. 44 Cost.) all'acquisizione da parte delle aziende più forti, capaci di praticare condizioni competitive grazie alle loro economie di scala. Questa tendenza può avere a sua volta ripercussioni sull'ambiente e la biodiversità delle colture e della nutrizione, in quanto proprio l'art. 44 cost. concepisce la piccola proprietà e il pluralismo economico come presidio dell'ambiente e dell'equità nei rapporti sociali, come avviene, ad esempio, quando le terre sono destinate a forme di agricoltura contadina o di economia sociale e solidale[51].

Questi rilievi mostrano ancora una volta le connessioni tra vulnerabilità esistenziali e situazionali: se è vero che la *food safety* sembra essere eminentemente espressione della vulnerabilità di tutte le persone al cibo insalubre o contaminato, è altresì vero che i modelli agricoli dominanti rendono l'accesso

[45] G. CORDINI, *Il diritto al cibo, le generazioni future e il mercato*, in *Diritto pubblico comparato ed europeo*, fascicolo speciale, 2019, 148 ss.

[46] A. LIGUSTRO, *Diritto al cibo e sovranità alimentare nella prospettiva dell'Organizzazione Mondiale del Commercio*, in *Diritto pubblico comparato ed europeo*, fasc. speciale, 2019, 396-397.

[47] P. DE MEO, F. PARASCANDOLO, *Si scrive cibo (agro-ecologico e territorializzato), si legge democrazia (di luogo)*, in *Scienze del territorio*, 8, 2020, 47.

[48] C. DEL CONT, *Non solo cibo, not just food: which compatibility between consumers' ethical and social preoccupations and trade and commercial law?*, intervento nel Florence Sustainability of Well-Being International Forum 2015: *Food for Sustainability and not just food*, in *Agriculture and Agricultural Science Procedia*, 8, 2016, 274.

[49] P. MCMICHAEL, *Food Regimes and Agrarian Questions*, Black Point – Winnipeg, 2013, 69 ss. Peraltro, si è osservato che anche le etichette geografiche vanno soprattutto a beneficio dei distributori, mentre l'impatto sui produttori e le produttrici della filiera può essere scarso o nullo: M. LO CASCIO, *Un prodotto Dop in terra di mafia. Le olive da tavola Nocellara in Sicilia*, in *Meridiana. Rivista quadrimestrale dell'Istituto meridionale di storia e scienze sociali*, 93, 2018, 8-14.

[50] P.H. HOWARD, *Concentration and Power in the Food System. Who Controls What We Eat?*, London – New York, 2016, 1-6.

[51] In alcuni casi queste reti spontanee vengono sostenute da apposite *food policy* a livello locale: P. DE MEO, F. PARASCANDOLO, *op. cit.*, 50 ss.; E. BUTELLI, *Pianificazione ambientale autosostenibile e alimentazione: il Piano del cibo della Provincia di Pisa*, in *Scienze del territorio*, 3, 2015, 128 ss.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

al cibo sano economicamente condizionato, in quanto realizzano l'accessibilità economica a discapito della salute delle persone consumatrici ed ecosistemi. A tal fine, giova ricordare che le vulnerabilità legate al cibo non consistono solo nell'insicurezza alimentare, ma anche nella cd. 'povertà alimentare', intesa come «l'incapacità di permettersi, o di avere accesso a un cibo che permetta una dieta salutare»[52], per effetto di fattori come la scarsità di risorse economiche, la mancanza di disponibilità locale di alimenti e la carenza dei trasporti o la mancanza di conoscenze, competenze e strumenti per cucinare[53].

Tali contraddizioni vengono appianate se si adotta un'ottica intersezionale che, come si è visto, è tipica della vulnerabilità. Qui rileva il secondo nodo evidenziato in apertura di questo paragrafo, e cioè le responsabilità pubbliche e private che nascono dalle vulnerabilità, consistenti nel dovere di ripristinare l'autodeterminazione delle persone colpite dalle disuguaglianze. Come chiarito dal Comitato sui diritti economici, sociali e culturali[54], gli Stati hanno il dovere di «migliorare i metodi di produzione, di conservazione e di distribuzione delle derrate alimentari […] in modo da conseguire l'accrescimento e l'utilizzazione più efficaci delle risorse naturali» nonché di «assicurare un'equa distribuzione delle risorse alimentari mondiali in relazione ai bisogni». Dunque, accessibilità e salubrità del cibo, che sempre più spesso sono messe in competizione nei fatti, devono giuridicamente coesistere.

La lotta alla povertà o insicurezza alimentare non è dunque soddisfatta attraverso la mera provvista di cibo alle persone indigenti. Le relative garanzie, infatti, devono tenere conto della 'cultural or consumer acceptability'[55] (*General Comment No. 12*), collegata altresì alle libertà religiose (artt. 19-20 Cost.)[56] e ai diritti culturali[57] (artt. 9, 33 e 34 Cost.), intesi anche quali diritti sociali[58]. Questa saldatura tra diritto al cibo e autodeterminazione individuale e collettiva è stata definita nel dibattito politico e

[52] La definizione è in UK DEPARTMENT OF HEALTH, *Choosing a better diet: a food and health action plan*, 2005, 7, in https://dera.ioe.ac.uk/7558/7/dh_4105709_Redacted.pdf (ultima consultazione 02/12/2024); ACTION AID, *La pandemia che affama l'italia. Covid-19, povertà alimentare e diritto al cibo*, Report 2020, 7 ss., in *https://actionaid.imgix.net/uploads/2020/10/AA_Report_Poverta_Alimentare_2020.pdf* (ultima consultazione 02/12/2024).

[53] *Ivi*.

[54] COMMITTEE ON ECONOMIC, SOCIAL AND CULTURAL RIGHTS, *General Comment No. 12: The Right to Adequate Food (Art. 11)*, 12 maggio 1999, E/C.12/1999/5, parr. 7-13. D'ora in poi, *General Comment No. 12*.

[55] Secondo il *General Comment No. 12*, l'accettabilità impone «di tenere in considerazione, per quanto possibile, i valori percepiti – non basati sugli aspetti nutritivi – che sono legati al cibo e al suo consumo, nonché le preoccupazioni informate del consumatore rispetto alla natura delle provviste di cibo accessibili».

[56] È ben noto che diversi culti includono prescrizioni legate al regime alimentare. In tal senso, la tutela del cibo è stata connessa altresì all'art. 9 CEDU e all'art. 10 della Carta dei Diritti Fondamentali dell'UE: R. D'ORAZIO, *La libertà di coscienza e il principio di eguaglianza alla prova delle «dottrine alimentari»*, in L. SCAFFARDI, V. ZENO-ZENCOVICH (a cura di), *Cibo e diritto. Una prospettiva comparata*, Roma, 2020, 45 ss.

[57] Tale riferimento va inteso anche in un'ottica evolutiva, «alla luce delle nuove scoperte scientifiche riguardanti la produzione di cibo (come nel caso della biotecnologia o dei cibi clonati), o anche in base alla rivalutazione delle cose commestibili (come per gli insetti)»: S. LANNI, *Not Just a Bug: Brief Remarks of Legal Anthropology for New Food Choices*, in L. SCAFFARDI, V. ZENO-ZENCOVICH (a cura di), *Cibo e diritto. Una prospettiva comparata*, Roma, 2020, 74. Per altri aspetti, la crescente sensibilità ai temi ambientali sta facendo emergere, anche nel discorso giuridico, il tema dell'opzione alimentare come scelta politica. In tal senso, il caso forse più noto è quello del vegetarianesimo e veganesimo, adottati come scelta di vita per contrarietà alle forme dell'allevamento intensivo: L. CHIEFFI, *op. cit.*, 247 ss.

[58] C. PICIOCCHI, *Le scelte alimentari come manifestazioni d'identità, nel rapporto con gli ordinamenti giuridici: una riflessione in prospettiva comparata*, in L. SCAFFARDI, V. ZENO-ZENCOVICH (a cura di), *Cibo e diritto. Una prospettiva comparata*, Roma, 2020, 126.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

giuridico come 'sovranità alimentare', ossia «il diritto dei popoli a un cibo sano e culturalmente appropriato prodotto con metodi sostenibili e il loro diritto a definire i propri sistemi alimentari e agricoli»[59]. Come accennato sopra, la sovranità alimentare si pone in tensione con i modelli dominanti di produzione e distribuzione, in quanto «sviluppa un modello di produzione sostenibile su piccola scala a beneficio delle comunità e del loro ambiente. La sovranità alimentare dà la priorità alla produzione e al consumo di cibo locale, dando a un paese il diritto di proteggere i suoi produttori locali dalle importazioni a basso costo e di controllare la sua produzione. Include la lotta per la terra e una vera riforma agraria che assicuri che i diritti di usare e gestire terre, territori, acqua, semi, bestiame e biodiversità siano nelle mani di coloro che producono cibo e non del settore corporativo»[60].

Di conseguenza, la produzione alimentare deve salvaguardare le colture e i saperi locali, nonché l'ambiente nelle zone interne e rurali, ma anche nelle aree periferiche e periurbane, dove l'agricoltura contadina sottrae suolo alla cementificazione e lo mette a disposizione per la cura della comunità attraverso la produzione di cibo genuino. Dunque, il cibo diviene anche il luogo dei diritti legati alla sussidiarietà orizzontale (art. 118, comma 4, Cost.) e alla partecipazione politica (art. 49 Cost.). Sono diversi gli esempi rilevanti: gli orti sociali, quali esperimenti pedagogici, di rigenerazione urbana ed economia solidale; i beni confiscati, dove le medesime attività sono altresì un presidio di lotta alle mafie e alla speculazione; gli spazi sociali e cooperativi che ospitano, ad esempio, dispense, empori e gruppi di acquisto solidale. Come si vedrà, tale aspetto è rilevante per il presente studio sull'IA, in quanto non dà per scontato che l'aumento della produttività sia coerente con la tutela dell'ambiente e implica la necessità che la tecnica – pur lavorando per conseguire standard di efficienza e produttività – non si ponga in chiave omologante, ma rispetti e metta a frutto i saperi agroecologici maturati nei diversi territori.

Infine, in relazione al terzo nodo – circa il ruolo 'vulnerante' o 'protettivo' del diritto – diversi elementi della normazione eurounitaria sul cibo sono sintomatici della rinuncia a una regia pubblicistica capace di tutelare il lavoro agricolo rispetto al mercato[61]: la progressiva erosione delle sovvenzioni accessibili ai piccoli produttori e produttrici[62]; l'assenza di una regolamentazione minima per i 'contratti di filiera'

[59] La Via Campesina – International Peasants' Movement, *The International Peasants' Voice Globalising hope, globalising the struggle!*, in https://viacampesina.org/en/international-peasants-voice/ (ultima consultazione 02/12/2024). Cfr. A. Rinella, H. Okoronko, *Sovranità alimentare e diritto al cibo*, in *Diritto pubblico comparato ed europeo*, 1, 2015, 92 ss.

[60] *Ivi*.

[61] I. Canfora, *Raggiungere un equilibrio nella filiera agroalimentare. Strumenti di governo del mercato e regole contrattuali*, in L. Scaffardi, V. Zeno-Zencovich (a cura di), *Cibo e diritto. Una prospettiva comparata*, Roma, 2020, 237 ss.

[62] Si vedano, ad esempio, le critiche che hanno riguardato la scorsa proposta di riforma della PAC. Tale politica è stata criticata in quanto incapace di sostenere la piccola agricoltura ecologica, e destinata soprattutto alle grandi industrie agricole e agli allevamenti intensivi: Cfr. L. Gaita, *Politica agricola comune, tutte le occasioni mancate del megapiano UE: sussidi senza limiti ai colossi, pochi incentivi a chi è ecosostenibile, fondi in crescita agli allevamenti intensivi*, in *IlFattoQuotidiano.it*, 12 settembre 2021, in *https://www.ilfattoquotidiano.it/2021/09/12/politica-agricola-comune-tutte-le-occasioni-mancate-del-megapiano-ue-sussidi-senza-limiti-ai-colossi-pochi-incentivi-a-chi-e-ecosostenibile-fondi-in-crescita-agli-allevamenti-intensivi/6317454/* (ultima consultazione 02/12/2024). Nel complesso, la riforma della PAC è stata vissuta come un'occasione persa per introdurre delle condizionalità sociali alla sovvenzione: Associazione Rurale Italiana, *Come applicare la Condizionalità Sociale nella Politica Agricola Comune (PAC)*, in *AssoRurale.it*, 13 dicembre 2021, in

Downloaded from www.biodiritto.org.
ISSN 2284-4503

– e in particolare quello tra il produttore e il primo acquirente – con la totale rimessione dei debiti al libero mercato, con i rapporti di forza iniqui sopra descritti[63]; la accordi commerciali e trattati di libero scambio[64], negoziati con ampio coinvolgimento di diversi attori privati[65]. Ciò ha consentito un ampio ingresso di un diritto di matrice privata, definito come *private food law*[66].

Tra i vari temi, sembra pertinente in questa sede citare il ruolo degli standard – spesso di origine privata – che hanno il compito di mediare produttività e qualità del cibo. Infatti, si può immaginare che una traiettoria regolativa simile potrebbe verificarsi rispetto agli standard tecnologici che, come si vedrà, sono uno dei futuri terreni regolativi della *smart agriculture*.

Ebbene, circa la qualità del cibo, la definizione degli standard è stata ampiamente lasciata alle aziende private: tali canoni sono stati sviluppati dalle imprese per autotutelarsi nelle pratiche commerciali e sono successivamente nella regolazione[67]. In particolare, ciò è avvenuto attraverso almeno due canali. Il primo è, appunto, il contratto: se un operatore di mercato richiede regolarmente gli standard ai propri fornitori, gli stessi standard diventano simili a modelli contrattuali o condizioni generali di contratto. Astrattamente, ogni strumento contrattuale è retto dal principio di relatività, e quindi ha efficacia solo tra le parti; tuttavia, è stato osservato che modelli e condizioni generali possono avere degli effetti molto simili a quelli di un atto normativo quando una parte ha un potere negoziale tale da poter imporre le proprie condizioni a un numero molto ampio di controparti e utenti finali, che non riescono a trovare altrove un mercato dove vendere con migliori condizioni[68]. Naturalmente, se i grandi compratori – la grande distribuzione – impongono commercialmente l'uso di uno standard, tutta la catena a

https://www.assorurale.it/2021/05/13/come-applicare-la-condizionalita-sociale-nella-politica-agricola-comune-pac/ (ultima consultazione 02/12/2024).

[63] A tal fine, è risultata insufficiente la scelta di favorire accordi collettivi tra produttori, in deroga alle normative sulla concorrenza: C. DEL CONT, *op. cit.*, 273-274; I. CANFORA, *op. cit.*, 238.

[64] Si veda, anche per un resoconto del percorso regolativo della WTO: A. LIGUSTRO, *op. cit.*, 398 ss.

[65] A. RINELLA, H. OKORONKO, *op. cit.*, 97 ss.

[66] Si deve l'espressione a B. VAN DER MEULEN (a cura di), *Private food law. Governing food chains through contract law, self-regulation, private standards, audits and certification schemes*, Wageningen 2018. Il curatore preferisce questo termine a quello più diffuso di 'autoregolazione' in quanto quest'ultimo trasmette erroneamente l'idea di una completa coincidenza tra i soggetti che producono della regola e quelli tenuti a osservarla, mentre tale coincidenza – visti i rapporti di forza presenti – non sempre si dà (*Ibidem*, p. 31). Si è altresì parlato, in tal senso, di 'ibridizzazione' della legge sul cibo: P. VERBRUGGEN, T. HAVINGA, *Hybridization of food governance: An analytical framework*, in P. VERBRUGGEN, T. HAVINGA (a cura di), *Hybridization of Food Governance. Trends, Types and Results*, Cheltenham – Northampton, 2017, 1-4.

[67] Gli standard regolano la manifattura di un prodotto per rendere chiare e coerenti le aspettative circa la qualità dello stesso lungo tutta la filiera, anche quando alcuni elementi della filiera non sono tenuti a conoscere le buone regole della produzione. In tal modo, in tali snodi non è necessario avere tutte le conoscenze tecniche necessarie, bensì è sufficiente richiedere il rispetto dello standard. Lo standard consente, inoltre, di esternalizzare le procedure di controllo, affidandole a specifici organismi che svolgono le verifiche e rilasciano un certificato. Normalmente, a creazione di uno standard richiede aziende tecnicamente specializzate. Queste ultime talvolta rilasciano lo standard in forma aperta, mentre in altri casi lo rendono disponibile solo a pagamento, apponendo misure di protezione della proprietà intellettuale. Un esempio di questo secondo caso è la certificazione ISO: *https://www.iso.org/home.html* (ultima consultazione 02/12/2024). La successiva illustrazione si deve a B. VAN DER MEULEN, *3. The Anatomy of Private Food Law*, in ID. (a cura di), *Private food law. Governing food chains through contract law, self-regulation, private standards, audits and certification schemes*, Wageningen 2018, 77 ss.

[68] Sul tema, cfr. G. DE MINICO, *Regole. Comando e consenso*, Torino, 2005, 133.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

monte si deve adeguare e omologare allo standard stesso, anche se non è tradotto in una normativa pubblica. A ciò si aggiunga che in diversi casi le stesse autorità pubbliche fanno propri gli standard privati, imponendoli normativamente[69].

Tutte le questioni evidenziate pongono la questione di come la postura del settore pubblico possa essere la tutela attiva della sovranità alimentare, come fondamentale diritto di autodeterminazione. In questa sede, si cercherà di contribuire (anche) a tale dibattito osservando i nuovi rischi determinati dalle tecnologie IA, unitamente alle possibili vie d'uscita.

3. La *climate smart agriculture* e i suoi nodi regolativi

I rilievi sopra svolti si sono resi necessari per inquadrare gli interessi in gioco e i loro possibili contrasti, al fine di identificare le condizioni regolative per il loro migliore contemperamento. Tale trattazione si è resa necessaria per evidenziare le possibili contraddizioni tra i valori in gioco, laddove le tecnologie di *smart agriculture* si presentano come uno strumento capace di assicurare al contempo la sicurezza alimentare e l'ottimizzazione delle risorse. Con tale intento, diversi documenti di *policy* si riferiscono alla *climate smart agriculture*[70], che sarebbe in grado di aumentare la quantità di cibo immesso sul mercato – in risposta alla crescita di popolazione sul pianeta – contribuendo al tempo stesso alla resilienza rispetto al cambiamento climatico e all'ottimizzazione delle risorse a vantaggio della tutela ambientale[71]. Ad avviso di chi scrive, tale condizione di *win-win* non può essere data per scontata[72], quasi che la tecnica possa essere una nuova 'mano invisibile' capace di realizzare automaticamente il benessere collettivo[73]. Viceversa, il contemperamento degli obiettivi può realizzarsi soltanto se la tecnologia è regolata in modo da limitare le esternalità negative della produzione; in caso contrario, è altresì possibile che la tecnica si traduca in un aggravamento delle condizioni ambientali e un esaurimento delle risorse. Inoltre, il contemperamento degli interessi deve tenere in considerazione che l'aumento di

[69] Ciò è accaduto, ad esempio, in tema di agricoltura biologica, ma può succedere anche in via residuale, quando la disciplina pubblicistica rinvia alle buone pratiche esistenti in ambito commerciale per regolare tutti gli aspetti non direttamente indicati dalla norma. Ad esempio, il D.M. 18 luglio 2008 del Ministero delle politiche agricole alimentari, forestali e del Turismo si riferisce alla possibilità di utilizzare, nelle more dell'adozione della disciplina nazionale, «norme private […] conformi alle procedure ed ai parametri minimi individuati nell'Allegato 1 del presente decreto».

[70] FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS, *Climate-smart agriculture sourcebook*, Roma, 2013, IX, in *https://www.fao.org/docrep/018/i3325e/i3325e.pdf* (ultima consultazione 02/12/2024). Alcuni esempi sono in: FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS, *"Climate-Smart" agriculture: policies, practices and financing for food security, adaptation and mitigation*, Roma, 2010, 1 ss., in https://www.fao.org/4/i1881e/i1881e00.htm (ultima consultazione 02/12/2024). Anche la World Bank ha raccomandato questo tipo di politiche: WORLD BANK GROUP, *World Bank Group Climate Change Action Plan 2021–2025: Supporting Green, Resilient, and Inclusive Development*, Washington, 2021, 25, in http://hdl.handle.net/10986/35799 (ultima consultazione 02/12/2024).

[71] M. TAYLOR, *Climate-smart agriculture: what is it good for?*, in *The Journal of Peasant Studies*, 45, 1, 2018, 89 ss.

[72] M. GARDEZI, R. STOCK, *Growing algorithmic governmentality: Interrogating the social construction of trust in precision agriculture*, in *Geoforum*, 122, 2021, 1.

[73] L'espressione è stata traslata dal linguaggio economico a quello giuridico in: G. DE MINICO, *Internet. Regola e anarchia*, Napoli, 2012, 12 e 96-206.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

produttività non esaurisce la necessaria risposta all'insicurezza o povertà alimentare, in quanto l'obiettivo da realizzare è quello più ampio della sovranità alimentare.

In breve, la CSA deve preoccuparsi non solo di far progredire la tecnica, ma anche di riequilibrare il quadro di *governance* e le relazioni di potere esistenti nella produzione e distribuzione alimentare[74]. A tal fine, è necessario che l'UE affronti regolativamente alcuni snodi problematici[75], tra i quali si evidenzieranno almeno tre questioni irrisolte: il *digital divide*; il controllo sui dati e il possibile *lock in* delle imprese agricole rispetto alle *agricultural technology provider* (ATP); la standardizzazione dei processi agricoli sul modello dell'agricoltura intensiva, connessa altresì all'accaparramento delle terre per la destinazione alla medesima agroindustria. Tali questioni riguardano in parte meri interessi economici di produttori e produttrici, specie delle piccole realtà, ma con riflessi anche sui diritti fondamentali legati alla preservazione del pluralismo economico. Inoltre, si pone l'esigenza di una *just transition*: l'obsolescenza delle competenze potrebbe incidere sulla sussistenza di lavoratori e lavoratrici o piccole imprese che si trovano già in condizione di precarietà e vulnerabilità[76].

Il *digital divide* è un problema ben noto nello sviluppo tecnologico in genere: le nuove disuguaglianze riguardano risorse come l'accesso a Internet, i dispositivi, le competenze digitali, le informazioni e altre. Nel settore agricolo, il semplice accesso alla rete veloce è un tema significativo per le realtà situate in aree interne e remote: queste ultime potrebbero non avere una connettività tale da consentire, ad esempio, il caricamento dei dati in *cloud*, spesso impiegato dalle tecnologie di *smart agriculture* per l'analisi delle informazioni[77]. Un secondo *gap* digitale riguarda le nuove conoscenze e competenze, la cui acquisizione rappresenterà un onere proporzionalmente maggiore per le imprese di ridotte dimensioni e per le categorie svantaggiate di lavoratori e lavoratrici[78]. In particolare, serviranno nuove *skill* per utilizzare i nuovi macchinari, ma anche – e a maggior ragione – per autodeterminarsi sia sulla scelta

[74] O. De Schutter – Un Special Rapporteur on the Right to Food, *Commentary VI: Agroecology: A solution to the Crises of Food Systems and Climate Change*, in United Nations Conference on Trade and Development, *Trade and environment review 2013*, Ginevra, 2013, 37-38, in https://unctad.org/publication/trade-and-environment-review-2013 (ultima consultazione 02/12/2024).

[75] In particolare, uno studio commissionato dal Servizio Studi del Parlamento europeo evidenzia sfide quali: la proprietà (ownership) dei dati forniti da chi produce e la redistribuzione dei loro benefici; il rischio che l'agricoltura europea diventi dipendente dalla produzione extraeuropea di tecnologie e macchinari per la PA; la necessità di colmare il digital divide, anche in termini di competenze; la possibile perdita di posti di lavoro; l'aumento, di conseguenza, del divario tra piccole e grandi imprese agricole, in quanto le prime potrebbero non avere il capitale o lo conoscenze necessarie per acquisire le tecnologie; le nuove tecnologie potrebbero sostituire il rapporto diretto tra l'uomo e la natura, che sarebbe affidato a un tracciamento elettronico dei prodotti lungo la filiera: J. De Baerdemaeker, *Artificial intelligence in the agri-food sector. Applications, risks and impacts*, Study commissioned by the European Parliamentary Research Service – Scientific Foresight Unit (STOA) – Panel for the Future of Science and Technology, 2023, 27-28, in https://www.europarl.europa.eu/stoa/en/document/EPRS_STU(2023)734711 (ultima consultazione 02/12/2024).

[76] Sul tema si veda, ad esempio: S. García-Magariño, U. Belintxon, *Cognitive and Energetic Sustainability for Development: Spain and Europe before the Green Deal*, in *Energies,* 14, 2021, 1; M. Mazzucato, G. Dibb, M. Mcpherson, *Il green deal non può aspettare*, in M. Mazzucato, *Non sprechiamo questa crisi*, Roma – Bari, 2020, 66-67.

[77] E. Jakku, B. Taylor, A. Fleming, C. Mason, S. Fielke, C. Sounness, P. Thorburn, *"If they don't tell us what they do with it, why would we trust them?" Trust, transparency and benefit-sharing in Smart Farming*, in *Journal of Life Sciences*, 90-91, 1, 2019, 6.

[78] «It will change the skills, over time, required to be a successful farmer»: dalla citazione di un'intervista a una persona coltivatrice in E. Jakku, B. Taylor, A. Fleming, C. Mason, S. Fielke, C. Sounness, P. Thorburn, *op. cit.*, 6.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

di adottare o meno una tecnologia, sia per esercitare un controllo umano sugli esiti suggeriti dall'IA. Infatti, non può darsi autodeterminazione e se non c'è possibilità di comprendere i criteri e gli esiti dei processi automatizzati e guardare in modo critico agli stessi.

Un ulteriore *divide* digitale riguarda il controllo dei dati, che è il tema su cui ci si concentrerà più ampiamente in questa sede, in quanto presenta un insieme di problematiche giuridiche tuttora irrisolte, con ampie ripercussioni anche sulle altre due questioni qui sollevate. Come è noto, l'IA si fonda ampiamente sulla mole di informazioni, più che sulla capacità di trattarle. Nel caso dell'agricoltura intelligente, l'impresa agricola deve trasmettere all'ATP un ampio numero di dati – riguardanti diversi aspetti della terra, del suolo, delle colture e dei processi produttivi – per ricevere il servizio. Tale modello di *business*, osservabile anche in altri settori *data-driven*, tende naturalmente a creare concentrazioni sul mercato, a causa degli effetti di rete: le aziende dominanti hanno più dati, quindi forniscono un servizio migliore e così crescono ancora[79]. Il potere di mercato aumenta ulteriormente quando grandi ATP e *data broker* intraprendono intese e concentrazioni finalizzate ad accrescere il proprio patrimonio di dati[80]. Questa dominanza ha effetti sul mercato delle ATP, ma anche, a cascata, sulle realtà agricole che producono i dati: queste ultime si vedono costrette a cedere tali informazioni per avere un servizio, senza poter né negoziare alla pari le relative condizioni, né trovare ATP *competitor* che assicurino condizioni più vantaggiose[81].

Gli effetti di tali concentrazioni possono apprezzarsi rispetto ad almeno due ordini di diritti, tra loro connessi: da un lato, le libertà economiche di produttori e produttrici, con particolare riferimento alle piccole aziende; dall'altro lato, la sovranità alimentare stessa, che si sostiene anche grazie alla varietà di piccole produzioni, di colture e di tecniche agricole. Entrambi i profili saranno analizzati di seguito, partendo dalle libertà economiche, fino ad arrivare ai diritti fondamentali coinvolti.

Dal punto di vista dell'impresa agricola, accettare di cedere i dati significa perdere il controllo su informazioni economicamente rilevanti, acquisite peraltro grazie al lavoro delle stesse realtà produttrici, che catturano informazioni guidando i trattori o svolgendo le proprie mansioni quotidiane[82]. Ciò pone un problema di sfiducia, ma anche di disparità di potere: se è vero che le imprese agricole restano formalmente titolari dei dati, tale titolarità è svuotata dal controllo delle ATP, che ne traggono profitto[83].

---

[79] European Commission, *Bayer v. Monsanto*, Case No. COMP/M.8084, §2470.

[80] Ad esempio, nel 2013 il colosso agricolo Monsanto/Bayer ha acquisito The Climate Corporation, una compagnia di *data science*, tramite cui ha accordi con produttori di macchine intelligenti come John Deere, Agco e CNHI e offre servizi di piattaforma alle aziende agricole per mettere insieme i dati da diverse fonti – cedendo anche i propri dati – e supportare le decisioni. Cfr. C. ATIK, *Towards Comprehensive European Agricultural Data Governance: Moving Beyond the "Data Ownership" Debate*, in *IIC - International Review of Intellectual Property and Competition Law*, 53, 2022, 706.

[81] C. ATIK, *op. cit.*, 705.

[82] J.L. FERRIS, *Data Privacy and Protection in the Agriculture Industry: Is Federal Regulation Necessary*?, in *Minnesota Journal of Law, Science and Technology*, 18, 1, 2017, 317.

[83] P.J. ZARCO-TEJADA, N. HUBBARD, P. LOUDJANI, *op. cit.*, 40. A maggior ragione in quanto dal momento che normalmente le aziende tech sono attori più grandi rispetto alle imprese agricole, specie alle piccole: esponente di un governo locale citato in E. JAKKU, B. TAYLOR, A. FLEMING, C. MASON, S. FIELKE, C. SOUNNESS, P. THORBURN, *op. cit.*, 7.

Downloaded from www.biodiritto.org.

ISSN 2284-4503

Tali squilibri possono creare situazioni di *lock in*, quando l'ATP crea barriere tecnologiche[84] o contrattuali rispetto al trasferimento dei dati presso altre ATP. Infatti, in questo caso la prospettiva di dover 'allenare' una nuova macchina, quindi affrontare nuovamente le minori efficienze che si verificano nella fase iniziale, potrebbe essere un deterrente[85]. La portabilità dei dati stessi è dunque cruciale affinché l'utente possa fruire di una reale libertà di scelta e, di conseguenza, perché funzioni la libera concorrenza tra le ATP. A tal fine, è necessario che tale pretesa sia esplicitamente tutelata dal diritto, anche con la previsione di standard e protocolli uniformi che facilitino la comunicazione tra diversi sistemi[86].

Le medesime dominanze possono portare ad altre forme di sfruttamento: in alcuni casi, ad esempio, le ATP impediscono all'utente di riparare le proprie macchine, dichiaratamente con ragioni di cybersecurity. Di conseguenza, si rende necessario comprare nuove macchine o rivolgersi ai servizi accreditati dalla stessa azienda per le riparazioni. Tali impedimenti rappresentano degli oneri ulteriori per le realtà agricole; non a caso, alcuni studi riferiscono circa la nascita una vera e propria comunità 'hacker' dell'agri tech, composta da chi è costretto/a ad hackerare le proprie stesse macchine per ripararle[87].

Infine, l'affidamento delle informazioni agli ATP pone un tema di sicurezza e riservatezza dei dati. Ad esempio, sarebbero considerevoli i danni di un cyberattacco o un problema tecnico che comportasse la perdita dei dati[88]. Secondo un sondaggio del 2016 della American Farm Bureau Federation, il 70% delle persone produttrici hanno manifestato preoccupazione che le aziende e il settore pubblico possano accedere ingiustamente ai dati per i propri scopi[89]. Ciò può avvenire illecitamente, a causa di *leak* volontari o involontari di informazioni, oppure in attuazione degli stessi accordi contrattuali, che le imprese agricole sono costrette ad accettare per fruire del servizio.

Tutte queste circostanze hanno indotto parte della dottrina a discorrere di 'spossessamento'[90] (*dispossession*) o di *data grabbing*, in analogia con il *land grabbing*: di fatto, se chi produce il dato è l'impresa agricola e la *big tech* si appropria della gran parte del profitto, allora la prima è messa a lavoro per il profitto della seconda[91], come in una nuova forma di latifondo digitale[92]. Anche se la fornitura di dati è equamente remunerata al singolo produttore, ciò non permette una redistribuzione del profitto lungo la catena di produzione del valore, in quanto il vero guadagno deriva dall'incrocio dei dati e dunque è maggiore rispetto alla somma del valore dei singoli *data set*[93].

---

[84] C. Atik, *op. cit.*, 705.
[85] E. Jakku, B. Taylor, A. Fleming, C. Mason, S. Fielke, C. Sounness, P. Thorburn, *op. cit.*, 7.
[86] P.J. Zarco-Tejada, N. Hubbard, P. Loudjani, *op. cit.*, 32-33.
[87] S. Rotz, E. Duncan, M. Small, J. Botschner, R. Dara, I. Mosby, M. Reed, E.D.G. Fraser, *The Politics of Digital Agricultural Technologies: A Preliminary Review*, in *Sociologia Ruralis*, 59, 2, 2019, 115.
[88] *Ibidem*, p. 116.
[89] American Farm Bureau Federation, *Farmers Want to Control Their Own Data*, Farm Bureau Survey, 2016, in *https://eu.farmforum.net/story/news/agriculture/2016/05/12/farm-bureau-survey-farmers-want-to-control-their-own-data/49219409/* (ultima consultazione 02/12/2024); M. Gardezi, R. Stock, *op. cit.*, 2.
[90] J. Thatcher, D. O'Sullivan, D. Mahmoudi, *Data colonialism through accumulation by dispossession: New metaphors for daily data*, in *Environment and Planning D: Society and Space*, 0, 0, 2016, 6.
[91] M. Gardezi, R. Stock, *op. cit.*, 2.
[92] A. Fraser, *Land grab/data grab*, cit., 885.
[93] *Ibidem*, 886.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Su vasta scala, tali impatti sulle libertà economiche possono cambiare il volto del settore agricolo, attentando alle piccole produzioni e, di conseguenza, alla stessa sovranità alimentare. Le tecnologie – più che semplici strumenti agricoli – divengono vere e proprie mediatrici di un ordine sociale, cambiando i rapporti di potere nella produzione di conoscenza e nella determinazione delle modalità produttive[94].

Di seguito si potranno osservare molteplici fattori di vulnerabilità, suscettibili di rendere ulteriormente precarie le piccole realtà produttive, con effetti su diversi diritti della personalità.

Da un punto di vista individuale, il trattamento di dati su vasta scala può compromettere il diritto alla *privacy*[95], allorché i dati sulle terre possono rivelare anche aspetti personali quali, ad esempio, il reddito, la localizzazione della terra e dell'abitazione[96].

Su scala più ampia, è forte l'interrogativo su come l'IA possa convivere con i principi agroecologici. Infatti, il *data mining* è fondato su procedimenti statistici e quindi, salvo specifici correttivi, si basa sulla standardizzazione delle decisioni e dei criteri decisionali: uno dei vantaggi dichiarati dell'agricoltura intelligente è che essa genera benefici ambientali su vasta scala, al di là delle intenzioni e della cultura della singola impresa agricola[97]. Questo funzionamento comporta un rischio di omologazione quando le imprese agricole si rivolgono in massa a un ridotto numero di ATP, adeguandosi all'esito dell'IA: come è noto, la parvenza di esattezza del calcolo matematico pur non essendo vincolante rappresenta un *nudge* per l'utente a seguirne i suggerimenti[98]. Ciò avviene a maggior ragione quando l'ATP non è trasparente sull'algoritmo e sui dati impiegati, rendendo meno comprensibili gli effetti dell'automazione sulle scelte valoriali e imprenditoriali sottese alla produzione.

Questa circostanza non sarebbe un problema se l'aumento della produttività potesse essere considerata un calcolo politicamente neutro; eppure, il discorso svolto nel paragrafo precedente mostra l'inattendibilità di questo assunto[99]. In primo luogo, non si può sempre assumere che l'ottimizzazione delle risorse in senso economico vada di pari passo con i valori ecologici: possono essere diversi, ad esempio, gli orizzonti temporali: l'uno immediato, l'altro anche di lungo periodo e a beneficio delle generazioni future[100]. Peraltro, anche da un punto di vista economico un'impresa può avere diversi obiettivi di produttività a seconda della sua taglia: ad esempio, mentre la piccola impresa – basata sul lavoro familiare e su una piccola proprietà terriera – tende a massimizzare il prodotto per ettaro, dati i limiti

[94] M. Gardezi, R. Stock, *op. cit.*, 2.

[95] E. Jakku, B. Taylor, A. Fleming, C. Mason, S. Fielke, C. Sounness, P. Thorburn, *op. cit.*, 6.

[96] J.L. Ferris, *op. cit.*, 316.

[97] P.J. Zarco-Tejada, N. Hubbard, P. Loudjani, *op. cit.*, 35.

[98] R. Stock, M. Gardezi, *Make bloom and let wither: Biopolitics of precision agriculture at the dawn of surveillance capitalism*, in *Geoforum*, 122, 2021, 199.

[99] «Investments that increase food production will not make significant progress in combating hunger and malnutrition if they do not lead to higher incomes and improved livelihoods for the poorest – particularly small-scale farmers in developing countries. And short-term gains will be offset by long-term losses if they cause further degradation of ecosystem, thus threatening the ability to maintain current levels of production in the future»: O. De Schutter, *op. cit.*, 34.

[100] «Efforts to improve the productivity of a given crop by finding more intensive ways to produce it through simplifying production and increasing its scale, for example, may have negative implications at a landscape level through unintended impacts such as biodiversity loss, interruption of nutrient or water cycling, degradation or contamination of neighbouring fields, and the foreclosing of common property resources»: M. Taylor, *op. cit.*, 97.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

della propria terra, la grande impresa massimizza la produzione per ora di lavoro, per accrescere l'output del capitale investito[101]. Proprio per tenere conto di questi fattori, tradizionalmente l'agricoltura contadina funziona attraverso la valorizzazione di conoscenze profonde – normalmente locali e tramandate nelle generazioni – sullo specifico ecosistema agricolo[102]. Peraltro, come è noto, è ancora irrisolto il tema di come l'IA stessa possa adeguarsi a criteri ecologici, considerate le risorse – specialmente energetiche – necessarie al suo funzionamento[103].

Se l'impostazione dell'algoritmo viene lasciata alle dinamiche di mercato, i grandi ATP adatteranno probabilmente le tecnologie ai loro compratori più forti, cioè i colossi dell'agroindustria, con enormi rischi per la preservazione di altre forme agricole. In breve, gli ATP possono avere effetti simili a quelli della grande distribuzione[104], che con la propria capacità di acquisto sono in grado di condizionare le modalità produttive, imponendo standard e quantitativi di produzione. In questo caso si tratta di *provider* e non di compratori, ma analogo è il potere di mercato, qualora pochi ATP diventino in grado di fornire tecnologie necessarie a soddisfare i requisiti di produzione imposti dalla grande distribuzione.

In parallelo alla standardizzazione dei processi produttivi, il controllo sui dati – specie in presenza di alleanze tra ATP e colossi agroindustriali – può altresì favorire l'accaparramento delle terre. Infatti, «il *land grabbing* ha bisogno di dati»[105]: conoscere le terre aiuta l'investitore a scegliere quelle più remunerative e oggi il trattamento dei dati consente di avere informazioni molto dettagliate sul punto, incrociando dati circa l'altitudine, il flusso dei fiumi, la variabilità dei suoli, la potenziale esistenza di depositi minerari e altre circostanze utili. Così, il *data grabbing* può ripercuotersi anche su un controllo materiale sulle terre, dunque sul pluralismo delle aziende produttrici e, in ultima analisi, sulla sovranità alimentare[106].

---

[101] *Ibidem*, 96.

[102] O. DE SCHUTTER, *op. cit.*, 36-37. Cfr. L. KARLSSON, L.O. NAESS, A. NIGHTINGALE, J. THOMPSON, *'Triple wins' or 'triple faults'? Analysing the equity implications of policy discourses on climate-smart agriculture (CSA)* (Version 1), University of Sussex, 2017, 11 e 14-15, in https://hdl.handle.net/10779/uos.23456360.v1 (ultima consultazione 02/12/2024); S. ROTZ, E. DUNCAN, M. SMALL, J. BOTSCHNER, R. DARA, J. MOSBY, M. REED, E.D.G. FRASER, *op. cit.*, 113. «Productivity, therefore, is not a neutral or self-evident concept but one that is constitutive of value judgements about the purposes of agriculture and the broader socio-ecological functions it serves»: M. TAYLOR, *op. cit.*, 97.

[103] Tale tema non sembra aver trovato una risoluzione definitiva ed è all'ordine del giorno nell'Unione Europea: *Communication from the Commission the European Green Deal*, COM(2019) 640 final, 11/12/2019; INDEPENDENT HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE SET UP BY THE EUROPEAN COMMISSION, *Ethics Guidelines for Trustworthy AI*, 2019, in https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (ultima consultazione 02/12/2024); *Council conclusions "Digitalisation for the Benefit of the Environment"*, 13957/20, 17/12/2020, in https://data.consilium.europa.eu/doc/document/ST-13957-2020-INIT/en/pdf (ultima consultazione 02/12/2024); EUROPEAN COMMISSION, *Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions Fostering a European approach to Artificial Intelligence*, (COM) 2021/205 final, 21/4/2021, Annex - *Coordinated Plan on Artificial Intelligence*, Bruxelles.

[104] Di fatto, la promozione della *Climate Smart Agriculture* ha altresì l'effetto di attrarre piccoli produttori e piccole produttrici nel mercato della grande distribuzione: L. KARLSSON, L.O. NAESS, A. NIGHTINGALE, J. THOMPSON, *op. cit.*, 12.

[105] A. FRASER, *Land grab/data grab: precision agriculture and its new horizons*, in *The Journal of Peasant Studies*, 46, 5, 2019, 905.

[106] *Ibidem*, 894.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

4. Dalla *food sovereignty* alla *data sovereignty*

Tutti i fattori fin qui presentati rendono evidente la necessità di un intervento regolativo sui divari digitali, con particolare riferimento alla *governance* dei dati. Si rende necessario, in particolare, un doppio ordine di salvaguardie, che riguardino la libertà imprenditoriale, ma guardino soprattutto agli impatti di ampia portata sulla *privacy* e sulle scelte di produzione e consumo.

Per cogliere tale necessità di autodeterminazione, la dottrina ha coniato la nozione di *data sovereignty*, come complemento della *food sovereignty*, con l'obiettivo che «actors in civil society, or in cooperative economic associations, develop principles and practices that explore whether the emergent value of data should be held in common, rather than privatized; destroyed, rather than analyzed and brought to market; or stored nearby, rather than exported»[107]. Il significato normativo di tale *data sovereignty* non è lontano dall'autodeterminazione informativa di cui si parlava in apertura (§1) e necessita di specifiche misure normative, solo parzialmente realizzate.

Una via per affrontare il problema potrebbe quella difensiva: scoraggiare produttori e produttrici rispetto alla condivisione dei dati, preservando così i propri *asset* e, laddove pertinente, la propria *privacy*. Tale soluzione, tuttavia, non sembra ottimale, in quanto vi sono diversi fattori di interesse generale che militano verso la condivisione stessa, con particolare riferimento al miglioramento delle tecnologie agricole, alla tracciabilità alimentare o ancora al monitoraggio ambientale. Rileva qui altresì l'interesse generale della comunità, che è coinvolta non solo in quanto si nutre dei cibi prodotti, ma anche in quanto è interna ai processi agroecologici, dove il legame tra chi produce e chi acquista è configurabile nel senso di una solidarietà e comunità di intenti, piuttosto che in un rapporto di consumo. Corrispondentemente, la *data sovereignty* richiede una condivisione di dati da parte delle ATP, non solo con le realtà produttive e tra le medesime realtà, ma anche con la generalità delle persone.

Per queste ragioni si ritiene, in questa sede, che la condivisione dei dati da parte delle imprese agricole non debba essere impedita o scoraggiata, bensì inquadrata in una cornice normativa capace di contemperare i diritti fondamentali individuali e collettivi in gioco. Ugualmente, e a maggior ragione, gli stessi motivi mirano a favore della condivisione dei grandi patrimoni di dati posseduti dalle ATP e dai *data broker*. In entrambi i casi, le soluzioni normative che si cercheranno di esplorare saranno basate sulla promozione di procedimenti negoziati per la gestione dei dati, ma anche sulla necessità di norme imperative, laddove gli interessi economici degli attori in gioco si pongano in conflitto con la salvaguardia dell'ambiente o di altri diritti fondamentali, individuali e collettivi, sovraordinati.

Si potrebbe obiettare che costringere le ATP ad aprire i propri forzieri di dati potrebbe ostacolare l'innovazione tecnologica: come è noto, le *enclosures* della proprietà immateriale servono a stimolare il progresso, in base all'assunto che le imprese non si dedicherebbero alla ricerca e allo sviluppo se non avessero la prospettiva di un profitto. Analogamente, le realtà agricole potrebbero decidere di non usare le tecnologie, dal momento che rischiano di perdere il controllo di un proprio *asset*[108]. L'obiezione non tiene conto, però, di un opposto rischio, che si verifica quando l'eccesso di *enclosure* finisce per parcellizzare la proprietà. In questo caso si parla di *tragedy of anti-commons*: la parcellizzazione della proprietà aumenta i costi di transazione anche quando la condivisione dei beni – in questo caso

[107] *Ibidem*, 907.
[108] C. ATIK, *op. cit.*, 725.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

dei dati – sarebbe utile all'avanzamento della scienza e della stessa economia. Peraltro, nel caso dei dati, si tratta di beni non rivali, i quali possono essere utilizzati al contempo da diversi attori economici. Proprio in virtù di tali circostanze, la *European Data Strategy* spinge verso la condivisione dei dati tra le imprese[109], con la formazione di *European Agricultural Data Spaces* (EADS). La motivazione alla base di questa *policy* è eminentemente economica: evitare – quanto meno all'interno del mercato unico eurounitario – la *data fragmentation* che oggi compromette l'innovatività e competitività dell'economia. Nella stessa direzione vanno le principali fonti normativi che concorrono a promuovere il *data sharing*, in particolare il *Data Act* (DA)[110] e al *Data Governance Act* (DGA)[111]. Del resto, da tempo alcune teorie economiche sottolineano che esistono degli strumenti alternativi di remunerazione dello sforzo di ricerca – specie nella forma di incentivi pubblici – e oggi l'UE è più che mai motivata a considerarsi 'partner' di investimenti virtuosi, a partire dai piani post-pandemici e dalla revisione del patto di stabilità, che hanno mitigato la più rigorosa versione dell'austerità adottata nel decennio scorso.

Sembra dunque di poter ribadire la tesi per cui il *data sharing* sia – a determinate condizioni – la miglior tutela della *data sovereignty*. In questo senso, la discussione si intreccerà con il dibattito sull'efficacia dell'attuale regolazione, ma non sarà del tutto sovrapponibile. Infatti, la strategia europea sui dati ha il mercato unico come suo principale obiettivo e, se in parte la tutela delle piccole aziende rappresenta una preoccupazione condivisa tra il diritto della concorrenza e i diritti umani, non vi è completa identità tra i due interessi. Di conseguenza, in questa sede si terrà conto di uno scarto e di una gerarchia tra le diverse posizioni tutelate: la tutela delle libertà economiche può e deve essere realizzata in modo da promuovere i diritti fondamentali coinvolti.

---

[109] Tale intento è confermato nei successivi regolamenti attuativi della strategia. Come recita il consid. 32 del *Data Act* (vd. nota 104): «L'obiettivo del presente regolamento non è solo promuovere lo sviluppo di prodotti connessi o di servizi correlati nuovi e innovativi e stimolare l'innovazione nei servizi post-vendita, ma anche stimolare lo sviluppo di servizi completamente nuovi che utilizzano i dati in questione, anche sulla base di dati provenienti da una varietà di prodotti connessi o servizi correlati. Allo stesso tempo, il presente regolamento mira a evitare di compromettere gli incentivi agli investimenti per il tipo di prodotto connesso da cui i dati sono ottenuti, ad esempio mediante l'uso di dati per sviluppare un prodotto connesso concorrente considerato intercambiabile o sostituibile dagli utenti, in particolare in base alle caratteristiche del prodotto connesso, al suo prezzo e all'uso previsto».

[110] Regolamento (UE) 2023/2854 del Parlamento europeo e del Consiglio, del 13 dicembre 2023, riguardante norme armonizzate sull'accesso equo ai dati e sul loro utilizzo e che modifica il regolamento (UE) 2017/2394 e la direttiva (UE) 2020/1828 (regolamento sui dati), PE/49/2023/REV/1, GU L, 2023/2854, 22/12/2023, in http://data.europa.eu/eli/reg/2023/2854/oj (ultima consultazione 02/12/2024). «In other words, the creation of a mixed public-private regulatory space will offer the infrastructural and regulatory framework within which data, including privately held datasets, will be voluntarily, or mandatorily when the required, exchanged for economic and societal benefit. This will effectively become what has been termed the European single market for data. By doing so, the EC aims to foster data sharing and re-use, which is expected to deliver growth and innovation, to support policy making and to preserve European values such as privacy, property, competition, consumer protection, pluralisms, safety, security, fairness, ethical standards and digital sovereignty»: C. DUCUING, T. MARGONI, *Introduction*, in C. DUCUING, T. MARGONI, L. SCHIRRU (a cura di), *White Paper on the Data Act Proposal*, CiTiP Working Paper, 2022, 11, in *https://ssrn.com/abstract=4259428* (ultima consultazione 02/12/2024)

[111] Regolamento (UE) 2022/868 del Parlamento europeo e del Consiglio del 30 maggio 2022 relativo alla governance europea dei dati e che modifica il regolamento (UE) 2018/1724 (Regolamento sulla governance dei dati), GU L 152/1, 3/6/2022, in https://eur-lex.europa.eu/legal-content/IT/TXT/HTML/?uri=CELEX:32022R0868#d1e2093-1-1 (ultima consultazione 02/12/2024).

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Per limitare i timori difensivi di chi produce i dati, diverse *policy* settoriali promettono di mantenere 'proprietà' (*ownership*) dei dati in capo a chi li produce[112]. Ammesso e non concesso che tale approccio sia il più efficace[113], il coinvolgimento dei diritti fondamentali dovrebbe comunque portare a intendere tale proprietà come inalienabile[114]. Infatti, è stato correttamente osservato che ricostruire la titolarità dei dati come proprietà – sebbene sembri rafforzare la tutela nei rapporti tra privati, come diritto assoluto invece che relativo – possa comportare dei rischi peggiori laddove l'agricoltore o agricoltrice, in base ai rapporti di forza vigenti, sia comunque costretto/a a cedere i propri dati[115]. Infatti, se nel contratto viene dedotta la titolarità stessa, in luogo di un semplice diritto di accesso, la cessione si configura come una rinuncia all'esercizio dei diritti relativi a tali informazioni. Naturalmente, tale esito è inaccettabile nel caso dei diritti fondamentali[116]. Sono necessari dunque dei limiti, volti a realizzare una *data sovereignty* individuale e collettiva[117].

Con questa lente si può leggere innanzitutto il *Data Act*, che non è una norma settoriale, ma quale norma generale si applica altresì ai dati connessi all'agricoltura. Tale regolamento dà risposta ad alcune situazioni critiche che si evidenziano altresì nella *smart agriculture*, con particolare riferimento ai rischi di *lock in* e di sfruttamento economico di chi produce i dati. Infatti, la norma prevede l'accesso dell'utente ai dati del prodotto (Artt. 3-4) e la loro portabilità, con la possibilità di condividere i dati

---

[112] J. DE BAERDEMAEKER, *op. cit.*, 58-59. Non è questa, tuttavia, la via seguita dal *Data Act*, come si vedrà di qui a poco.

[113] Il dibattito è molto fecondo sul punto, se è vero che esistono diverse opinioni critiche su tale impostazione. Si veda, ad es., J. DE BEER, *Ownership of open data: governance options for agriculture and nutrition*, 2016, 5-6, in *https://ssrn.com/abstract=3015958*; L. WISEMAN, J. SANDERSON, L. ROBB, *Rethinking Ag data ownership*, in *Farm Policy Journal*, 15, 1, 2018, 73-74; J. DREXL, *Data access and control in the era of connected devices*, 2018, 29-30 in *https://www.ip.mpg.de/fileadmin/ipmpg/content/aktuelles/aus_der_forschung/beuc-x-2018-121_data_access_and_control_in_the_area_of_connected_devices.pdf*, (ultima consultazione 02/12/2024).

[114] Sull'inalienabilità dei diritti fondamentali: A. BALDASSARRE, *Diritti della persona e valori costituzionali*, Torino, 1997, 84-86. In ogni caso, resta complesso cogliere il limite tra esercizio del diritto e titolarità dello stesso: G.B. ABBAMONTE, *The Protection of Computer Privacy under EU Law*, in *Columbia Journal of European Law*, 21, 2014-2015, 77; J.E.J. PRINS, *Property and privacy: European perspectives and the commodification of our identity*, in *Information Law Series,* 16, 2006, 241; G. RESTA, *La disponibilità dei diritti fondamentali e i limiti della dignità (note a margine della Carta dei diritti)*, in *Rivista di diritto civile*, 6, 2002, 807-808 e 816; S. ROSE-ACKERMAN, *Inalienability and The Theory of Property Rights*, in *Faculty Scholarship Series*, 580, 1985, 937-941 e 960-963, in https://openyls.law.yale.edu/handle/20.500.13051/4962. In merito, si è osservato che un indice più sicuro per individuare i limiti della disponibilità dei propri diritti è proprio l'osservazione dei valori concretamente in gioco: G. RESTA, *Il diritto alla protezione dei dati personali*, in F. CARDARELLI, S. SICA, V. ZENO-ZENCOVICH (a cura di), *Il codice dei dati personali: temi e problemi*, Milano, 2004, 52-53. È stato chiarito in dottrina che c'è un confine molto labile tra attribuzione di un diritto indisponibile e violazione del diritto stesso, e che l'imposizione al soggetto delle opzioni valoriali dell'autorità può essere mascherata da tutela della dignità del soggetto stesso: G. MANIACI, *La dittatura dei diritti indisponibili*, in *Diritto e questioni pubbliche*, 14, 2014, 675 e 697. In merito, è bene sottolineare che – come illustrato dalla dottrina civilistica – l'indisponibilità del bene non deve porsi in contrasto con l'autodeterminazione della persona, in quanto si tratta di un concetto diverso dall'incapacità di agire del suo titolare, e attiene al rango del valore costituzionale cui il bene è servente: O. DESSI, *L'indisponibilità dei diritti del lavoratore secondo l'art. 2113 c.c.*, Torino, 2011, 19.

[115] C. ATIK, *op. cit.*, 711.

[116] In questo senso, la disciplina si atteggerebbe in modo simile a quella sulla *privacy* – nel senso di richiedere una tutela rafforzata rispetto a quella meramente contrattuale – in quanto viene in rilievo la tutela di diritti fondamentali, ancorché non si tratti di dati personali: J.L. FERRIS, *op. cit.*, 333.

[117] C. ATIK, *op. cit.*, 713.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

con terze parti nel caso in cui si voglia cambiare fornitrice (Artt. 5-6). Sono previste altresì alcune condizioni che disciplinano la cessione dei dati medesimi, quali: l'obbligo per il *provider* di mettere a disposizione i dati a condizioni eque, ragionevoli e non discriminatori (cd. FRAND) e in modo trasparente al nuovo *provider* (art. 8); la previsione di un compenso ragionevole e non discriminatorio per la messa a disposizione dei dati (art. 9); la disciplina di clausole abusive tra imprese (art. 13).

Gli *stakeholder* hanno segnalato che la normativa dovrebbe essere seguita da una legislazione settoriale perché sia utile a dirimere i nodi centrali del *data sharing* in ambito agricolo. Un tentativo, in tal senso, è stato l'*EU Code of Conduct on Agricultural Data Sharing by Contractual Agreement*[118], siglato nel 2020 da aziende tecnologiche e agricole, che ha previsto alcune garanzie contrattuali come: il diritto di chi ha prodotto i dati sui dati stessi, con il conseguente diritto di mantenere il controllo sui dati e ricevere una compensazione rispetto al loro trattamento, di accedere ai dati stessi[119], di chiederne la pseudonimizzazione, di impedirne il trasferimento a terzi. Tali disposizioni dettano garanzie di rilievo, ma restano norme non vincolanti, che peraltro sono risultate scarsamente applicate nella realtà contrattuale, in quanto poco conosciute dagli operatori del settore[120]. Inoltre, una delle sfide ancora segnalati dagli *stakeholder* del settore è la costruzione di standard tecnologici per la portabilità, i quali dovrebbero caratterizzarsi – ad avviso di chi scrive – come *open standard*, per non replicare processi di 'cattura' privata della regolazione, capaci di inibire la *data sovereignty*[121].

Per altri versi, il *Data Act* riguarda la condivisione di dati tra imprese, ma non dirime tutti i conflitti che si possono determinare circa l'accesso alle informazioni, in quanto il suo obiettivo è soprattutto l'eliminazione delle incertezze normative (consid. 4) e degli squilibri contrattuali (consid. 5) che ostacolano il mercato unico[122]. Non riguardando direttamente i diritti fondamentali, il DA non esplicita l'indisponibilità dei diritti ivi garantiti; anzi, in alcuni casi prevede che il contratto possa contenere eccezioni alle relative regole. Inoltre, il Regolamento non disciplina l'apertura generalizzata dei dati[123], bensì solo quella che avviene con il consenso dell'utente e per il suo vantaggio (consid. 35); peraltro, il trasferimento coattivo di dati al soggetto pubblico è consentito dal DGA soltanto in casi specifici di eccezionale

[118] https://www.cema-agri.org/images/publications/brochures/EU_Code_of_conduct_on_agricultural_data_sharing_by_contractual_agreement_2020_ENGLISH.pdf (ultima consultazione 02/12/2024).

[119] J.K. ARCHER, C.A. DELGADILLO, *Key Data Ownership, Privacy and Protection Issues and Strategies for the International Precision Agriculture Industry*, in *Proceedings of the 13th International Conference on Precision Agriculture, online version,* Monticello, 2016, 6 e 12.

[120] R. GIAFFREDA, *D3.1: Definition of requirements for Agriculture Data Space building blocks*, rapporto commissionato all'interno del Progetto *Agri Data Space - Building a European framework for the secure and trusted data space for agriculture*, 31 marzo 2023, 5-6.

[121] L. NAGEL, D. LYCKLAMA (a cura di), *Design Principles for Data Spaces*, Position paper – Horizon 2020 project "OPEN DEI Aligning Reference Architectures, Open Platforms and Large-Scale Pilots in Digitising European Industry", International Data Spaces Association, 2021, 72.

[122] Peraltro, ci si è domandato se la nozione di *data sovereignty* adottata dall'UE non sia connotata da accenti protezionistici di natura commerciale: S. TORREGIANI, *Il Data Act: una versione europea del Data Nationalism?*, in *Rivista italiana di informatica e diritto*, 2, 2023, 139-140.

[123] Questa può essere vista come una prospettiva futura della regolazione: E. CREMONA, *Quando i dati diventano beni comuni: modelli di* data sharing *e prospettive di riuso*, in *Rivista italiana di informatica e diritto*, 2, 2023, 125-126.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

necessità legate all'interesse pubblico[124] (art. 14). In questo senso, una normativa *ad hoc* sarebbe necessaria per contemperare la condivisione generalizzata dei dati con altri diritti fondamentali connessi alla sovranità alimentare, nonché con la tutela della riservatezza del produttore o della produttrice. Tale iniziativa dovrebbe procedere di pari passo con un *empowerment* della comunità e delle stesse imprese agricole, in particolare delle piccole imprese e delle realtà contadine, nell'autodeterminazione delle condizioni di cessione dei dati. In questo senso, si rende necessario un processo iterativo, che accompagni l'autoregolazione e al tempo stesso impari dalla stessa, senza per questo tradursi in una delega in bianco ai grandi attori privati[125].

In questa sede si cercheranno di tracciare alcune direttrici sul possibile funzionamento di questo processo, seguendo la medesima indicazione di partire dalle pratiche già in essere, prima ancora dell'emanazione di una legislazione settoriale sugli EADS.

Nell'ordinamento eurounitario, un utile punto di riferimento è il DGA. Quest'ultimo riguarda soprattutto la condivisione dei dati detenuti dagli enti pubblici e, per quanto riguarda gli altri dati, condivisi per il pubblico interesse, si muove sul binario dell'apertura volontaria (consid. 4), disciplinando in particolare i *data intermediaries* (art. 2, n. 11; art. 10) e il *data altruism* (consid. 46, capo IV).

I *data intermediaries* sono definiti come «servizi[o] che mira[no] a instaurare, attraverso strumenti tecnici, giuridici o di altro tipo, rapporti commerciali ai fini della condivisione dei dati tra un numero indeterminato di interessati e di titolari dei dati, da un lato, e gli utenti dei dati, dall'altro, anche al fine dell'esercizio dei diritti degli interessati in relazione ai dati personali» (art. 2). L'ordinamento vuole che tali soggetti siano terzi rispetto al trattamento dei dati, con l'esclusione della possibilità per tali soggetti di processare e valorizzare essi stessi i dati, invece che fungere da intermediari in un rapporto tra i titolari dei dati e gli utenti dei dati (art. 2). A tali fine, sono previste delle regole comportamentali, ma anche di separazione strutturale (art. 12). Sono altresì esclusi i servizi utilizzati da un titolare dei dati per consentire l'utilizzo dei propri dati e i servizi di condivisione dei dati di natura non commerciale offerti da enti pubblici (art. 2). Viceversa, i servizi di intermediazione dei dati possono offrire servizi specifici strumentali all'intermediazione stessa, «come la conservazione temporanea, la cura, la conversione, l'anonimizzazione e la pseudonimizzazione, fermo restando che tali strumenti e servizi sono utilizzati solo su richiesta o approvazione esplicita del titolare dei dati o dell'interessato e gli strumenti di terzi offerti in tale contesto non utilizzano i dati per altri scopi» (art. 12).

---

[124] J. CHU, *Chapter V of the Data Act - Which should be the legal basis for B2G data sharing: 'exceptional need' or 'public interest'?*, in C. DUCUING, T. MARGONI, L. SCHIRRU (a cura di), *White Paper on the Data Act Proposal*, CiTiP Working Paper, 2022, 50-52, in https://ssrn.com/abstract=4259428 (ultima consultazione 02/12/2024). Cfr. COMMISSION STAFF, *Impact Assessment Report Accompanying the document Proposal for a Regulation of the European Parliament and of the Council on European Data Governance (Data Governance Act)*, Commission Staff Working Document, SWD/2020/295 final; EUROPEAN PARLIAMENTARY RESEARCH SERVICE, *Governing data and artificial intelligence for all: Models for sustainable and just data governance*, 2022, in https://www.europarl.europa.eu/RegData/etudes/STUD/2022/729533/EPRS_STU(2022)729533_EN.pdf (ultima consultazione 02/12/2024); COMMISSION – DIRECTORATE GENERAL FOR COMMUNICATIONS NETWORKS, *Content and Technology, Towards a European strategy on business-togovernment data sharing for the public interest: final report prepared by the High-Level Expert Group on Business-toGovernment Data Sharing'*, 2021, in https://op.europa.eu/en/publication-detail/-/publication/d96edc29-70fd-11eb-9ac9-01aa75ed71a1 (ultima consultazione 02/12/2024)

[125] L. NAGEL, D. LYCKLAMA (a cura di), *op. cit.*, 18.

Downloaded from www.biodiritto.org.

ISSN 2284-4503

Il regolamento prevede altresì la presenza di Autorità competenti per i servizi di intermediazione dei dati in funzione di vigilanza (artt. 13-14).

Tali soluzioni hanno il vantaggio del pragmatismo: a oggi, diverse iniziative di condivisione di dati (*data sharing initiatives – DSI*) si configurano giuridicamente come *data intermediaries*, coerentemente con la natura della cessione dei dati da parte delle aziende agricole: pur potendo implicare delle esternalità positive, la cessione non è motivata dall'altruismo, bensì dall'acquisto di un servizio.

L'istituto, così costruito, si presta ai fini della *data sovereignty* nella misura in cui può servire a creare 'corpi intermedi' – sul modello delle organizzazioni sindacali – capaci di mediare negoziazioni collettive che coinvolgono anche i diritti fondamentali. Una simile entità sembra poter prendere la forma delle cooperative di dati (consid. 31, art. 10, lett. c), DGA)[126]. Queste ultime sono descritte in modo generale dal diritto eurounitario[127], con flessibilità anche nella loro forma giuridica, fermo restando lo scopo di aiutare i propri membri nell'esercizio dei loro diritti in relazione a determinati dati. La norma non prevede esplicitamente che l'intermediazione possa avere una natura economica solidale, ma neanche impone che si tratti di un ente a scopo di lucro; pertanto, sarebbe ben possibile costituire un organismo di natura cooperativa capace di tenere in considerazione non soltanto gli interessi economici di chi produce, ma anche i diritti fondamentali delle diverse soggettività coinvolte. Peraltro, la norma sembra già essere costruita presupponendo che esista un'iniquità di relazione rispetto al *provider* che utilizza i dati, in quanto si estende anche alla cooperativa di dati l'impossibilità di fornire servizi ai propri soci, nettamente derogatoria rispetto all'idealtipo della cooperativa. Tale regola, ancorché possa apparire eccessivamente rigida[128], sembra ragionevole se si considerano i concreti squilibri di potere tra chi produce e chi processa i dati, che rendono difficilmente pensabile una cooperazione e fanno pensare, piuttosto, che sia necessaria un'aggregazione di interessi collettiva di chi conferisce i dati, indipendente dai *provider* stessi, perché la cooperativa possa tutelare efficacemente gli interessi dei soci.

Tali indicazioni sono sufficienti a instaurare alcune pratiche che potrebbero essere l'inizio di una capacitazione dei produttori e delle produttrici; tuttavia, alcune proposte *de iure condendo* potrebbero enfatizzare l'utilità dello strumento per riequilibrare situazioni di svantaggio.

Da un punto di vista sostanziale, è necessario che il soggetto pubblico determini a livello eurounitario i principi d'uso dei dati, con un quadro etico e normativo capace di tenere conto delle situazioni di potere e delle conseguenti necessità ridistributive[129]. In questo senso, è ormai ineludibile che i *data spaces* previsti dalla *data strategy* per la condivisione dei dati siano in grado di sostenere persone fisiche, entità giuridiche e gruppi dettando un insieme unitario di regole e principi di *design*, fondate

[126] Sui diversi tipi di *data intermediaries*: D. POLETTI, *Gli intermediari dei dati*, in *European Journal of Privacy Law and Technologies*, 1, 2022, 49-51.

[127] F. BRAVO, *Le Cooperative di Dati*, in *Project Papers del progetto di terza missione Cooperative di Dati dell'Alma mater studiorum di Bologna*, 3, 2023, 4 ss., in https://site.unibo.it/cooperative-di-dati/it/attivita-di-ricerca/pubblicazioni/bravo_le_cooperative_di_dati.pdf (ultima consultazione 02/12/2024).

[128] *Ibidem*, 17.

[129] In questo senso: EUROPEAN LAW INSTITUTE, AMERICAN LAW INSTITUTE, *Principles for a data economy: data transactions and data rights*, 2018, in https://www.europeanlawinstitute.eu/projects-publications/completed-projects-old/data-economy/ (ultima consultazione 02/12/2024). Cfr. L. PETRONE, *Il mercato digitale europeo e le cooperative di dati*, in *Project Papers del progetto di terza missione Cooperative di Dati dell'Alma mater studiorum di Bologna*, 3, 2023, 6-7.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

sull'obiettivo dell'autodeterminazione e della *data sovereignty*[130]. Ciò richiede altresì un reale protagonismo di chi produce i dati, che favorisce la fiducia nell'ecosistema, e un diritto della concorrenza attento alle implicazioni delle dominanze economiche sui diritti fondamentali, oltre che su quelli consumeristici[131]. In tal modo, si potrebbe costruire un regime di accesso ai dati basato sul legittimo interesse a fare il miglior uso dei dati[132].

Da un punto di vista soggettivo, si è accennato prima che non vi è un espresso riconoscimento di cooperative composte da soggettività vulnerabilizzate – ad esempio, le realtà contadine e i relativi gruppi di consumo – sebbene tale riconoscimento potrebbe essere un primo potenziale passo per riconoscere uno statuto di vantaggio a tali entità, in chiave di azione positiva. Infatti, come si è osservato a livello teorico, la modalità più efficace per rispondere alle vulnerabilità è creare spazi di autodeterminazione per le persone coinvolte, con cui queste ultime possono acquisire la capacità di difendere collettivamente i propri interessi.

In secondo luogo, l'efficacia delle previsioni si arresta allorché il *data intermediary* non può comunque sovradeterminare la persona titolare dei dati, anche quando vi dovesse essere una chiara prevalenza dell'interesse generale. La previsione di uno statuto specifico per le cooperative di dati potrebbe in futuro dare luogo alla previsione di vere e proprie deleghe a una negoziazione collettiva da parte di specifiche categorie di *data intermediaries*[133], sul modello delle organizzazioni sindacali, in vista

---

[130] L. NAGEL, D. LYCKLAMA (a cura di), *op. cit.*, 10; J. SONNEN, J. MOELLER e T. HUELSMANN, in M. FARALDI, *How To Build A Common European Agricultural Data Space Workshop Report*, 16 settembre 2020, 14-15, in https://digital-strategy.ec.europa.eu/en/events/expert-workshop-common-european-agricultural-data-space (ultima consultazione 02/12/2024). Similmente, i *data spaces* sono stati prefigurati nel progetto UE Agri Data Space come infrastruttura di facilitazione di alleanze tra iniziative di *data sharing*: M. EISENTRÄGER, I. SEIFERT, D. FOTAKIDIS, G. FIROGENIS, *Governance Scheme and Business Models of a Common European Agricultural Data Space*, Rapporto commissionato all'interno del Progetto Agri Data Space - Building a European framework for the secure and trusted data space for agriculture, 6-7.

[131] Per un'interpretazione costituzionalmente orientata del diritto della concorrenza: J. DREXL, *Economic Efficiency versus Democracy: On the Potential Role of Competition Policy in Regulating Digital Markets in Times of Posttruth Politics*, in *Max Planck Institute for Innovation & Competition Research Paper*, 16-16, 6 dicembre 2016, 20 ss., in *https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2881191* (ultima consultazione 02/12/2024). Si veda, ad esempio, sull'ipotesi di considerare la lesione della *privacy* come un abuso, in quanto fattore che diminuisce la qualità del prodotto senza che il consumatore possa avere migliore trattamento presso i *competitor*: AUTORITÉ DE LA CONCURRENCE, BUNDESKARTELLAMT, *op. cit.*, 25; G. DE MINICO, Big Data *e la debole resistenza delle categorie giuridiche*. Privacy *e* lex mercatoria, in *Diritto pubblico*, 1, 2019, 104; N. NEWMAN, *Search, Antitrust and the Economics of the Control of User Data*, in *Yale Journal on Regulation*, 31, 2, 2014, 441-442; A. SOLA, *Sull'intreccio privacy-concorrenza in tale ambito regolatorio, si veda: Primi cenni di regolazione europea nell'economia dei dati*, in *MediaLaws*, 3, 2021, 208-209.

[132] J. DREXL, *Data access and control in the era of connected devices*, cit., 43.

[133] L. PETRONE, *op. cit.*, pp. 14-15 Analogamente suggeriscono gli esiti del progetto *Agri Data Space*: «Consent flow should be streamlined by the data space to avoid bottlenecks in data sharing. To do so different mechanisms might be implemented, like consent automation based on predefined policies or its delegation to entities like data cooperatives that do so based on the delegators' interests»: M. RYAN, M. BIZOT-ESPIARD, *Design principles and guidelines for agricultural data spaces based on legislation and ethical principles*, Rapporto commissionato all'interno del Progetto *Agri Data Space - Building a European framework for the secure and trusted data space for agriculture*, 31 marzo 2024, 6-7. In tema di *data altruism*, ma – ad avviso di chi scrive – nella medesima logica: W. VEIL, *Data altruism: how the EU is screwing up a good idea*, in *AlgorithmWatch discussion paper*, 4-5, in https://algorithmwatch.org/en/eu-and-data-donations/ (ultima consultazione 02/12/2024).

Downloaded from www.biodiritto.org.
ISSN 2284-4503

dell'apertura dei dati agricoli. In tal caso, dovrebbero essere previste misure vincolanti sulla 'porta aperta', sul voto capitario, sulla parità di trattamento delle persone socie, sul voto capitario, non determinato dall'investimento economico, sui diritti partecipativi a favore di produttori e produttrici, sullo standing dei soggetti diversi titolati a intervenire – ad esempio, consumatori e consumatrici, organizzazioni ambientaliste, reti contadine, etc. – come esponenti dell'interesse generale.

Infine, in questo quadro si colloca altresì la necessità di investimenti volti ad appianare il *gap* delle competenze digitali, che oggi rende piccoli/e produttori e produttrici soccombenti rispetto alle ATP. Si rende dunque necessario un intervento pubblico per il *reskilling* e l'*upskilling*, soprattutto delle categorie più precarie, volto non solo all'immissione nel mercato del lavoro, ma anche alla dotazione di strumenti per l'analisi critica delle informazioni e l'autodeterminazione effettiva nelle scelte sui dati. A ciò si aggiunge che tali operazioni richiedono in ogni caso risorse quali dispositivi e macchine con adeguata capacità di calcolo, nonché tempo da dedicare, sottraendolo alle attività strettamente agricole[134].

Nei termini usati da una certa dottrina politologica, le imprese *high tech* operano una vera e propria 'curatela' sui dati, intesa come pratica di selezionare, rappresentare e conservare i contenuti, anche in base a valutazioni contingenti e prospettive future[135]. La medesima 'curatela' dovrebbe poter essere esercitata dalla comunità, che potrebbe sviluppare proprie tecnologie per scopi di interesse generale. Per questo, 'free access isn't necessarily fair access'[136], laddove non sia equamente ridistribuita la capacità e la possibilità giuridica di farsi protagonista, invece che semplice 'pubblico' di tale curatela. Fuor di metafora, serve una democratizzazione delle competenze per consentire fare scelte tecniche di merito sulla base delle proprie opzioni valoriali[137], esattamente nella maniera in cui oggi fanno uso della propria strumentazione agricola. In assenza di simili azioni positive, il diritto non potrà che aumentare le vulnerabilità, rinunciando alla piena esplicazione del potenziale della *data sovereignty*.

5. Conclusioni

Il lavoro svolto mette a nudo gli effetti dell'informatizzazione nel settore agricolo, tentando di guardare in modo particolare ai diritti fondamentali, sotto la specifica lente dei rimedi alle diverse vulnerabilità che sussistono in merito al godimento del diritto al cibo. Tale paradigma ha mostrato di essere una chiave di lettura utile per discutere rischi e potenzialità della tecnica, in quanto è capace di cogliere le interdipendenze dei diversi attori sociali nell'affrontare il progresso tecnologico. Allo stesso tempo, si tratta di una nozione che coglie la necessità di un intervento pubblico attivo contro la precarietà di alcune specifiche categorie, dando centralità all'autodeterminazione.

La *food sovereignty* e la *data sovereignty* necessitano di strumenti normativi nuovi per garantire l'acquisizione di potere decisionale da parte di tante soggettività fragili del mondo agricolo e contadino,

[134] A. FRASER, *Land grab/data grab*, cit., 890.

[135] A. FRASER, *The digital revolution, data curation, and the new dynamics of food sovereignty construction*, cit., 209.

[136] M. CAROLAN, *'Smart' farming techniques as political ontology: Access, sovereignty and the performance of neoliberal and not-so-neoliberal worlds*, in *Sociologia Ruralis*, 57, 2, 2017, 20; S. ROTZ, E. DUNCAN, M. SMALL, J. BOTSCHNER, R. DARA, I. MOSBY, M. REED, E.D.G. FRASER, *op. cit.*, 114.

[137] E. JAKKU, B. TAYLOR, A. FLEMING, C. MASON, S. FIELKE, C. SOUNNESS, P. THORBURN, *op. cit.*, 9.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

che rischiano di vivere ulteriori percorsi di marginalizzazione all'interno di mercati che tendono fortemente alla concentrazione. Di fronte ai cambiamenti climatici, la salvaguardia dell'agroecologia e dell'agricoltura contadina costituisce un presidio essenziale di fronte alla standardizzazione di modi di produzione tendenti a creare economie di scala favorevoli al profitto di breve periodo, non sempre capaci di cogliere le implicazioni sociali e ambientali di lungo termine. Alla tecnica spetta prendere posizione, decidendo di supportare queste strutture o affidarsi alla 'mano invisibile', illudendosi che la lotta al cambiamento climatico e la resilienza dei sistemi produttivi accadano automaticamente quando l'IA si mette a servizio della produttività.

Di fronte a un ingresso dell'IA che si fonda oggi su oligopoli e concentrazioni di mercato *data based*, si pone la necessità di rafforzare il controllo dei dati da parte degli altri attori in campo, come presupposto di una trustworthy AI, dotata di sistemi sicuri, affidabili e robusti. Se tali valori sono presi sul serio, la costruzione della fiducia non può essere delegata a un processo di *marketing* e 'sensibilizzazione' sull'uso della *smart agriculture*[138], ma deve basarsi su sistemi di trasparenza e controllo effettivo delle informazioni da parte di produttori e produttrici.

In tale frangente, i *data spaces* previsti dalla strategia europea sui dati, unitamente al nuovo *framework* regolamentare per la condivisione dei dati stessi, promettono di costruire un'infrastruttura che ridistribuisce il valore generato da tali masse di informazioni. La sfida è dunque sfruttare appieno il potenziale dei nuovi diritti garantiti dal DA – come i diritti di accesso e portabilità dei dati – e dal DGA, che riconosce e promuove la costituzione di forme cooperative per la negoziazione sui dati. In futuro, l'auspicio è che tali strumenti si rendano capaci di alimentare processi equi che vanno verso la liberazione delle informazioni e la loro apertura a fini di interesse generale, ivi inclusa la sovranità alimentare. Ciò richiederà, nel futuro, sforzi sempre maggiori perché il settore pubblico possa riconoscere e sostenere tali aggregazioni, dettando al tempo stesso un quadro chiaro di principi per la condivisione. Ci sarà bisogno altresì di nuove competenze, da favorire mediante un intervento pubblico massivo, volto a sviluppare una visione critica sull'uso dei dati e sugli algoritmi, strumentale a rinnovare la possibilità di comprendere le implicazioni delle tecnologie scelte e decidere sulle stesse.

Si tratta di un percorso appena avviato, costellato di questioni aperte sulla *governance* dei futuri spazi di dati. Tuttavia, è opinione di chi scrive che solo la pratica continua dell'autodeterminazione, mediante strumenti cooperativi di aggregazione di interessi tra produttori e produttrici di dati, aprirà la strada verso soluzioni sempre più condivise e rispettose degli umani e degli ecosistemi.

[138] A. FRASER, *Land grab/data grab*, cit., 897.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Il contributo dell'intelligenza artificiale simbiotica nella protezione delle vittime vulnerabili e nel contrasto della violenza di genere

Lorenzo Pulito [*]

THE CONTRIBUTION OF SYMBIOTIC ARTIFICIAL INTELLIGENCE TO THE PROTECTION OF VULNERABLE VICTIMS AND IN THE FIGHT AGAINST GENDER-BASED VIOLENCE

ABSTRACT: Given the importance of assessing the risk of recurrence and escalation of violence for the prevention and fight against gender-based violence, the present paper emphasises the need to introduce specific assessment procedures, targeted and calibrated to the vulnerability of victims. Furthermore, the essay reconstructs the framework of algorithmic systems currently used in this field, the challenges they pose on a technical as well as ethical and legal level, and their potential developments. Finally, using the symbiotic paradigm, the paper provides suggestions for the acceptance of artificial intelligence in this sensitive area.

KEYWORDS: Risk assessment; gender-based violence; vulnerability; symbiotic artificial intelligence; significant human control.

ABSTRACT: Premessa l'importanza che la valutazione del rischio di ripetizione ed escalation della violenza assume per la prevenzione e il contrasto della violenza di genere, il contributo sottolinea la necessità di adottare procedure valutative specifiche, calibrate e mirate alla particolare vulnerabilità delle vittime. Inoltre, il saggio ricostruisce il quadro dei sistemi algoritmici attualmente in uso in tale dominio, delle sfide che sollevano, sia sul piano tecnico, che su quello etico e giuridico, e dei loro potenziali sviluppi. Infine, ricorrendo al paradigma simbiotico, fornisce suggerimenti per l'accettabilità dell'intelligenza artificiale in questo delicato ambito.

PAROLE CHIAVE: Valutazione del rischio; violenza di genere; vulnerabilità; intelligenza artificiale simbiotica; controllo umano significativo.

[*] *Ricercatore di diritto processuale penale, Università di Bari. Mail: lorenzo.pulito@uniba.it.* *Questo lavoro è stato parzialmente sostenuto dal progetto FAIR – Future AI Research (PE00000013), nell'ambito del programma MUR del PNRR finanziato dal NextGenerationEU. Contributo sottoposto a doppio referaggio anonimo.*

Downloaded from www.biodiritto.org.
ISSN 2284-4503

1. Introduzione

La violenza di genere rappresenta una delle violazioni dei diritti umani più sistematiche a livello mondiale ed impegna i legislatori ad individuare soluzioni per combatterla più efficacemente, tanto in ambito internazionale (dove la Convenzione di Istanbul[1], ratificata dall'Italia con l. 27 giugno 2013, n. 77[2], ha gettato i pilastri sinteticamente riassumibili nelle "tre P" – prevenzione, protezione e punizione) ed europeo (è di recentissima emanazione la Direttiva 2024/1385[3], atto normativo che segue l'adesione dell'Ue alla stessa Convenzione di Istanbul), quanto a livello nazionale (dove è stato da ultimo "rafforzato" il "Codice Rosso"[4], la cui disciplina – "ispirata" dal noto caso Talpis[5] – è già stata *medio tempore* incisa dalle norme immediatamente precettive della c.d. "Riforma Cartabia"[6] e dal *restyling* della l. 8 settembre 2023 n. 122[7]).

La Convenzione di Istanbul obbliga gli Stati aderenti a predisporre le necessarie azioni legislative o altre misure per proteggere tutte le vittime da ogni ulteriore atto di violenza (art. 18), esigendo indagini e procedimenti efficaci (art. 49), all'interno dei quali deve essere offerta una protezione adeguata ed immediata (art. 50) dalla vittimizzazione reiterata. In base a quanto disciplinato dal suo art. 51, la corretta valutazione del rischio di reiterazione dei comportamenti violenti (*risk assessment*) si colloca a monte della sua efficiente gestione (*risk management*), attuabile attraverso la messa in atto delle misure di protezione, che potranno consistere in interventi che offrono rifugio immediato alla vittima o in provvedimenti che determinano un allontanamento dell'autore della violenza ed una inibizione ad avvicinare la vittima (artt. 23, 52 e 53).

---

[1] Convenzione del Consiglio d'Europa sulla prevenzione e la lotta contro la violenza nei confronti delle donne e la violenza domestica, approvata il 7 aprile 2011 ed aperta alla firma l'11 maggio 2011 a Istanbul. Sulla Convenzione di Istanbul, v., tra gli altri, S. De Vido, M. Frulli (a cura di), *Preventing and Combating Violence Against Women and Domestic Violence. A Commentary on the Istanbul Convention*, Cheltenham, 2023.

[2] Su cui cfr. G. Battarino, *Note sulla attuazione in ambito penale e processuale penale della Convenzione di Istanbul sulla prevenzione e la lotta contro la violenza nei confronti delle donne e la violenza domestica*, in *Dir. pen. cont.*, 2 ottobre 2013, 1 ss.

[3] Direttiva (UE) 2024/1385 del Parlamento europeo e del Consiglio del 14 maggio 2024 sulla lotta alla violenza contro le donne e alla violenza domestica, in G.U.U.E., 24 maggio 2024, L. 1385. Per un primo inquadramento, M. Ferrari, *Violenza contro le donne: l'Unione europea adotta finalmente la direttiva (UE) 2024/1385*, in *Eurojus*, 17 giugno 2024. Sull'originaria Proposta di direttiva cfr. A. Pitrone, *Il lungo (ed incidentato) percorso della lotta alla violenza contro le donne nell'Unione europea. Dalla questione dell'adesione alla Convenzione di Istanbul alla proposta di una direttiva "ad hoc"*, in *Ordine internazionale e diritti umani*, 3, 2022, 692 ss.

[4] L. 19 luglio 2019 n. 69, c.d. "Codice Rosso". A riguardo, N. Triggiani, *L'ultimo tassello nel percorso legislativo di contrasto alla violenza domestica e di genere: la legge 'Codice Rosso'*, in *Proc. pen. giust.*, 2, 2020, 451 ss.

[5] Corte EDU, 2 marzo 2017, *Talpis c. Italia*, ric. n. 41237/14. Per un commento, v. M. Castellaneta, *I ritardi e le misure inadeguate per combattere il fenomeno della violenza contro le donne rappresentano delle violazioni*, in *Guida dir.*, 14, 2017, 102 ss.

[6] D. Bianchi, *Le modifiche al codice penale immediatamente precettive: prescrizione del reato e sospensione condizionale*, in *Dir. pen. proc.*, 11, 2021, 1468 ss.

[7] G. Amato, *Intervento sicuramente apprezzabile ma non centra l'obiettivo perseguito*, in *Guida dir.*, 36, 2023, 17 ss.; Id., *Il supporto dell'informatizzazione necessario per garantire il controllo*, *ivi*, 23 ss.; A. Marandola., *Codice Rosso rafforzato*, in *Dir. pen. proc.,* 11, 2023, 1420 ss.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

L'apprezzamento del rischio rappresentato dall'autore del reato è richiesto, sin dal primo contatto tra vittima e autorità, dalla citata Direttiva 2024/1385, in via aggiuntiva rispetto agli obblighi della valutazione individuale a norma dell'art. 22 della Direttiva 2012/29/UE[8].
L'uso di sistemi di intelligenza artificiale predittiva – e in particolare di approcci di *machine learning* – potrebbe ottimizzare tali processi valutativi, ma solleva diverse sfide, sia sul piano tecnico, che su quello etico[9] e giuridico, con ampie implicazioni per gli individui e la società. Il contributo mira a comprendere queste ultime, in modo da mitigare efficacemente i rischi associati, nonché ad individuare le condizioni per l'accettabilità di tali sistemi intelligenti, al fine di assicurare valutazioni affidabili e rispettose dei diritti umani.

2. Vittime vulnerabili e obblighi di protezione nella cornice convenzionale

Sebbene Fondamentale in materia risulta anche l'esegesi della Corte europea dei diritti dell'uomo che, spesso richiamando la Convenzione di Istanbul, ha «da tempo proceduto ad una progressiva messa a fuoco delle centralità della procedura di valutazione del rischio»[10].
Dalle disposizioni contenute nella Convenzione europea dei diritti umani (in particolare, dagli artt. 2, 3 e 4) la Corte di Strasburgo fa discendere in capo agli Stati ben precisi obblighi sostanziali e procedurali, individuati come strumento per contrastare la criminalità di genere e volti a garantire una tutela effettiva dei diritti convenzionali.
Per la Corte europea, in ossequio al principio di uguaglianza sostanziale, quando ad essere titolari dei diritti sono i soggetti più vulnerabili, gli Stati contraenti sono obbligati a fornire un livello di protezione più elevato, adeguato ai loro bisogni e alle loro particolari condizioni[11].
Ne deriva che, proprio «ai fini dell'adempimento degli obblighi di tutela delle vittime vulnerabili», quali sono quelle di violenza di genere, sessuale e domestica, specialmente se straniere[12], «appare cruciale l'adozione di efficaci procedure di valutazione e gestione del rischio»[13], sia da parte della polizia giudiziaria che dei giudici.

[8] Direttiva 2012/29/UE del Parlamento europeo e del Consiglio, che istituisce norme minime in materia di diritti, assistenza e protezione delle vittime di reato e che sostituisce la decisione quadro 2001/220/GAI, in G.U.U.E., 14 novembre 2012, L 315/57. In dottrina, v. L. LUPÁRIA (a cura di), *Victims and criminal justice. European standards and national good practices,* Milano, 2015.

[9] Sulle sfide etiche e sulle opportunità offerte dalle tecnologie digitali e intelligenti nell'affrontare la violenza domestica cfr. P. NOVITZKY, J. JANSSEN, B. KOKKELER, *A systematic review of ethical challenges and opportunities of addressing domestic violence with AI-technologies and online tools*, in *Heliyon*, 6, 2023, 1 ss.

[10] V. BONINI, *Protezione della vittima e valutazione del rischio nei procedimenti per violenza domestica tra indicazioni sovranazionali e deficit interni*, in *Sist. pen*., 3, 2023, 53.

[11] Corte EDU, 24 giugno 2008, *Chapman c. Regno Unito*, ric. n. 27970/02, ha rappresentato il primo caso in cui la Corte ha avvertito l'esigenza che dei soggetti potessero godere di una tutela specifica in quanto "vulnerabili". Sul punto, cfr. E. LICATA, *Vulnerabilità e tutela dei "core rights": gli obblighi di protezione in materia di criminalità di genere derivanti dalla Cedu*, in *Sist. pen*., 2, 2024, 63 s.

[12] Si parla in tal caso di doppia vulnerabilità. Su tale profilo e sulle interconnessioni problematiche tra violenza di genere e migrazione, si v. l'opera di A. DI STASI, R. CARDIN, A. IERMANO, V. ZAMBRANO (a cura di), *Donne migranti e violenza di genere nel contesto giuridico internazionale ed europeo*, Napoli, 2023.

[13] E. LICATA, *op. cit.*, 65.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

La Corte alsaziana individua le caratteristiche che la valutazione del rischio deve presentare nell'ambito della violenza di genere e domestica, ampliandola – rispetto al c.d. "Osman test"[14] – con una serie di parametri valutativi volti a intercettare più adeguatamente l'esposizione di queste vittime al pericolo di subire nuove aggressioni, destinato altrimenti a rimanere nell'ombra se si adoperassero i meccanismi valutativi tradizionali, "tarati" esclusivamente sulla persona dell'accusato e sulle caratteristiche oggettive dell'azione criminosa già commessa[15].
L'operazione valutativa, oltre ad essere improntata ai canoni della prontezza e della tempestività, dovrebbe giovarsi di strumenti standardizzati riconosciuti dalla comunità internazionale[16] ed essere condotta in modo «*autonomous, proactive and comprehensive*»[17].

3. L'ultimissimo tassello nel percorso legislativo interno di contrasto alla violenza domestica e di genere e gli spazi di valutazione del rischio di ripetizione ed *escalation* della violenza

La l. 24 novembre 2023, n. 168, recante «Disposizioni per il contrasto della violenza sulle donne e della violenza domestica», ribattezzata legge "Codice Rosso Rafforzato" o "legge Roccella"[18], rappresenta l'ultimissimo tassello di un lungo e articolato percorso legislativo nazionale diretto alla tutela delle

[14] Corte EDU, Grande Camera, 28 ottobre 1998, *Osman c. Regno Unito*, ric. n. 23452/94.

[15] Corte EDU, Grande Camera, 15 giugno 2021, *Kurt c. Austria*, ric. n. 62903/15, par. 164.

[16] Per una ricognizione degli strumenti si rimanda alla pubblicazione ufficiale dell'Unione europea elaborata dall'European Institute for Gender Equality (EIGE), *A guide to risk assessment and risk management of intimate partner violence against women for police*, Lussemburgo, 2019, 13 s., reperibile in https://eige.europa.eu/publications-resources/publications/guide-risk-assessment-and-risk-management-intimate-partner-violence-against-women-police# (ultima consultazione 10/07/2024).

[17] Corte EDU, Grande Camera, 15 giugno 2021, *Kurt c. Austria*, par. 168. L'autonomia e la proattività della valutazione richiedono che la stessa non si basi esclusivamente sulla percezione del rischio riportata dalla vittima; la globalità implica che la valutazione riguardi una pluralità di fattori, quali quelli individuali, relazionali, statici e dinamici.

[18] Sulla quale v., tra i primi volumi pubblicati, C. Cecchella, C. Parodi, *Il nuovo codice rosso. L. n. 168/2023 contro la violenza sulle donne e la violenza domestica. Profili penali e civili*, Milano, 2023; V. De Gioia, G. Molfese, *Il nuovo codice rosso. Commento alla L. 24 novembre 2023, n. 168 recante le nuove disposizioni per il contrasto della violenza sulle donne e della violenza domestica (Riforma Roccella),* Piacenza, 2024; J.F. Lorefice, M. Storzini, *Il nuovo codice rosso. Fondamenti in materia di misure cautelari e misure di prevenzione*, Milano, 2024; F. Piccioni, *Abusi e violenza domestica. Il nuovo Codice rosso e le opportunità di difesa*, Rimini, 2024. Nell'impossibilità di tratteggiare in questa sede i molteplici aspetti dell'intervento normativo, si rinvia, tra i primi commenti apparsi in rivista, a G. Amato, *Una corsia accelerata e preferenziale per definire le fattispecie più gravi*, in *Guida dir.*, 46, 2023, 59 ss.; Id., *Sul tema delle priorità nelle Procure serve una legge per i criteri generali*, *ivi*, 63 ss.; Id., *Violazione atti cautelari "attenuati": rimediata un'incoerenza normativa*, *ivi*, 76 s.; Id., *Strumenti cautelari d'urgenza, arresto differito e via dalla casa*, *ivi*, 78 ss.; L. Biarella, *La gestione dei rischi aumenta l'efficacia*, *ivi*, 43 ss.; V. Cannas, *Violenza di genere e misure di prevenzione: continuano le modifiche al codice rosso*, in *Dir. pen. proc.*, 2, 2024, 158 ss.; A. Famiglietti, *Disposizioni per il contrasto della violenza sulle donne e della violenza domestica*, in *Proc. pen. giust.*, 1, 2024, 9 ss.; A. Marandola, *Violenza di genere: accelerazione del rito e formazione per l'adeguamento della legislazione domestica agli standard europei*, in *Dir. pen. proc.*, 2, 2024, 167 ss.; Ead., *L'accelerazione delle richieste cautelari e specializzazione dei magistrati*, in *Giur. it.*, 2024, 973 ss.; A. Marandola, L. Risicato, *Pregi e limiti del ''Codice Rosso''*, *ivi*, 959 s.; M. Pierdonati, *Il rafforzamento della tutela delle persone vulnerabili e l'eterogenesi dei fini*, *ivi*, 989.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

vittime di violenza domestica e di genere: tale complesso di disposizioni costituisce, ormai, un vero e proprio sottosistema normativo sia nell'ambito del codice penale, ove è possibile individuare un ampio catalogo dei reati che si caratterizzano quale manifestazione di violenza domestica e di genere, sia nell'ambito del codice di procedura penale, ove si è prevista una disciplina differenziata ed un *iter* sempre più preferenziale[19].

La stratificazione degli interventi normativi ha investito numerose disposizioni del codice di rito, incidendo particolarmente sul piano delle misure cautelari (applicabili in deroga ai criteri generali e da adottarsi, in maniera inedita, secondo precise cadenze temporali acceleratorie[20]) e delle misure precautelari (ora "affrancate" dalla necessità che il soggetto sia colto in stato di flagranza e, in taluni casi, eseguibili anche quando questa è "differita" [21]).

Nonostante la moltitudine degli interventi normativi, continua a mancare una disciplina dedicata specificatamente alla valutazione del rischio di ripetizione ed *escalation* della violenza relazionale, che, attingendo dai criteri forniti dalla giurisprudenza europea, delinei una cornice organica che individui i soggetti tenuti a compierla, ne regolamenti i tempi, i modi e le caratteristiche.

L'adempimento «finisce per inscriversi all'interno di schemi normativi pensati e costruiti per altri processi valutativi» che, per limitarsi agli strumenti a caratura processuale, sono quelli concernenti: 1) l'apprezzamento dei presupposti di applicazione delle misure cautelari; 2) l'adozione di pre-cautele che prevedono spazi di discrezionalità, come nei casi di cui all'art. 381 c.p.p., all'art. 382-*bis* c.p.p., da ultimo introdotto e che consente l'arresto in flagranza differita[22], e all'art. 384-*bis* c.p.p., che – così come novellato - consente al pubblico ministero, anche fuori dei casi di flagranza, di disporre l'allontanamento urgente dalla casa coniugale senza attendere il provvedimento del giudice[23]; 3) in fase di esecuzione della pena, l'accesso a misure alternative alla detenzione o a benefici penitenziari che comportino una restituzione di spazi di libertà al soggetto *in vinculis*.

4. Lo *Spousal Assault Risk Assessment* (SARA)

A fronte del proliferare degli spazi normativi che richiedono la valutazione del rischio a cui la vittima vulnerabile è esposta e la prognosi sulla reiterazione delle condotte violente, il *risk assessment* resta affidato a prassi virtuose, che si avvalgono della consolidata esistenza di protocolli valutativi, sviluppati nell'ambito degli studi e delle ricerche sulla violenza di genere.

Tra questi vi è il SARA (*Spousal Assault Risk Assessment*)[24], un metodo messo a punto in Canada e ritenuto valido dalla comunità scientifica. Il modello, nella versione semplificata SARA-S, prende in

[19] Così, in relazione al c.d. "Codice Rosso", N. Triggiani, *op. cit.*, 2, 2020, 451 ss.

[20] L. Algeri, *Tempi rapidi per l'adozione delle misure cautelari a protezione delle vittime*, in *Dir. pen. proc.*, 2, 2024, 172 ss.

[21] N. Rombi, *Novità in materia precautelare e cautelare e nuovi obblighi informativi*, in *Giur. it.*, 2024, 967 ss.

[22] L. Algeri, *L'arresto in flagranza differita per reati di violenza di genere e domestica*, in *Dir. pen. proc.*, 2, 2024, 181 ss.

[23] A. Marandola, *I nuovi presidi a tutela della vittima: rimedi pre-cautelari, cautelari e obblighi informativi*, in *Dir. pen. proc.*, 2, 2024, 186 ss.

[24] Per la trattazione tecnica del quale si v. A.C. Baldry, *Dai maltrattamenti all'omicidio. La valutazione del rischio di recidiva e dell'uxoricidio*, Milano, 2016.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

considerazione dieci fattori di rischio (raggruppati in due sezioni, che tengono conto rispettivamente le condotte violente del partner e l'adattamento psico-sociale dell'accusato) e cinque fattori di vulnerabilità della vittima; infine, la *check-list* richiede anche di verificare se ricorrono ulteriori circostanze significative (quali *child abuse* e, come peraltro richiesto espressamente dall'art. 51, par. 2, della Convenzione di Istanbul, disponibilità di armi da fuoco). Si procede a stabilire il livello di presenza o meno di ogni singolo fattore, collocandolo allo stato attuale (ultime quattro settimane) e nel passato (prima di un mese); successivamente, si riporta il livello di rischio di recidiva, che può risultare basso, medio o elevato, sia nell'immediato (entro due mesi) che nel lungo termine (oltre i due mesi). Al valutatore viene anche chiesto di verificare se esiste un rischio di violenza letale e di un'*escalation* dell'atto violento.

È rimarcata la necessità di effettuare una nuova valutazione del rischio quando si verificano le cosiddette "circostanze critiche", in cui la vittima necessita di maggiori misure protettive (come nel caso di separazione, di contrasti circa l'affidamento o la visita dei figli, di scarcerazione del maltrattante).

L'efficacia operativa di questo sistema, riconosciuta dagli stessi organi di governo della magistratura[25], dipende dal fatto che la valutazione finale del rischio non viene effettuata soltanto sulla base della quantità di fattori presenti, ma soprattutto «sulla tipologia degli stessi e sulla loro interazione ed evoluzione nel tempo»[26].

Oltre a quello appena illustrato, tra i più noti strumenti valutativi riconosciuti a livello internazionale si annoverano B-SAFER (*brief spousal assault form for the evaluation of risk*), DASH (*domestic abuse, stalking and harassment and honour-based violence*)[27], MARAC (*multi-agency rirsk assessment conference*), DVSI (*domestic violence screening instrument*), ODARA (*Ontario domestic assault risk assessment*)[28], nonché VioGén (*Sistema de seguimiento integral en los casos de violencia de género*)[29], che è stato costruito proprio a partire dal modello SARA[30].

[25] Si allude, in particolare, alla risoluzione sulle linee guida in tema di organizzazione e buone prassi per la trattazione dei procedimenti relativi a reati di violenza di genere e domestica (delibera CSM 9 maggio 2018).

[26] L. Algeri, *Tempi rapidi per l'adozione delle misure cautelari a protezione delle vittime*, cit., 177.

[27] A riguardo, v. E. Turner, G. Brown, J. Medina-Ariza, *Predicting Domestic Abuse (Fairly) and Police Risk Assessment*, in *Psychosocial Intervention*, 3, 2022, 145 ss. La validità predittiva del Dash, strumento che pone l'accento sul giudizio professionale strutturato, è messa in dubbio da alcuni ricercatori, come può leggersi in M. Bland, *Algorithms Can Predict Domestic Abuse, But Should We Let Them?*, in H. Jahankhani, B. Akhgar, P. Cochrane, M. Dastbaz (a cura di), *Policing in the Era of AI and Smart Societies. Advanced Sciences and Technologies for Security Applications*, Cham, 2020, 147.

[28] Su questo strumento si v. J. Hegel, K.D. Pelletier, M.E. Olver, *Predictive Properties of the Ontario Domestic Assault Risk Assessment (ODARA) in a Northern Canadian Prairie Sample*, in *Criminal Justice and Behavior*, 3, 2022, 411 ss., i quali richiamano anche l'interessante caso giurisprudenziale canadese *Ewert c. Canada*.

[29] Sullo strumento e sul suo *background* normativo cfr. J.L. González-Álvarez, J.J. López-Ossorio, C. Urruela, M. Rodríguez-Díaz, *Integral Monitoring System in Cases of Gender Violence. VioGén System*, in *Behavior & Law Journal*, 1, 2018, 29 ss. Sulla sua validazione, J.J. López-Ossorio, J.L. González-Álvarez, J.M. Muñoz Vicente, C. Urruela Cortes, A. Andrés Pueyo, *Validation and Calibration of the Spanish Police Intimate Partner Violence Risk Assessment System (VioGén)*, in *Journal of Police and Criminal Psychology*, 4, 2019, 439 ss.

[30] Lo rilevano A. Valdivia, C. Hyde-Vaamonde, J. García-Marcos, *Judging the algorithm: A case study on the risk assessment tool for gender-based violence implemented in the Basque country*, in *arXiv*, 2022 (arXiv:2203.03723), 3.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

5. Simbolismo vs. effettività: *risk assessment* e ausilio algoritmico. Il sistema VioGén

Sebbene la considerazione dell'esistenza di prassi virtuose sia valsa ad evitare (anche) all'Italia censure da parte del GREVIO (*Group of Expert on Action against Violence against Women and Domestic Violence*), l'organismo indipendente di monitoraggio dell'implementazione della Convenzione di Istanbul non ha mancato di sollecitare il nostro Paese ad intraprendere azioni migliorative[31].

Del resto, l'arsenale di strumenti introdotto dal legislatore rischia di entrare in contrasto con l'ideale di un diritto effettivo se tali arnesi sono lasciati nella teca e non vengono adoperati per la mancanza o inadeguatezza della valutazione del rischio di esposizione a nuovi atti di violenza.

Non è casuale che, nonostante gli sforzi legislativi, l'Italia sia stata condannata più volte (ed anche in tempi recenti) dalla Corte di Strasburgo, ma le censure non sono dipese da inadeguatezze normative, bensì ruotano sempre intorno a deficit nel *risk assessment*, vuoi perché non eseguito dall'autorità giudiziaria[32], inerte anche nell'avviare un'indagine effettiva nella quale condurre un'adeguata valutazione del rischio[33], vuoi perché quest'ultima non era stata immediata e proattiva come avrebbe dovuto[34].

È incidentalmente interessante notare come, in ben due casi, la Corte abbia evidenziato la scarsa conoscenza delle caratteristiche strutturali della violenza relazionale da parte dell'autorità giudiziaria[35], che invece rappresenta una «precondizione generale [per] riconoscere i tratti di una realtà ad elevata complessità [e] approcciarsi ad essa in modo libero da pregiudizi e *bias* ancora molto radicati sul piano culturale»[36].

Alla luce di tali decisioni risulta pertinente domandarsi se «le difficoltà che la complessità e la globalità della valutazione del rischio comportano» non possano «essere superate impiegando risorse tecnologiche che consentono, attraverso l'ausilio della macchina, una più veloce analisi di quantità anche assai imponenti di dati»[37].

Il quesito richiede di ritornare al summenzionato sistema attuariale VioGén, un algoritmo adoperato in Spagna e che trova base giuridica nella *Ley Orgánica* 1/2004.

Il funzionamento del sistema VioGén è basato su due questionari (*Protocolo Dual*): *Valoración Policial del Riesgo* (VPR) e *Valoración Policial de la Evolución del Riesgo* (VPER). Questi protocolli di va-

[31] Si v. il par. 233 del documento del GREVIO intitolato *Baseline Evaluation Report Italy*, 2020, disponibile al seguente url: https://rm.coe.int/grevio-report-italy-first-baselineevaluation/168099724e (ultima consultazione 10/07/2024).

[32] Corte EDU, 7 aprile 2022, *Landi c. Italia*, ric. n. 10929/19. A riguardo, si rinvia al commento di A.A. Dei Cas, *La Corte europea condanna ancora l'Italia per violazione degli obblighi positivi derivanti dall'art. 2 nei confronti di vittime di violenze domestiche*, in *Arch. pen. (web)*, 2, 2022, 1 ss.

[33] Corte EDU, 16 giugno 2022, *De Giorgi c. Italia*, ric. n. 23735/19.

[34] Corte EDU, 7 luglio 2022, *M.S. c. Italia*, ric. n. 32715/19.

[35] Nel caso Landi la Corte ha rilevato come non fosse stata effettuata una valutazione del rischio di letalità capace di considerare lo specifico contesto della violenza domestica, così censurando l'ignoranza da parte dell'autorità giudiziaria delle specificità strutturali e delle dinamiche di tale forma di violenza domestica; nel caso De Giorgi la violenza domestica era stata erroneamente sminuita ed interpretata alla stregua di una mera conflittualità coniugale.

[36] V. Bonini, *op. cit.*, 65.

[37] V. Bonini, *op. cit.*, 66.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

lutazione vengono esaminati e revisionati da un *team* di esperti multidisciplinari; dal 2019 la valutazione del rischio è stata effettuata tramite VPR5.0-H e VPER4.1.

Quando una donna sporge denuncia contro il suo aggressore, gli agenti di polizia compilano il modulo VPR5.0-H. Questo comprende cinque domini con trentacinque indicatori di rischio: ogni elemento viene valutato come "presente" e "non presente". In questo modo la raccolta delle informazioni è standardizzata su tutto il territorio nazionale. Una volta compilato il modulo, il sistema assegna un punteggio di rischio violenza di genere, i cui livelli sono "non apprezzato" (*no apprezzado*), "basso" (*bajo*), "medio" (*medio*), "alto" (*alto*) ed "estremo" (*extremo*). Gli agenti di polizia possono solo modificare il punteggio portandolo a un livello di rischio più elevato. A seguito dello studio dell'*Equipo Nacional de Revisión Pormenorizada de Homicidios en el contexto de la Violencia de Género* (EHVdG), VioGén è stato adattato per rilevare i casi con rischio di aggressione letale[38]. Questi vengono segnalati come di "particolare rilevanza"; in tale evenienza (così come nel caso di bambini a rischio o in situazione di vulnerabilità) una *Diligencia Automatizada* è allegata alla VPR al fine di richiamare l'attenzione di pubblici ministeri e giudici e raccomandare ulteriori approfondimenti.

Lo sviluppo delle metodologie computazionali e la disponibilità di *big data* digitali non solo rendono possibile applicare tecniche di *machine learning* alla previsione dei crimini di genere[39], ma – si sostiene – un approccio di apprendimento automatico riesce a valutare meglio i rischi rispetto a quello tradizionale[40]. Infatti, i risultati empirici relativi alla valutazione di un modello creato applicando tecniche di *machine learning* ed alimentato con dati estratti dal sistema VioGén dimostrerebbero una netta *outperformance* dell'approccio automatizzato, con un sensibile miglioramento rispetto al sistema di valutazione del rischio preesistente[41].

Si tratta di un ambito di impiego molto promettente, ma che – come condivisibilmente sostenuto – potrà avere successo solo a condizione di considerare le numerose sfide che si pongono[42] e che riguardano, tra le altre, la trasparenza[43], la *accountability*[44] e la (eccessiva) dipendenza dall'algoritmo.

---

[38] Sul funzionamento di VioGén e sul suo duplice meccanismo algoritmico, si rinvia a A. Giraldi, *Intelligenza artificiale e predictive policing nella rinnovata fase d'indagine*, in A. Massaro (a cura di), *Intelligenza artificiale e giustizia penale*, Roma, 2020, 83.

[39] Si v. il documento *The External Audit of the VioGén System*, elaborato da Eticas foundation, disponibile al seguente url: https://eticasfoundation.org/wp-content/uploads/2024/07/ETICAS-FND-The-External-Audit-of-the-VioGen-System-1.pdf (ultima consultazione 10/07/2024), 2022, 9.

[40] J. Grogger, S. Gupta, R. Ivandic, T. Kirchmaier, *Comparing Conventional and Machine-Learning Approaches to Risk Assessment in Domestic Abuse Cases*, in *Journal of Empirical Legal Studies*, 1, 2021, 114.

[41] Á. González-Prieto, A. Brú, J.C. Nuño, J.L. González-Álvarez, *Hybrid machine learning methods for risk assessment in gender-based crime*, in *Knowledge-Based Systems*, 2, 2023, 1 ss.

[42] M. Bland, *op. cit.*, 148.

[43] Con il noto problema della *black-box*. Utile punto di partenza sul nodo della trasparenza è il quadro ALGO-CARE per il processo decisionale di M. Oswalda, J. Graceb, S. Urwinc, G.C. Barne, *Algorithmic risk assessment policing models: lessons from the Durham HART model and 'Experimental' proportionality*, in *Information & Communications Technology Law*, 2, 2018, 223 ss.

[44] Nel senso che, in relazione a tutti gli usi della tecnologia decisionale algoritmica, l'obiettivo deve essere quello di «*to augment human legal intelligence, not to replace it*», M. Hildebrandt, *Law as computation in the era of artificial legal intelligence. Speaking law to the power of statistics*, in *The University of Toronto Law Journal*, vol. 68, suppl. 1 (*Artificial Intelligence, Technology, and the Law*), 2018, 33.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Stessi aspetti problematici che, insieme ad altri non meno delicati – come i riflessi negativi sulla presunzione di innocenza, il pericolo di cristallizzazione degli individui nel loro passato[45] nonché la supervisione umana minima e incoerente sul risultato del sistema automatizzato[46] –, sono già emersi in relazione al più "tradizionale" VioGén e che è facilmente pronosticabile assumeranno maggior enfasi critica con la transizione di questo strumento verso modelli di *machine learning*.

Tanto più che l'utilizzo di questi *tools* – in questa sede circoscritto al presupposto di una previa denuncia e, dunque, senza tenere conto delle ulteriori questioni che sollevano gli strumenti di *predictive policing*[47] e il loro impiego in chiave pre-investigativa[48] – non appare essere appannaggio delle sole autorità di *law enforcement*.

Strumenti algoritmici, come EPV-R, assistono infatti il processo decisionale del giudice: nella fattispecie, lo strumento, anch'esso ispirato al SARA, viene utilizzato nei tribunali baschi, dove i giudici decidono il livello di protezione da accordare alla vittima di violenza di genere sulla scorta dell'*output* algoritmico. Anche questo strumento di valutazione solleva questioni legali con riferimento ad almeno «*three factors: (1) opaque implementation, (2) efficiency's paradox and (3) feedback loop*»[49], per mitigare i quali si è ritenuto «*essential that judicial users understand their important overseeing role, that an algorithm cannot replace them and that its automatic assessment can be wrong*»[50], coltivando una visione antropocentrica dell'interazione tra intelligenza umana e artificiale[51].

Il criterio di non esclusività del dato algoritmico per la decisione giudiziaria, con la correlata necessità che lo stesso sia corroborato da ulteriori e diversi elementi di prova, secondo l'avvertimento della Suprema Corte del Wisconsin nel caso *Loomis*, è stato già eletto da fonti di *soft law*[52] e sostenuto da

[45] E. Falletti, *Human rights protection and high-risk AI systems: the Spanish model in gender-based violence prevention*, in *SSRN*, 16 aprile 2022, 6 s.

[46] Così *The External Audit of the VioGén System*, cit., 32 s., dove si raccomanda di «*accompanying the VioGén score with the justification of the police officers*», in quanto ciò «*would support the accountability of the Police for the risk assessment*».

[47] Con tale locuzione si indicano «le attività rivolte allo studio e all'applicazione di metodi statistici con l'obiettivo di "predire" chi potrà commettere un reato o dove e quando potrà essere commesso un reato, al fine di prevenirne la commissione»: così L. Camaldo, *Intelligenza artificiale e investigazione penale predittiva,* in *Riv. it. dir. proc. pen*., 1, 2024, 234. Sulla polizia predittiva, v. L. Algeri, *Intelligenza artificiale e polizia predittiva*, in *Dir. pen. proc*., 6, 2021, 724 ss.; K. Blount, *Using artificial intelligence to prevent crime: implications for due process and criminal justice*, in *AI & Soc*, 1, 2024, 359 ss., J.L.M. McDaniel, K. G. Pease (eds.), *Predictive Policing and Artificial Intelligence*, Abingdon, 2021; R. Orlandi, *Uso poliziesco dell'intelligenza artificiale. L'insegnamento del Bundesverfassungsgericht*, in *Cass. pen*., 7/8, 2023, 2167 ss.; B. Perego, *Predictive policing: trasparenza degli algoritmi, impatto sulla privacy e risvolti discriminatori*, in *BioLaw Journal – Rivista di BioDiritto*, 2, 2020, 447 ss.; E. Pietrocarlo, *Predictive policing: criticità e prospettive dei sistemi di identificazione dei potenziali criminali*, in *Sist. pen*., 28 settembre 2023, 1 ss.

[48] Diversamente, M. Bland, *op. cit.*, 142 ss., tratta la valutazione del rischio di reiterazione del reato accostandola alla nozione di "pre-crime".

[49] A. Valdivia, C. Hyde-Vaamonde, J. García-Marcos, *op. cit.*, 11 s.

[50] A. Valdivia, C. Hyde-Vaamonde, J. García-Marcos, *op. cit.*, 14.

[51] K. La Regina, *I.A. e ragionamento giuridico: la giustizia prevedibile*, in G.M. Baccari, P. Felicioni (a cura di), *La decisione penale tra intelligenza emotiva e intelligenza artificiale*, Milano, 2023, 181.

[52] Cfr. Commissione europea per l'efficienza della giustizia (CEPEJ), *Carta etica europea sull'utilizzo dell'intelligenza artificiale nei sistemi giudiziari e negli ambiti connessi*, Strasburgo, 3-4 dicembre 2018. In argomento, v. S. Quattrocolo, *Intelligenza artificiale e giustizia: nella cornice della Carta etica europea, gli spunti per*

Downloaded from www.biodiritto.org.
ISSN 2284-4503

quelle unionali[53]; inoltre, risulta accolto in una certa misura anche dall'*AI Act*[54], che chiarisce sin dall'esordio (sia al *considerando* 1, che all'art. 1) il suo obiettivo primario di «promuovere la diffusione di un'intelligenza artificiale (IA) antropocentrica».

Stando al regolamento sull'intelligenza artificiale, caratterizzato da un approccio «della individuazione e della gestione del "rischio" di violazione dei diritti fondamentali delle persone fisiche»[55], i software come quelli in esame sono riconducibili ai sistemi ad "alto rischio"[56] e devono sottostare a regole precise per la valutazione d'impatto, utilizzare dati ad alto standard qualitativo, avere risultati tracciabili, essere sottoposti a costante supervisione umana durante il loro utilizzo, possedere un elevato livello di robustezza, sicurezza e precisione[57]. Inoltre, il Regolamento 2024/1689 non pregiudica l'esercizio dei diritti fondamentali riconosciuti dagli Stati membri, oltre che a livello di Unione, sicché operano i divieti vigenti sul piano interno, come ad esempio quello di perizia psicologica e criminologica di cui all'art. 220, comma 2 c.p.p.[58]. Sempre curandosi di garantire i diritti della difesa e la presunzione di innocenza (come l'*AI Act* prevede ai *considerando* 48 e 59) e, laddove i dati da cui lo strumento predittivo provengano da risposte o informazioni fornite dallo stesso indagato o imputato, di non sopraffare la garanzia del *nemo tenetur se detegere*[59].

6. Intelligenza artificiale simbiotica e nuovi assetti paradigmatici

un'urgente discussione tra scienze penali e informatiche, in *Leg. pen. (web)*, 18 dicembre 2018, 4.

[53] Cfr. Direttiva 2016/680/UE del Parlamento europeo e del Consiglio del 27 aprile 2016 relativa alla protezione delle persone fisiche con riguardo al trattamento dei dati personali da parte delle autorità competenti a fini di prevenzione, indagine, accertamento e perseguimento di reati o esecuzione di sanzioni penali, nonché alla libera circolazione di tali dati e che abroga la decisione quadro 2008/977/GAI del Consiglio, in G.U.U.E., 4 maggio 2016, L 119/89. A riguardo, v. G. Baccari, *Il trattamento (anche elettronico) dei dati personali per finalità di accertamento dei reati*, in A. Cadoppi, S. Canestrari, A. Manna, M. Papa (diretto da), *Cybercrime*, Torino, 2019, 1611 ss. In particolare, l'art. 11, par. 1, della citata direttiva impone agli Stati membri di introdurre il divieto di decisioni basate «unicamente» su un trattamento automatizzato di dati, compresa la profilazione, che producano effetti giuridici negativi o incidano significativamente sui diritti fondamentali dell'interessato.

[54] Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio del 13 giugno 2024 che stabilisce regole armonizzate sull'intelligenza artificiale e modifica i regolamenti (CE) n, 300/2008, (UE) n, 167/2013, (UE) n, 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 in G.U.U.E., 12 luglio 2024, L 1689.

[55] S. Quattrocolo, *Prova e intelligenza artificiale*, in C. Conti, A. Marandola (a cura di), *Prova scientifica*, Milano, 2023, 471.

[56] Cfr. l'art. 5, lett. d) nonché l'art. 6, par 2, che rinvia all'allegato III, di cui si vedano i punti 6. e 8. Secondo E. Falletti, *op. cit.,* VioGén è inquadrabile come "sistema IA ad alto rischio".

[57] L'esemplificazione di tali requisiti è di S. Quattrocolo, *Prova e intelligenza artificiale*, cit., 472.

[58] Sulla possibilità di ricondurre gli *outputs* di software di *risk assessment* al novero delle prove generate automaticamente e, segnatamente, al paradigma della prova peritale, cfr. S. Quattrocolo, *Prova e intelligenza artificiale*, cit., 480. In senso contrario, ritiene che «i risk assessment tools non hanno una funzionalità tesa ad analizzare la personalità dell'autore del fatto» e che, pertanto, sarebbe «improprio ricondurli nell'ambito di operatività del divieto di cui all'art. 220 comma 2 c.p.p.», M. Montagna, *Prognosi personologica, commisurazione della pena e applicazione di misure di sicurezza*, in G.M. Baccari, P. Felicioni, *La decisione penale tra intelligenza emotiva e intelligenza artificiale*, cit., 241, la quale reputa pur sempre necessario «verificare, di volta in volta, la loro compatibilità con i principi fondamentali del sistema penale italiano», dovendosi impedire ogni forma di determinismo penale e che dal diritto penale del fatto si giunga a un diritto penale del profilo d'autore.

[59] M. Montagna, *op. cit.*, 245.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

La necessità di preservare la centralità dell'uomo ha indotto la dottrina a sostenere come sarebbe preferibile assegnare agli strumenti predittivi del rischio in esame compiti di mero *screening* iniziale, di guisa che tali *tools* verrebbero ad eseguire una «prima attività di rilevazione potenziale del rischio in modo generalizzato e standardizzato, senza creare un ingolfamento delle attività procedimentali», mentre soltanto laddove tale *step* «si concluda con la rilevazione di un numero predeterminato di *red flags*, si impo[rrebbe] una valutazione del rischio più accurata che può involgere accertamenti più approfonditi e il ricorso a competenze valutative specifiche», in modo da «nulla toglie[re] alla (anzi, semmai restituendo) pienezza della valutazione umana»[60].

Tuttavia, siffatta tranquillante proposta pare limitativa delle potenzialità di questi strumenti, fruibili non solo per far scattare una *red flag*, ma anche per assistere l'autorità (tanto di *law enforcement* come pure quella giudiziaria) nel processo decisionale sulle misure da adottare, connettendo il singolo caso agli n-esimi altri "interiorizzati" dalla macchina e ricercando tra loro le interrelazioni più nascoste. Sono attività che la mente umana, computazionalmente limitata e condizionata da strategie cognitive definite euristiche, non potrebbe parimenti eseguire.

Ma finché si ragioni assumendo gli artefatti artificiali come strumenti, anziché come agenti integrati in una squadra e dotati di autonomia, adattabilità e capacità collaborative, sì da creare un sistema multi-agente più ampio e più capace delle singole entità[61], sarà difficile coniugare l'efficienza computazionale dei sistemi algoritmici con l'approccio incentrato sull'uomo.

Occorre, piuttosto, individuare strategie di partnership collaborativa tra esseri umani e macchine[62], da concepirsi come entità che si aiutano a vicenda per un obiettivo comune.

Il potenziamento della collaborazione uomo-macchina e del *teaming* sociotecnico, con relazioni reciprocamente vantaggiose, che aumentino (e valorizzino) le capacità cognitive umane anziché sostituirle, è l'obiettivo dell'intelligenza artificiale simbiotica (SAI)[63].

Il paradigma simbiotico, che si ispira all'idea preconizzata dall'informatico e psicologo statunitense Joseph Carl Robnett Licklider nel saggio dal titolo *Man-Computer Symbiosis* (1960)[64], individua nel processo cooperativo uomo-macchina la soluzione antagonista «[a]lla sostituzione e [a]lla delega in bianco alle macchine»[65].

La sua adozione consente di superare l'interrogativo se collocare l'intelligenza artificiale alla stregua di un aiutante del giudice o di un suo sostituto. D'altro canto, la risposta fornita a tale quesito dal re-

[60] V. Bonini, *op. cit.*, 66 s.

[61] T. O'Neill, N. McNeese, A. Barron, B. Schelble, *Human–autonomy teaming: a review and analysis of the empirical literature*, in *Human Factors*, 5, 2022, 904 ss.

[62] N. Lettieri, A. Guarino, R. Zaccagnino, D. Malandrino, *Keeping judges in the loop: a human–machine collaboration strategy against the blind spots of AI in criminal justice*, in *Soft Computing*, 16, 2023, 11275 ss.

[63] A. Carnevale, A. Lombardi, F.A. Lisi, *Exploring Ethical and Conceptual Foundations of Human-Centred Symbiosis with Artificial Intelligence*, in G. Boella, F.A. D'Asaro, A. Dyoub, L. Gorrieri, F.A. Lisi, C. Manganini, G. Primiero (a cura di), *Proceedings of the 2nd Workshop on Bias, Ethical AI, Explainability and the role of Logic and Logic Programming co-located with the 22nd International Conference of the Italian Association for Artificial Intelligence (AI*IA 2023),* CEUR, 3615, 2023, 30.

[64] J.C.R. Licklider, *Man-computer symbiosis*, in *IRE Transactions on Human Factors in Electronics*, 1, 1960, 4 ss.

[65] P. Marra, *Giustizia digitale simbiotica e sue prospettive procedurali*, in *L'Ircocervo - First Italian digital journal of Legal Methodology, General Theory of Law and Doctrine of the State*, M.N. Campagnoli, P. Marra (a cura di), *Artificial Intelligence and Neuro-cognitive Sciences in Law: From Symbiosis to Substitution*), 1, 2024, 26.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

cente regolamento europeo che, imponendo al «processo decisionale finale d[i] rimanere un'attività a guida umana» (così il *considerando* 61), promette «la superiorità dell'intelletto umano su quello automatizzato», ha lasciato insoddisfatti, in quanto «nell'articolato normativo [si] dimentica la promessa», non individuandosi nel testo «le condizioni necessarie perché ciò accada»[66].

Lo sforzo si dovrebbe concentrare, invece, sul passaggio dalla supervisione al *teaming*, vera cifra che consente di rendere realmente significativo il controllo umano sull'algoritmo[67].

Quest'ultimo risulta concretamente possibile in quanto diviene il prodotto del contesto associativo, nel quale si registra un allineamento di valori tra gli esseri umani e l'intelligenza artificiale; l'integrazione tra il comportamento dei primi e quello dei sistemi automatizzati consente agli agenti umani non solo di sfruttare le capacità di quest'ultimi per eseguire compiti, ma anche di garantirne il controllo, preservando la "consapevolezza situazionale" da parte dell'uomo[68].

Poiché la valutazione del rischio nell'ambito della violenza di genere richiede di essere ripetuta ciclicamente, con crescente grado di complessità, sono immediatamente percepibili i vantaggi di questo approccio, superandosi la dinamica dell'"*automation bias*", la sfida dell'"*automation surprise*" per l'operatore umano e il pericolo che costui abdichi alla propria responsabilità decisionale[69], problemi tutti in varia misura registrati nell'esperienza del sistema VioGén.

La cooperazione, per essere effettiva, «non può che fondarsi sulle capacità comunicative, che non si riducono alla semplice trasmissione di contenuti, del tutto inutile se questi non vengono processati ed effettivamente compresi»[70]; pertanto, la collaborazione uomo-computer richiede che gli *outputs* siano trasparenti e spiegabili[71]: simbioticamente, l'utente umano dovrebbe avere la possibilità d'intervenire sulla valutazione del sistema di intelligenza artificiale, di esaminare gli argomenti e, trovandosi in disaccordo con il "ragionamento" della macchina, anche di modificarli[72]. Conseguentemente, sarà anche più facile impugnare le decisioni che abbiano utilizzato l'*output* algoritmico, rispondendo più adeguatamente al problema della contestabilità, una delle nuove sfide democratiche poste

---

[66] G. De Minico, *Giustizia e intelligenza artificiale: un equilibrio mutevole*, in *Rivista AIC*, 2, 2024, 107.

[67] A. Tsamados, L. Floridi, M. Taddeo, *Human control of AI systems: from supervision to teaming*, in *AI Ethics*, 2024, p. 1 ss. G. Ubertis, *Intelligenza artificiale, giustizia penale, controllo umano significativo*, in *Sist. pen.*, 4, 2020, 83 s., chiarisce quali siano le condizioni alle quali assoggettare l'impiego della macchina in sede giurisdizionale, affinché sia mantenuto un controllo umano significativo.

[68] Sul problema della "loss of situational awareness" v. A. Tsamados, L. Floridi, M. Taddeo, *op. cit.*, 3 ss.

[69] A. Tsamados, L. Floridi, M. Taddeo, *op. cit.*, 5.

[70] P. Marra, *op. cit.*, 24.

[71] Nel senso che un approccio antropocentrico è fondamentale per l'intelligenza artificiale simbiotica e che, pertanto, la trasparenza e la *explainability* sono requisiti fondamentali per stabilire la fiducia nei sistemi SAI e renderli accettabili eticamente e legalmente, v. F. Lisi, A. Carnevale, A. Dyoub, A. Lombardi, P. Marra, L. Pulito, *Acceptability of symbiotic artificial intelligence: Highlights from the FAIR project*, in S. Di Martino, C. Sansone, E. Masciari, S. Rossi, M. Gravina (a cura di), *Proceedings of the Ital-IA Intelligenza Artificiale - Thematic Workshops co-located with the 4th CINI National Lab AIIS Conference on Artificial Intelligence (Ital-IA 2024)*, CEUR, vol. 3762, 2024, 115.

[72] Sui modelli linguistici argomentativi di grandi dimensioni per processi decisionali spiegabili e contestabili, pensati anche per scenari complessi legali, v. G. Freedman, A. Dejl, D. Gorur, X. Yin, A. Rago, F. Toni, *Argumentative Large Language Models for Explainable and Contestable Decision-Making*, in *arXiv*, 2024 (arXiv:2405.02079), 1 ss.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

dall'automazione del processo decisionale[73]. Del resto, lo stesso *AI Act* non manca di sottolineare più volte come occorra garantire che «le previsioni, le raccomandazioni o le decisioni del sistema di IA possano essere efficacemente ribaltate e ignorate»[74].

7. Osservazioni conclusive (a partire da ChatGPT)

Interrogando *ChatGPT*, a domanda se «l'intelligenza artificiale simbiotica può aiutare le forze dell'ordine e i giudici nella valutazione del rischio nei casi di violenza di genere e nel contrasto a tale fenomeno», il *chatbot* ha fornito una risposta positiva e – almeno apparentemente – rassicurante, sostenendo che «l'intelligenza artificiale simbiotica può essere un potente alleato nella lotta contro la violenza di genere, fornendo strumenti avanzati per la valutazione del rischio, il monitoraggio, la prevenzione e l'azione legale»; ha enunciato diversi esempi; ha suggerito di considerare alcune sfide e implicazioni etiche, e ha concluso affermando che «è fondamentale implementare queste tecnologie in modo etico e responsabile, garantendo la protezione dei diritti delle vittime e la giustizia equa».
L'artefatto mostra di possedere una "consapevolezza" e una sensibilità sulle questioni di genere, terreno sfidante per la piena garanzia dei diritti e l'effettività delle tutele, che a tante volte sembra smarrirsi nell'essere umano.
Se, in teoria, le macchine offrono potenza di calcolo e capacità di automazione in grado di gestire in modo efficiente attività ripetitive e ad alta intensità di dati, mentre gli esseri umani forniscono le capacità cognitive ed emotive necessarie per la creatività e l'empatia[75], proprio quest'ultima – fondamentale per rispondere ai bisogni di chi è vulnerabile – risulta troppo spesso mancare nei casi di violenza di genere, dove si registrano da parte degli operatori comportamenti giudicanti, che possono persino indurre le vittime a rinunciare di denunciare[76].
Guardare alla simbiosi come costrutto socio-tecnico e come modello procedurale del processo decisionale[77] significa altresì combinare gli approcci quantitativi con quelli qualitativi.
Il processo di operazionalizzazione cui si è accennato, necessario per allineare i valori etici e le regole giuridiche tra l'intelligenza artificiale e gli esseri umani, dovrebbe rappresentare anche l'occasione per ripensare e migliorare l'assetto esistente, allo scopo di alimentare il costante processo di rafforzamento delle tutele, di cui la Direttiva 1385/2024 rappresenta l'ultimo epilogo a livello europeo: in

[73] A.A. Tubella, A. Theodorou, V. Dignum, L. Michael, *Contestable Black Boxes*, in *arXiv*, 2020 (arXiv:2006.05133), 2.

[74] Cfr. il *Considerando* 141, nonché artt. 60, par. 4, lett. k) e 61, par. 1, lett. d), Regolamento (UE) 2024/1689.

[75] F. Lisi, A. Carnevale, A. Dyoub, A. Lombardi, P. Marra, L. Pulito, *Acceptability of symbiotic artificial intelligence: Highlights from the FAIR project*, *op. cit.*, 1.

[76] Cfr. *The External Audit of the VioGén System*, cit. 23. Un interessante studio sull'utilizzo di modelli di apprendimento automatico per prevedere il *disengagement* dai procedimenti legali nei casi di violenza domestica è stato portato avanti da E. Escobar-Linero, M. García-Jiménez, M.E. Trigo-Sánchez, M.J. Cala-Carrillo, J.L. Sevillano, M. Domínguez-Morales, *Using machine learning-based systems to help predict disengagement from the legal proceedings by women victims of intimate partner violence in Spain*, in *PLoS ONE*, 7 giugno 2023, 1 ss.

[77] P. Marra, L. Pulito, A. Carnevale, A. Lombardi, A. Dyoub, F.A. Lisi, *A Procedural Idea of Decision-making in the Context of Symbiotic AI*, in A. Dix, M. Roach, T. Turchi, A. Malizia, B. Wilson (a cura di), *Proceedings of the 1st International Workshop on Designing and Building Hybrid Human-AI Systems co-located with 17th International Conference on Advanced Visual Interfaces (AVI 2024)*, CEUR, vol. 3701, 2024, 4. s.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

particolare, questo atto normativo si sforza di eliminare le barriere che impediscono l'accesso delle donne alla giustizia, propedeutico alla stessa valutazione del rischio, e si preoccupa della dimensione culturale del fenomeno della violenza di genere.

Così come per le macchine, è indispensabile allora che anche gli operatori siano sottoposti ad adeguato *training*, in modo da creare le opportune condizioni di supporto per la vittima[78].

Oltre alla necessaria specializzazione, andrà garantito il supporto legale e psicologico già al momento in cui vengono fornite le risposte ai questionari, delle quali i *tools* si nutrono, e che il "punteggio" sia spiegato dalla macchina e giustificato dagli operatori, anche per evitare effetti di vittimizzazione secondaria[79]. Infine, occorrerà implementare i sistemi algoritmici adoperati in tale dominio mediante metodologie di co-progettazione e meccanismi di feedback con le organizzazioni femminili come un modo per esplorare le vulnerabilità attuali ed emergenti. La valutazione delle prestazioni di tali sistemi non potrà limitarsi alla loro analisi tecnica, ma è necessario che tenga conto delle esperienze, dell'applicazione pratica e delle percezioni degli *stakeholders* che si imbattono negli stessi o che li utilizzano nella loro attività lavorativa[80].

[78] Cfr. *The External Audit of the VioGén System*, cit. 25.

[79] Cfr. *The External Audit of the VioGén System*, cit., 30.

[80] Cfr. *The External Audit of the VioGén System*, cit., 30. Un primo quadro per comprendere e valutare i sistemi SAI è presentato in A. Carnevale, A. Lombardi, F.A. Lisi, *A human-centred approach to symbiotic AI: Questioning the ethical and conceptual foundation*, in *Intelligenza artificiale*, 1, 2024, 9 ss.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Intelligenza artificiale e diritti delle donne: siamo dinanzi ad un algoritmo maschilista?

Susanna Viggiani *

ARTIFICIAL INTELLIGENCE AND WOMEN'S RIGHTS: ARE WE UP AGAINST A SEXIST ALGORITHM?
ABSTRACT: The algorithms used by AI systems are developed by scientists and computer scientists, who train the algorithm on the basis of a set of data, which - sometimes - can hide their own biases and errors. One of the most widespread and pervasive prejudices is that of gender, which tends to manifest itself in a variety of contexts, from automated decision-making processes in human resources and credit access systems to its use in so-called 'revenge porn', whose victims are, in fact, mostly women. Such forms of discrimination result in a violation of the principles and freedoms that underpin our democracies. Therefore, to struggle against the perpetuation of such stereotypes and prejudices, it is important to be awareness-raising and to address diversity as a fundamental priority of policies and institutional structures.

KEYWORDS: Algorithms; artificial intelligence; principle of equality; discrimination; women.

ABSTRACT: Gli algoritmi impiegati dai sistemi di AI sono elaborati da scienziati ed informatici, i quali addestrano l'algoritmo sulla base di una serie di dati, che – talvolta - possono nascondere i loro stessi pregiudizi ed errori. Uno dei pregiudizi maggiormente diffuso e pervasivo è quello di genere, il quale tende a manifestarsi in svariati contesti, dai processi decisionali automatizzati nelle risorse umane, nei sistemi di accesso al credito sino ai suoi utilizzi nel c.d. "*Revenge porn*", le cui vittime sono, infatti, per la maggioranza donne. Tali forme di discriminazione si traducono in una violazione dei principi e delle libertà che sono alla base delle nostre democrazie. Per questi motivi, per combattere il perpetuarsi di tali stereotipi e pregiudizi, è necessario esserne consapevoli e affrontare la diversità come priorità fondamentale delle politiche e delle strutture istituzionali.

PAROLE CHIAVE: Algoritmi; intelligenza artificiale; principio di uguaglianza; discriminazioni; donne.

* *Consulente Legale e Privacy, specializzata SPISA. Mail: viggianisusanna@gmail.com.* *Contributo sottoposto a doppio referaggio anonimo.*

Downloaded from www.biodiritto.org.
ISSN 2284-4503

1. Il divieto di discriminazione

Il divieto di discriminazione affonda le proprie radici nell'idea più generale di uguaglianza, quale pilastro fondamentale dello Stato di diritto[1]. Il diritto antidiscriminatorio è il risultato dell'incontro di norme di diritto nazionale, di norme di recepimento di direttive europee[2] e di norme primarie dell'Unione Europea. A livello nazionale, il divieto di discriminazione trova la sua consacrazione nel principio di uguaglianza, di cui all'art. 3 commi 1 e 2 della Costituzione, nella duplice forma di uguaglianza formale e sostanziale. Tali principi sostengono non solo che tutti i cittadini hanno pari dignità dinanzi alla legge senza discriminazione di sesso, razza, religione, ma altresì il divieto di adottare comportamenti differenziati in situazioni eguali. Il divieto di discriminazione cui si riferisce la nostra Costituzione non prevede, però, una parificazione assoluta e indiscriminata, nel senso che il divieto si rivolge a tutte quelle caratteristiche immutabili del soggetto o scelte personalissime dello stesso che ne rafforzano la dignità[3]. Ne consegue, dunque, che il *Leitmotiv* del divieto di discriminazione si traduce in comportamenti volti ad impedire distinzioni produttive di diseguaglianze.

Nel contesto del diritto europeo, il divieto di discriminazioni ha assunto caratteri particolari, dettati dalle specificità delle competenze e delle funzioni di cui l'Unione Europea è investita. Innanzitutto, l'art. 14 CEDU sancisce il divieto di discriminazione: «il godimento dei diritti e delle libertà riconosciuti nella presente Convenzione deve essere assicurato senza discriminazione alcuna, di sesso, di razza, di colore, di lingua, di religione, di opinione politica o di altro genere, di origine nazionale o sociale, di appartenenza a una minoranza nazionale, di ricchezza, di nascita o di altra condizione». A partire dagli anni 2000, all'art. 14 CEDU viene affiancato il Protocollo Addizionale n. 12, il quale riconosce il divieto di discriminazione ancorandolo non più ai soli diritti sanciti dalla Convenzione, ma a tutti i diritti previsti a livello nazionale[4]. Tuttavia, la legittimazione del diritto antidiscriminatorio si avrà solo con la Carta dei diritti fondamentali e specificatamente all'art. 21. Il testo dell'art. 21 dispone, infatti, che «è vietata qualsiasi forma di discriminazione fondata, in particolare, sul sesso, la razza, il colore della pelle o l'origine etnica o sociale, le caratteristiche genetiche, la lingua, la religione o le convinzioni personali, le opinioni politiche o di qualsiasi altra natura, l'appartenenza ad una minoranza nazionale, il patrimonio, la nascita, la disabilità, l'età o l'orientamento sessuale». Il divieto di discriminazione concretamente, quindi, si declina in svariati comportamenti, potenzialmente in grado di

[1] G. Dodaro, *Uguaglianza e diritto penale. Uno studio sulla giurisprudenza costituzionale*, Milano, 2012, 9.

[2] Art. 2, § 2, Dir. 2000/43/CE in materia di discriminazioni per razza e origine etnica; Art. 2, § 2, Dir. 2000/78/CE in materia di discriminazioni per religione, convinzioni personali, handicap, età, tendenze sessuali; Art. 2, § 2, Dir. 2006/54/CE in materia di discriminazioni di genere. Nella disciplina interna di recepimento: art. 2 D.lgs. n. 215/2003 in materia di discriminazioni per razza e origine etnica; art. 2 D.lgs. n. 216/2003 in materia di discriminazioni per religione, convinzioni personali, handicap, età, tendenze sessuali; art. 25 D.lgs. n. 198/2006 in materia di discriminazioni di genere.

[3] S. Rodotà, *Il diritto di avere diritti*, Roma-Bari, 2012, 184 ss.; M. Militello, *Principio di uguaglianza e di non discriminazione tra Costituzione italiana e Carta dei diritti fondamentali dell'Unione Europea*, in *Biblioteca "20 Maggio"*, 2010, 1, 158, «*in virtù del legame esistente tra la tutela antidiscriminatoria e la connotazione sociale e storica dei divieti si ricava una lettura obbligata per cui i divieti di discriminazione, più che sancire l'irrilevanza di determinate qualità soggettive, sono destinati ad impedire che esse si traducano in distinzioni produttive di diseguaglianze*».

[4] Nel Preambolo si parla di un «*principio fondamentale, secondo il quale tutte le persone sono uguali innanzi alla legge e hanno diritto alla stessa protezione da parte della legge*».

Downloaded from www.biodiritto.org.
ISSN 2284-4503

generare una disuguaglianza rilevante. L'elaborazione dell'idea di discriminazione nelle sue diverse accezioni si deve alle Direttive di matrice europea, che hanno consentito di approfondire la nozione di discriminazione distinguendo, al suo interno, tra discriminazioni dirette ed indirette, molestie, molestie sessuali, ritorsioni e ordini di discriminare[5]. In particolare, la discriminazione diretta[6] si realizza in tutte quelle ipotesi in cui un soggetto è trattato meno favorevolmente di quanto sia, sia stata o sarebbe trattata un'altra persona in una situazione analoga. Al contrario, si verifica una discriminazione indiretta[7] se una disposizione o una prassi apparentemente neutra contribuisce a mettere un certo soggetto in una posizione di particolare svantaggio rispetto ad altre persone. Più di recente ai concetti cardine di discriminazione diretta e indiretta si è aggiunto il concetto di discriminazione organizzativa, che, ai sensi del Codice delle Pari Opportunità – d.lgs. 198/2006 – all'art. 25 comma 2*bis* chiarisce che costituisce discriminazione ogni trattamento o modifica dell'organizzazione delle condizioni e dei tempi di lavoro che, in ragione del sesso, dell'età anagrafica, delle esigenze di cura personale o familiare, dello stato di gravidanza nonché di maternità o paternità, anche adottive, ovvero in ragione della titolarità e dell'esercizio dei relativi diritti, pone o può porre il lavoratore in almeno una delle seguenti condizioni: a) posizione di svantaggio rispetto alla generalità degli altri lavoratori; b) limitazione delle opportunità di partecipazione alla vita o alle scelte aziendali; c) limitazione dell'accesso ai meccanismi di avanzamento e di progressione nella carriera.

Tali fenomeni risultano oggi accentuati dall'ingresso della tecnologia nell'agire pubblico e privato, e più segnatamente, dall'impiego di strumenti di intelligenza artificiale (d'ora in avanti IA o AI). I sistemi di intelligenza artificiale stanno contribuendo all'emersione di maggiori criticità a fronte di trattamenti differenti operati non più solo dall'azione umana, ma influenzati altresì dal funzionamento di una macchina[8] capace di utilizzare algoritmi di apprendimento automatico, in grado di analizzare grandi volumi di dati di addestramento per identificare correlazioni, schemi e metadati. Gli algoritmi che permettono all'intelligenza artificiale di funzionare sono algoritmi talvolta artefatti, poiché risentono

[5] F. Amato, *Le nuove direttive comunitarie sul divieto di discriminazione. Riflessioni e prospettive per la realizzazione di una società multietnic*a, in *Lavoro e diritto*, 2003, 127 ss.

[6] CGUE, *causa C-507/18, sentenza della Grande Camera del 23 aprile 2020, su rinvio pregiudiziale relativo al caso NH c. Associazione Avvocatura per i diritti LGBTI – Rete Lenford*: L'Associazione avvocatura per i diritti LGBTI aveva convenuto in giudizio davanti al Tribunale di Bergamo l'avvocato NH per alcune sue dichiarazioni considerate contrarie al divieto di non discriminazione dei lavoratori in base agli orientamenti sessuali. A seguito della decisione del suddetto Tribunale che dichiarava illecite le sue dichiarazioni, l'Avvocato NH ha presentato ricorso giungendo fino alla Corte di Cassazione italiana che ha effettuato il rinvio pregiudiziale. La Corte di giustizia nella sentenza afferma che nella nozione di "condizioni di accesso all'occupazione e al lavoro" rientrano anche le dichiarazioni rese da una persona durante una trasmissione audiovisiva in base alle quali egli non assumerebbe mai nella sua impresa una persona con un determinato orientamento sessuale. Ciò vale anche nel caso in cui in quel momento non sia in corso alcuna selezione, sempre che vi sia un collegamento non ipotetico tra dette dichiarazioni e le condizioni di accesso all'occupazione all'interno di detta impresa.

[7] *Corte di cassazione, sezione lavoro, ordinanza 21 aprile 2020 n.7982*: In tema di requisiti per l'assunzione, sussiste una discriminazione indiretta qualora sia previsto come requisito una statura minima identica per uomini e donne, in contrasto con il principio di uguaglianza, presupponendo erroneamente la non sussistenza della diversità di statura mediamente riscontrabile tra uomini e donne.

[8] L. Floridi, *La quarta rivoluzione. Come l'infosfera sta trasformando il mondo*, Milano, 2017: «L'impatto dell'intelligenza artificiale supera, come noto, la dimensione più circoscritta e propria del fenomeno discriminatorio, tanto da aver indotto autorevole dottrina a coniare l'espressione ormai famosa di realtà «on-life» a enfatizzare come la realtà virtuale si sta oppure si è ormai imposta al fianco di quella materiale e concreta».

Downloaded from www.biodiritto.org.
ISSN 2284-4503

dei significati, dei concetti, delle idee, dei giudizi e dei pregiudizi che l'essere umano apprende sin dalla nascita. Ne deriva che, certe decisioni vengono adottate mediante il supporto di sistemi di IA che sbagliano, perché non sono in grado di profilare attendibilmente per via di dati incompleti, obsoleti o *biased*, di errori nella costruzione degli algoritmi o di limitazioni al loro uso. Si parla, in tali casi, di discriminazioni algoritmiche, perché la determinazione o la pratica che definisce svantaggi per taluni soggetti è adottata o attuata (anche) mediante l'impiego di algoritmi, compresi quelli dell'IA[9].

2. Tipologie di algoritmi e lavoro

L'avvento dell'intelligenza artificiale sta rivoluzionando molti settori della nostra esistenza, tra i quali, in via principale, tutto ciò che attiene al mondo del lavoro. Il tema delle discriminazioni dettate dagli algoritmi, infatti, sta prendendo sempre più spazio nel mondo del lavoro, sia rispetto all'accesso al mercato del lavoro sia in relazione alle condizioni contrattuali. Pensati e scritti dall'essere umano, gli algoritmi si rivelano potenzialmente in grado di riprodurre nella sfera digitale i pregiudizi e gli stereotipi già esistenti nella realtà. Per questo motivo, è fondamentale la qualità dei *dataset* utilizzati, i quali devono essere sufficientemente completi e ampi da non ricreare i pregiudizi e le discriminazioni già presenti nella realtà sociale. Tuttavia, non tutti gli algoritmi operano nello stesso modo, per cui, risulta necessario, anzitutto, comprendere cosa siano gli algoritmi e, perché e come il loro utilizzo possa autorizzare processi decisionali discriminatori. Con riferimento alle differenti tipologie di algoritmi, è possibile distinguere tra Algoritmi *rule-based* e Algoritmi di *machine learning.* Entrambi rientrano all'interno della definizione di AI prevista dal Regolamento sull'IA all'art. 3 e al Cons. n. 12[10]. La distinzione è rilevante soprattutto per capire le modalità e i differenti gradi con cui si possono presentare i

[9] G. Gometz, *Intelligenza artificiale, profilazione e nuove forme di discriminazione,* in *Il lato oscuro della legge* (a cura di F. Mancuso e V. Giordano), *Teoria e storia del diritto privato,* www.teoriaestoriadeldirittoprivato.com

[10] Art. 3 "Definizioni"- Reg. UE 1689/2024 – AI Act: «sistema di IA": un sistema automatizzato progettato per funzionare con livelli di autonomia variabili e che può presentare adattabilità dopo la diffusione e che, per obiettivi espliciti o impliciti, deduce dall'input che riceve come generare output quali previsioni, contenuti, raccomandazioni o decisioni che possono influenzare ambienti fisici o virtuali.».
Cons. 12- Reg. UE 1689/2024 – AI Act: «La nozione di "sistema di IA" di cui al presente regolamento dovrebbe essere definita in maniera chiara e dovrebbe essere strettamente allineata al lavoro delle organizzazioni internazionali che si occupano di IA al fine di garantire la certezza del diritto, agevolare la convergenza internazionale e un'ampia accettazione, prevedendo nel contempo la flessibilità necessaria per agevolare i rapidi sviluppi tecnologici in questo ambito. Inoltre, la definizione dovrebbe essere basata sulle principali caratteristiche dei sistemi di IA, che la distinguono dai tradizionali sistemi software o dagli approcci di programmazione più semplici, e non dovrebbe riguardare i sistemi basati sulle regole definite unicamente da persone fisiche per eseguire operazioni in modo automatico. Una caratteristica fondamentale dei sistemi di IA è la loro capacità inferenziale. Tale capacità inferenziale si riferisce al processo di ottenimento degli output, quali previsioni, contenuti, raccomandazioni o decisioni, che possono influenzare gli ambienti fisici e virtuali e alla capacità dei sistemi di IA di ricavare modelli o algoritmi, o entrambi, da input o dati. Le tecniche che consentono l'inferenza nella costruzione di un sistema di IA comprendono approcci di apprendimento automatico che imparano dai dati come conseguire determinati obiettivi e approcci basati sulla logica e sulla conoscenza che traggono inferenze dalla conoscenza codificata o dalla rappresentazione simbolica del compito da risolvere. La capacità inferenziale di un sistema di IA trascende l'elaborazione di base dei dati consentendo l'apprendimento, il ragionamento o la modellizzazione. Il termine "automatizzato" si riferisce al fatto che il funzionamento dei sistemi di IA prevede l'uso di macchine.»

Downloaded from www.biodiritto.org.
ISSN 2284-4503

rischi per i diritti dei lavoratori, con particolare riguardo alle necessità che si pongono nella costruzione di garanzie di trasparenza ed eliminazione di ogni forma di discriminazione.

Gli algoritmi del primo tipo si qualificano come sistemi basati sulla logica, statici perché si alimentano di una serie di istruzioni fisse e modificabili solo in fase di programmazione e il cui risultato è prevedibile *ex ante*, perché tutte le variabili e i risultati possibili sono già programmati nell'algoritmo. Nel contesto lavoristico, un esempio di tale algoritmo si rinviene nell'algoritmo *Frank*, conosciuto per il suo utilizzo nel caso *Deliveroo*. Tale algoritmo si basava su un sistema di apprendimento automatico, impiegato per valutare le prestazioni dei lavoratori e classificarli sulla base dei parametri di affidabilità e partecipazione e dare, quindi, la precedenza sugli ordini ai migliori in classifica. Sul punto è intervenuto il Tribunale di Bologna[11], il quale ha accertato la natura discriminatoria ex art. 2 d.lgs. 216/2003 della condotta di *Deliveroo* in relazione alle condizioni di accesso alla prenotazione delle sessioni di lavoro, in quanto l'algoritmo non prendeva in considerazione la legittimità delle motivazioni di astensione dal lavoro – quali malattia, stato di necessità o esercizio del diritto di sciopero – e penalizzava ingiustamente i lavoratori tramite un abbassamento delle statistiche, a cui conseguiva una riduzione delle occasioni lavorative e quindi retributive.

Altro esempio di questo tipo, si può ravvisare, nel caso *Foodinho*[12], il quale impiegava una piattaforma, tramite la quale ai *riders* venivano assegnati "punteggi di eccellenza" in ragione di specifici criteri di produttività, tra cui il numero di consegne e la disponibilità nelle fasce orarie ad alta richiesta e nel fine settimana, tenendo conto, inoltre, della mancata presentazione negli slot prenotati. Sulla base di tali punteggi, infatti, la piattaforma consentiva ai lavoratori di scegliere in anticipo la collocazione delle successive prestazioni. Nel caso di specie, il Tribunale di Palermo ha ritenuto sussistente una discriminazione indiretta, in quanto prevedeva l'attribuzione di un punteggio negativo ai *riders* nelle ipotesi di c.d. *late cancellation*, ossia di cancellazione o annullamento della prenotazione di uno slot con un preavviso inferiore alle 24 ore, senza valutare le motivazioni che avevano dato luogo alla cancellazione[13].

Sono, invece, algoritmi del secondo tipo, tutti quegli algoritmi basati su un metodo statistico/probabilistico, dinamici, perché l'insieme delle istruzioni è calcolato nel tempo in modo automatizzato nella fase di addestramento e il cui risultato non è prevedibile *ex ante*, perché deriva da correlazioni di natura probabilistica, ma può essere, almeno teoricamente, compreso e spiegato solo *ex post*. Il risultato ottenuto, però, nella maggior parte dei casi, risulta poco trasparente, poiché pienamente comprensibile solo a un soggetto esperto, con la conseguenza della necessità di un intervento successivo volto a facilitarne la comprensione da parte di un soggetto non esperto[14].

Nel contesto lavorativo, un esempio potrebbe ravvisarsi nel sistema impiegato da *Amazon* per la selezione del personale: l'algoritmo era stato addestrato a stilare la graduatoria dei migliori candidati

[11] Tribunale di Bologna, sez. Lavoro, ordinanza 31 dicembre 2020, (discriminazione algoritmica di lavoratori), https://giurcost.org/casi_scelti/GM/TribunaleBologna31dicembre2020.pdf

[12] Tribunale di Palermo, sez. Lavoro, sentenza 17 novembre 2023, https://onelegale.wolterskluwer.it/document/10SE0002795397

[13] A. Perulli, *La discriminazione algoritmica: brevi note introduttive a margine dell'Ordinanza del Tribunale di Bologna*, in *Lavoro Diritti Europa*, 1, 2021, 2.

[14] M. Peruzzi, *Il diritto antidiscriminatorio al test di intelligenza artificiale*, in *Lab. & Law Issues*, 2021, 1, pp. 50 ss.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

per posizioni di ingegneri informatici osservando i *curricula* ricevuti nei dieci anni precedenti. La maggior parte degli stessi proveniva da uomini, maggiormente impiegati in settori tecnologici. Sulla base di queste statistiche, l'algoritmo aveva "insegnato" a sé stesso a preferire i candidati uomini rispetto alle candidate donne, nel senso che, pur essendo stato addestrato a non utilizzare direttamente il sesso come criterio selettivo, era riuscito a riconoscerlo da altre informazioni, comunque presenti nei *curricula*, utilizzando poi questi indici di genere come criteri utili ad effettuare la selezione. In alcuni casi, veniva favorito, addirittura, chi utilizzava alcuni termini - come verbi in forma attiva - che, nel campione storico di *curricula* alla base del modello decisionale algoritmico, erano statisticamente usati più dagli uomini che dalle donne.

Nel mercato del lavoro automatizzato, infatti, si riproducono potenzialmente gli stessi atteggiamenti discriminatori che si riscontrano nei lavori tradizionali, con riguardo a tutti i fattori di discriminazione, poiché le menti che programmano gli algoritmi sono menti umane[15]. I sistemi di IA discriminano, infatti, non perché il sistema sia di per sé maligno, ma perché eredita comportamenti sbagliati che poi ripete. Ciò che è noto, è che le discriminazioni algoritmiche assumono oggi una portata drasticamente pervasiva, capace di determinare conseguenze distruttive sulla società. Simili algoritmi, infatti, se guidati da dati imprecisi, parziali o non rappresentativi del fenomeno a cui si applicano, possono produrre risultati non trasparenti e distorti e condurre, perciò, a varie forme di discriminazione. Come affermato rispettivamente dai Tribunali di Bologna[16] e di Palermo[17], l'incoscienza della macchina e l'applicazione indifferenziata dei parametri di valutazione non giustificano la discriminazione, proprio per il particolare svantaggio che tali parametri implicano nei confronti dei portatori di determinati fattori di rischio, quali anzitutto l'affiliazione sindacale, il genere, la religione e la disabilità.

3. Fattori scatenanti la discriminazione

Gli algoritmi funzionano secondo la logica *garbage in – garbage out*, secondo cui dati incongrui, inesatti o non aggiornati possono generare solamente risultati decisionali inaffidabili. A differenza dei tradizionali sistemi informatici, l'AI, infatti, non si limita ad eseguire istruzioni predefinite, ma impara e genera contenuti basandosi sui dati forniti. Per questo, alimentare un sistema AI con dati non accurati porta a risultati imprevedibili e potenzialmente rischiosi.

Il primo fattore rilevante che contribuisce al perpetuarsi delle discriminazioni algoritmiche è sicuramente rappresentato dalla componente umana. Sebbene gli algoritmi operino sempre di più in modo autonomo, il ruolo degli esseri umani resta comunque cruciale per il loro sviluppo e funzionamento. È la persona, immersa in una realtà non scevra di pregiudizi, a fornire i dati alla macchina e, pure a fronte di tecniche che palesano un livello di progressiva autonomia dalla decisione o programmazione originaria, esse rimangono pur sempre il prodotto di un'azione umana, non sempre imparziale. Non a caso, infatti, l'AI Act[18] adotta un approccio basato sul rischio: maggiore è il rischio che l'applica-

[15] C. Alessi, *Lavoro tramite piattaforma e divieti di discriminazione nell'UE*, in C. Alessi, M. Barbera, L. Guaglianone (a cura di) *Impresa, lavoro e non lavoro nell'impresa digitale*, Bari, 2019, 663 ss.

[16] Tribunale di Bologna, sez. Lavoro, ordinanza 31 dicembre 2020.

[17] Tribunale di Palermo, sez. Lavoro, sentenza 17 novembre 2023.

[18] https://artificialintelligenceact.eu/

Downloaded from www.biodiritto.org.
ISSN 2284-4503

zione dell'AI può causare per i diritti e le libertà fondamentali degli interessati, più rigidi sono gli obblighi di sicurezza e trasparenza previsti sia in capo ai produttori che agli utilizzatori. L'AI Act classifica i sistemi di AI secondo quattro livelli di rischio e associa ad ognuno di essi delle salvaguardie che ne compensino la pericolosità. In particolare, vengono individuati sistemi a rischio inaccettabile, in tale ipotesi vi rientrano i sistemi o le applicazioni di IA che influenzano in maniera significativa gli utenti, distorcendone il comportamento, mediante tecniche manipolative, ingannevoli e/o sfruttandone le diversità e vulnerabilità. Pratiche di questo genere possono causare una lesione dei diritti fondamentali delle persone e, di conseguenza, lo sviluppo e la diffusione degli stessi è vietata all'interno dell'UE; sistemi a rischio alto, si fa riferimento a tutti quei sistemi appositamente identificati[19], che potrebbero comportare delle conseguenze sulla salute, sulla sicurezza o sui diritti fondamentali delle persone, e che per essere ammessi all'interno dell'UE, dovranno soddisfare requisiti rigorosi previsti dagli artt. 8 - 49 dell'AI Act. Per questa particolare tipologia di sistemi, si prevede che la sorveglianza umana debba essere affidata a persone fisiche differenti, allo scopo di verificare in che misura il sistema elaborato rischia di incidere sui diritti fondamentali dei cittadini, procurandone anche potenzialmente discriminazioni. L'art. 14 AI Act, nello specifico, teorizza il principio etico-giuridico di *Human-In-The-Loop*[20] e cioè quel principio in base al quale è necessario garantire che gli individui siano informati delle scelte che li riguardano e dell'impiego di strumenti di AI e più nel dettaglio è quella metodologia secondo cui gli esseri umani etichettano i dati, il che aiuta il modello a ottenere dati di addestramento di alta qualità e in quantità elevata. Il modello si basa sul connubio tra le capacità delle macchine e l'intelligenza umana, le quali attraverso varie interazioni contribuiscono ad alimentare il modello di apprendimento automatico. In altre parole, l'intelligenza umana dovrebbe intervenire quando la macchina ha difficoltà a risolvere un problema.

Secondo il Gruppo di Esperti[21], i sistemi di AI dovrebbero essere progettati per aumentare, integrare e potenziare le abilità cognitive, sociali e culturali umane, senza sostituirsi completamente ad esso. L'apporto del supervisore umano sulla macchina può manifestarsi in varie forme, non solo come previsto dal c.d. *Human In The Loop*, ma altresì nelle sue declinazioni di *Human On The Loop* che assicura un controllo minimo, in fase di progettazione o di monitoraggio e *Human In Command*, che consente un monitoraggio costante sul sistema e sui suoi effetti, lasciando ampia discrezionalità al supervisore con la conseguenza, però, che il supervisore umano potrebbe decidere di ignorare la decisione

[19] Capo III "Sistemi di IA ad alto rischio" - Reg. UE 1689/2024- artt. 6 e ss. Si definisce ad alto rischio se sono soddisfatte entrambe le condizioni seguenti: a) il sistema di IA è destinato ad essere utilizzato come componente di sicurezza di un prodotto, o il sistema di IA è esso stesso un prodotto, disciplinato dalla normativa di armonizzazione dell'Unione elencata nell'Allegato I dell'AI Act; b) il prodotto, il cui componente di sicurezza a norma della lett. a) è il sistema di IA, o il sistema di IA stesso in quanto prodotto, è soggetto a una valutazione della conformità da parte di terzi ai fini dell'immissione sul mercato o della messa in servizio di tale prodotto ai sensi della normativa di armonizzazione dell'Unione elencata nell'Allegato I. Oltre ai sistemi di IA ad alto rischio di cui sopra, sono considerati ad alto rischio anche i sistemi di IA di cui all'Allegato III dell'AI Act.

[20] Art. 14 par. 2 "*Sorveglianza Umana*" «La sorveglianza umana mira a prevenire o ridurre al minimo i rischi per la salute, la sicurezza o i diritti fondamentali che possono emergere quando un sistema di IA ad alto rischio è utilizzato conformemente alla sua finalità prevista o in condizioni di uso improprio ragionevolmente prevedibile, in particolare qualora tali rischi persistano nonostante l'applicazione di altri requisiti di cui alla presente sezione».

[21] High-Level Expert Group on AI, *Ethics guidelines for trustworthy AI*, https://digitalstrategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

Downloaded from www.biodiritto.org.
ISSN 2284-4503

assunta mediante l'AI. Non a caso, il paragrafo 4, lettera b) dell'art. 14 AI Act stabilisce che le persone, preposte alla sorveglianza, devono restare consapevoli «della possibile tendenza a fare automaticamente affidamento o a fare eccessivo affidamento sull'output prodotto da un sistema di IA ad alto rischio ("*automation bias*[22]"), in particolare per i sistemi di IA ad alto rischio utilizzati per fornire informazioni o raccomandazioni per le decisioni che devono essere prese da persone fisiche». Da qui l'esigenza dell'intervento di un'altra persona fisica, in grado di verificare la logicità e la legittimità delle scelte dettate in fase di input e degli esiti e intervenire sulla decisione automatizzata[23].

Le altre due tipologie di rischio vengono individuate in sistemi di AI a rischio di trasparenza, cioè quei sistemi destinati ad interagire direttamente con le persone fisiche, o che generano o manipolano immagini, contenuti, audio o video, e che possono comportare specifici rischi di furti di identità, manipolazioni o inganni (es. *chatbots* o *deep fakes*). Per queste ipotesi, sono previsti specifici obblighi di informazione e trasparenza. Gli utenti devono essere informati quando interagiscono con un sistema di Intelligenza Artificiale o quando un contenuto è stato generato da un'IA, al fine di consentire loro di prendere decisioni informate e interagire con la tecnologia in modo consapevole.

I sistemi a rischio minimo, infine, sono quei sistemi che presentano rischi minimi o nulli per i diritti e/o la sicurezza dei cittadini. Tali sistemi sono attualmente esenti da obblighi specifici, tuttavia, potranno essere introdotti dei codici di buone pratiche.

Altro fattore produttivo di discriminazione si rinviene nella qualità e nella difficile comprensibilità dei dati utilizzati nel processo decisionale algoritmico che, se non quantitativamente o qualitativamente adeguati, possono viziarlo e replicare discriminazioni insite nel loro stesso processo di programmazione e addestramento. L'opacità o l'effetto scatola nera, infatti, rende difficoltoso determinare dove si trova la radice della discriminazione: sistemi decisionali automatizzati o software semiautomatizzati possono contenere *bias* non intenzionali introdotti da loro programmatori, o, se intenzionali, possono essere nascosti o mascherati in un codice molto complesso e di non facile comprensione. Motivo per cui, diviene più difficile per le vittime di tali discriminazioni rendersi conto delle stesse. L'art. 13 dell'AI Act, al fine di scongiurare questi pericoli, si concentra sugli aspetti di trasparenza e sulle informazioni da procurare dai fornitori specialmente di sistemi di IA definiti ad alto rischio. L'AI Act sviluppa tre concetti fondamentali: trasparenza, interpretabilità e spiegabilità, al fine di evitare qualsiasi forma di discriminazione e rendere più facilmente comprensibile il funzionamento del sistema algoritmico impiegato. Per trasparenza si intende che i sistemi di AI devono esser progettati e sviluppati in modo tale da rendere i fornitori in grado di interpretare l'output del sistema e usarlo in maniera appropriata. L'interpretabilità si riferisce alla capacità di una persona di comprendere il funzionamento interno di un sistema di IA e nello specifico si traduce in un funzionamento del modello sufficientemente trasparente per permettere agli utenti di discernere come gli input siano trasformati in output. La spiegabilità, invece, si concentra sulla capacità di articolare gli esiti di un sistema di IA in termini comprensibili all'uomo. Può avvalersi dell'impiego di strumenti e metodi supplementari volti a fornire chiarimenti su come il sistema di IA giunga a determinate decisioni con l'obiettivo di colmare

[22] Per *automation bias* si intende la tendenza da parte dell'essere umano, coinvolto nell'interazione con la macchina, ad affidarsi ai suoi output, fino a trascurare o ignorare altre informazioni che derivano da fonti diverse.

[23] G. Lo Sapio, *La trasparenza sul banco di prova dei modelli algoritmici*, in *Federalismi.it*, 11, 2021, 242 ss.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

il divario tra la complessità dell'IA e la comprensione umana, consentendo agli utenti di apprendere le motivazioni dietro le decisioni dell'IA.
Infine, un ultimo elemento produttivo di discriminazione sarebbe ravvisabile, almeno nei c.d. *algoritmi machine learning*, nell'utilizzo di un *proxy*, un indicatore statistico di un'altra caratteristica a cui vengono ricollegati effetti sfavorevoli, che risulta, però, più difficilmente percepibile anche da chi programma ed eventualmente supervisiona il funzionamento dell'algoritmo stesso.

4. Proxy discriminations di genere

Un *proxy* è un elemento utilizzato da un sistema di intelligenza artificiale per fare distinzioni tra individui e/o gruppi sociali. La *proxy discrimination* – letteralmente discriminazione "per delega" – è da intendersi quale discriminazione provocata da un criterio apparentemente neutro, basato su una caratteristica (*proxy*) strettamente collegata ad un fattore protetto dalla normativa antidiscriminatoria. La discriminazione si verifica, quindi, ogni volta che il *proxy* perpetua pregiudizi influenzando in modo negativo determinati individui e gruppi, senza fondare la distinzione sui fattori classici della discriminazione, ma, piuttosto, basandosi su loro correlazioni presumibilmente non discriminatorie. Già nel caso *Coleman* del 2008, una madre aveva sostenuto di aver subito un trattamento discriminatorio sul posto di lavoro in ragione del fatto che il figlio fosse disabile. I giudici di Lussemburgo avevano rilevato che la tutela offerta dalla normativa dell'Unione rispetto al motivo della disabilità non andava riferita esclusivamente al figlio, ma poteva essere estesa anche alla madre, poiché il trattamento discriminatorio da essa subito era stato comunque posto in essere in ragione di quella disabilità[24]. In tali casi, il problema principale si rinviene nella prova della discriminazione: se la lavoratrice, ad esempio, riesce a dimostrare che un determinato criterio – basato su un certo *proxy* – adottato dal datore di lavoro produce un effetto pregiudizievole su tutti i lavoratori appartenenti ad una determinata categoria protetta, il giudice potrà anche dedurre l'esistenza di una discriminazione[25].
Tuttavia, valutare questi elementi non è così semplice, poiché gli utenti non sempre hanno gli strumenti per comprendere le modalità e i dati statistici circa i gruppi di utenti raggiunti o esclusi da un determinato trattamento. Invero, oggi sia le pubbliche amministrazioni che le aziende private stanno impiegando sempre più frequentemente algoritmi progettati per aiutare o sostituire le persone incaricate di prendere decisioni. Sistemi informatici costruiti, però, sulla base di pregiudizi discriminano sistematicamente e ingiustamente, negando opportunità o beni ovvero attribuendo un risultato indesiderato sulla base di motivazioni irragionevoli o inappropriate[26]. Le decisioni basate su algoritmi non sufficientemente trasparenti e addestrati possono avere un impatto di vario tipo sui diritti umani e le

[24] Corte di giustizia (Grande Sezione), sentenza del 17 luglio 2008, *causa C-303/06, Coleman*, EU:C:2008:415.

[25] A.E.R. Prince, D. Schwarcz, *Proxy Discrimination in the Age of Artificial Intelligence and Big Data*, in *Iowa Law Review*, 2020, 105.

[26] B. Friedman, H. Nissembaum, *Bias in Computer Systems*, ACM Transactions on Information Systems, 3/1996, 332: «Accordingly, we use the term bias to refer to computer systems that systematically and unfairly discriminate against certain individuals or groups of individuals in favor of others. A system discriminates unfairly if it denies an opportunity or a good or if it assigns an undesirable outcome to an individual or group of individuals on grounds that are unreasonable or inappropriate».

Downloaded from www.biodiritto.org.
ISSN 2284-4503

libertà fondamentali, in particolare – ai fini della trattazione - sugli stereotipi di genere[27]. Le vittime delle discriminazioni algoritmiche risultano essere in prevalenza donne. Da sempre, le donne sono una delle categorie di individui vulnerabili che subiscono l'impatto delle trasformazioni sociali. Secondo i risultati del Global Gender Gap Report 2024 del World Economic Forum[28], le donne stanno scontando a caro prezzo gli squilibri sistemici del mercato del lavoro. Questi squilibri non solo significano che ci sono meno donne in ruoli di *leadership*, ma anche che quando ci sono *shock* economici, le donne sono più colpite.

L'intelligenza artificiale non rispecchia egualmente entrambi i sessi e, anzi, si traduce, talvolta, in una discriminazione strutturale ai danni delle donne. Tale affermazione è supportata da analisi statistiche che convergono nell'attestare la natura non *gender-neutral* dei modelli di AI impiegati nella grande maggioranza delle attività[29]. In particolare, in tale ipotesi si realizza una *proxy discrimination*[30], poiché il problema nasce dalla presenza nei *data-sets* di *redundant encodings*, ovverosia si cerca di mascherare l'appartenenza ad un determinato sesso o stato (stato di gravidanza o di invalidità) in altri dati associati alla medesima categoria. Invero, come noto, le tecniche di intelligenza artificiale funzionano sulla base delle associazioni tra dati, per cui la macchina riesce a selezionare tutti quegli elementi che consentono di raggiungere più facilmente il risultato desiderato anche ricorrendo ad altri fattori capaci di determinare, direttamente o indirettamente, l'affiliazione ad una categoria protetta.

Il caso *Amazon* - sopra esposto - è considerato una pietra miliare delle problematiche connesse all'*inclusive recruitment,* dal momento che ha consentito di denunciare – forse per la prima volta - la mancata inclusione di persone diverse tra loro all'interno dei team, nel rispetto delle pari opportunità[31]. Eppure, appare opportuno ricordare, che la stessa Unione europea ha posto la parità tra uomo e donna tra i suoi principi fondanti e come guida per i lavori dell'Eurofound[32], soprattutto per ciò che attiene alla parità di genere in ogni ambito lavorativo, inclusa la parità retributiva. Esemplare, ancora, è il caso di LinkedIn, secondo cui gli algoritmi impiegati dai motori di pubblicazione di annunci di lavoro e, quindi, utilizzati per abbinare i candidati alle rispettive opportunità, producevano risultati distorti, privilegiando candidati uomini rispetto alle donne: gli uomini risultavano maggiormente propensi a cercare nuove opportunità rispetto alle donne. Le principali società di reclutamento online abbinano candidati qualificati con le posizioni disponibili. Molte piattaforme, però, per pianificare il *matching* tra le posizioni disponibili e i candidati, utilizzano algoritmi c.d. di raccomandazione, che elaborano le informazioni ricevute dalle organizzazioni e da chi cerca lavoro per stilare un elenco di soggetti in li-

[27] M. D'AMICO, *Una parità ambigua. Costituzione e diritti delle donne*, Milano, 2020.

[28] https://www.weforum.org/publications/global-gender-gap-report-2024/.

[29] Si richiamano il *Progetto Gender Shades* sul carattere discriminatorio ai danni delle donne, in particolare afroamericane di alcuni sistemi di riconoscimento facciale (su cui, anche,); la vicenda di Amazon in tema di reclutamento, su cui J. DUSTIN, *Amazon scraps secret AI recruiting tool that showed bias against wome*n, in *Reuters*, 11 ottobre 2018; J. LAURET, *Amazon's sexist AI recruiting tool: how did it go so wrong?,* in *Becominghuman.ai*, 16 agosto 2019.

[30] A.E.R. PRINCE, D. SCHWARCZ, cit.

[31] C. DELLA GIUSTINA, *Quando il datore di lavoro diviene un algoritmo: la trasformazione del potere del datore di lavoro in algocrazia. Quale spazio per l'applicazione dei principi costituzionali*?, in *Media Laws,* 2021, 2.

[32] https://european-union.europa.eu/institutions-law-budget/institutions-and-bodies/search-all-eu-institutions-and-bodies/eurofound_it.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

nea con la posizione lavorativa ricercata, «raccomandando» appunto solamente determinate categorie di soggetti, in prevalenza uomini. Anche, il rapporto dell'Unesco sugli impatti dell'IA nella vita lavorativa delle donne[33] denuncia, ad esempio, la riduzione della capacità delle donne africane di accedere al credito a causa dell'uso di sistemi di *credit scoring* che valutano l'impronta digitale di un individuo. Le differenze nell'uso e nell'accesso delle donne africane a Internet - il cosiddetto *digital divide* – diventano, in questo senso, un fattore discriminante rispetto alla possibilità di ottenere finanziamenti, che potrebbero, invece, rivelarsi di grande utilità per il riscatto di queste donne.

Le discriminazioni algoritmiche di genere, infatti - come sopra accennato- non si limitano solo all'ambito lavorativo: si pensi ai motori di ricerca come Google che associano la parola «infermiera» a una donna e la parola «dottore» a un uomo, confinando determinate categorie di lavori e di potere in capo ai solo soggetti uomini. Questi episodi di discriminazione da parte degli strumenti di intelligenza artificiale sono accompagnati dal rischio costante per le donne di essere esposte a violenza online, *cyber-stalking* o bullismo. Secondo quanto denunciato dal rapporto dell'Unesco[34], la tecnologia vocale con voce femminile dà troppo spesso risposte dal tono sottomesso rispetto alle domande. La maggior parte degli assistenti vocali presenta una voce femminile, trasmettendo un segnale di donne garanti, docili e desiderose di aiutare, sempre disponibili al solo e semplice tocco di un pulsante o con un comando vocale. La sottomissione con cui interagiscono, influenza il modo in cui la gente reagisce alle voci femminili e come le donne rispondono alle richieste e si esprimono. Questo rafforza i pregiudizi di genere e restituisce un'immagine di donna sottomessa e tollerante nei confronti di trattamenti inadeguati. Il rischio è che questa scelta progettuale perpetui uno stereotipo discriminatorio e trasmetta il messaggio, soprattutto alle generazioni più giovani, che alle donne vada attribuito un ruolo di subordinazione e incondizionata disponibilità, creando i presupposti per un'ulteriore violenza di genere.

5. Donne e molestie online: deep fake, cyberstalking, deep nude e revenge porn

Esistono diverse forme di violenza virtuale contro le donne, fra cui *cyberstalking*, pornografia non consensuale, molestie basate sul genere, stigmatizzazione a sfondo sessuale, stupro e minacce di morte, pubblicazione online di informazioni personali e private. La violenza virtuale contro le donne può manifestarsi come violenza sessuale, psicologica ed economica, in cui l'attuale o futura occupazione della vittima potrebbe esser compromessa da informazioni pubblicate *online*.

A seguito dell'avvento delle nuove tecnologie, lo *stalking* è oggi un fenomeno ancora più insidioso e invasivo: si parla di *cyberstalking* quando mediante l'uso di dispositivi di comunicazione elettronica si intende molestare un'altra persona. I comportamenti dello *stalker* che potrebbero connotare l'attività di *cyberstalking* sono legati: alla sorveglianza *online* nei confronti della vittima, mediante attivazione delle funzioni di geo- localizzazione; alla ricerca di contatto, attraverso pedinamento elettronico e aggiramento compiuto anche collegandosi ad amicizie sui *social network;* ad un controllo, ad esempio della posta elettronica o dei suoi profili social o conto correnti bancari, spesso all'insaputa della vittima. In Italia il *cyberstalking* ricade nella fattispecie disciplinata dall'art. 612-bis c.p. che regola-

[33] The effects of AI on the working lives of women, https://unesdoc.unesco.org/ark:/48223/pf0000380861

[34] https://www.unesco.org/en/forum-against-racism-discrimination.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

menta il delitto di *stalking*, il quale sancisce che è punito con la reclusione da un anno a sei anni e sei mesi «chiunque, con condotte reiterate, minaccia o molesta taluno in modo da cagionare un perdurante e grave stato di ansia o di paura ovvero da ingenerare un fondato timore per l'incolumità propria o di un prossimo congiunto o di persona al medesimo legata da relazione affettiva ovvero da costringere lo stesso ad alterare le proprie abitudini di vita». Con riferimento specifico al *cyberstalking* viene in rilievo l'aggravante prevista dal secondo comma che contempla un aumento della pena allorché «il fatto è commesso attraverso strumenti informatici o telematici», incluso Whatsapp come chiarito dalla Corte di Cassazione[35].

Per *deepfake* – secondo la definizione offerta dal Garante per la protezione dei dati personali[36] – si intendono tutte quelle foto, video e audio creati grazie a software di intelligenza artificiale che, partendo da contenuti reali quali immagini e audio, riescono a modificare o ricreare, in modo estremamente realistico, le caratteristiche e i movimenti di un volto o di un corpo e ad imitare fedelmente una determinata voce. In particolare, il legame tra *deepfake* e furto d'identità è molto stretto. Il reato di furto d'identità è regolato dall'art. 494 c.p. secondo cui «chiunque, al fine di procurare a sé o ad altri un vantaggio o di arrecare ad altri un danno, induce taluno in errore, sostituendo la propria all'altrui persona, o attribuendo a sé o ad altri un falso nome, o un falso stato, ovvero una qualità a cui la legge attribuisce effetti giuridici, è punito, se il fatto non costituisce un altro delitto contro la fede pubblica, con la reclusione fino a un anno[37]».

Il furto d'identità assume una gravità particolare quando si collega al contesto sessuale. La produzione di immagini fittizie con connotazioni sessuali è comunemente denominata *deepnude.* Il *deepnude* è una tecnica che permette di manipolare e spogliare artificialmente le figure femminili - sembra funzionare solo con queste - trasformandole in foto di nudo in relazione alla corporatura del soggetto. La semplicità con cui tali software possono essere installati e utilizzati è stata posta in rilievo nel caso di cronaca che coinvolgeva un gruppo di studenti di una scuola media di Latina. Gli studenti mediante l'impiego di un'applicazione denominata "*BikiniOff*", avevano posto in essere la manipolazione di fotografe di cinque studentesse e di una docente. Il *deepnude* della docente, nel caso specifico, era risultato così convincente da comparire su rinomati siti pornografici[38]. Nonostante, infatti, le immagini siano elaborate artificialmente, è innegabile che queste - considerato quanto esse si presentano realistiche - possano intaccare la dignità di una persona che si ritrovi a sua insaputa letteralmente spogliata sul web. Al momento nel nostro ordinamento non esiste una specifica tutela. In verità, sul punto era stata anche presentata nel 2021 una proposta di Legge, mai portata a compimento, volta a contrastare tale fenomeno con l'obiettivo di introdurre un ulteriore comma all'art. 612 del codice penale, finalizzato a prevedere una multa e la reclusione da due a sette anni per chi «invia, cede, pub-

[35] Cassazione penale, sez. V, sentenza 28/01/2019 n° 3989, https://www.altalex.com/documents/news/2019/02/07/stalking

[36] Garante per la Protezione dei Dati Personali "*Deepfake: dal Garante una scheda informativa sui rischi dell'uso malevolo di questa nuova tecnologia*", https://garanteprivacy.it/home/docweb/-/docweb-display/docweb/9512278.

[37] Il furto d'identità non è il solo reato perpetrabile creando, usando o condividendo un *deepfake*. Ad esempio, se il contenuto d*eepfake* va a ledere anche la reputazione dell'individuo, al reato di furto d'identità si aggiunge quello di diffamazione (art. 595 c.p.).

[38] D. Barbera, T*utti i rischi di usare Bikinioff, il chatbot che spoglia le donne*, 19 aprile 2023.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

blica o diffonde immagini manipolate di nudo appartenenti a persone fisiche riconoscibili, attraverso l'utilizzo di strumenti tecnologici e di applicazioni, allo scopo di trarre in inganno l'osservatore.»

Sempre legato al profilo sessuale, si registra un fenomeno denominato *Porno Deepfake*: una tecnica, rientrante nell'idea di *Revenge porn*[39], che attraverso l'uso dell'intelligenza artificiale rielabora immagini o video ritraenti persone reali al fine di trasformarli in materiali multimediali a carattere pornografico, falso ma altamente realistico[40]. Tali prodotti manipolati sono poi diffusi *online* attraverso i siti porno, i social network e le app di messaggistica istantanea. In tale ipotesi, però, l'ordinamento italiano prevede una tutela, in quanto il *Revenge porn* è considerato reato ai sensi dell'art. 612 ter del codice penale consistente nella diffusione in rete di immagini sessualmente esplicite, senza il consenso della persona raffigurata. La vittima è solitamente una donna, mentre il reato viene realizzato spesso dagli ex partner mediante la diffusione di video o immagini. La Cassazione ha recentemente avuto modo di precisare che per configurare il reato in questione, la divulgazione può riguardare non solo immagini o video che ritraggono atti sessuali ovvero organi genitali, ma anche altre parti erogene del corpo umano in condizioni e contesti tali da evocarne la sessualità[41].

Nel 2023, ad esempio, *The Guardian*[42] ha pubblicato un'indagine approfondita, che descrive come molti algoritmi presentino dei *bias* di genere che comporterebbero la diffusione di innumerevoli foto con corpi femminili. Le foto di donne vengono classificate come più spinte o sessualmente suggestive rispetto a foto analoghe di uomini. Anche i pancioni delle donne incinte diventano problematici per questi strumenti di Intelligenza Artificiale, in quanto ad esempio l'algoritmo di Google ha valutato la foto come molto probabile che contenga contenuti scabrosi e quello di Microsoft era convinto al 90% che l'immagine fosse di natura sessualmente suggestiva. Proprio allo scopo di limitare la diffusione non consensuale di tali contenuti pornografici e a tutela della dignità delle donne, il Garante per la protezione dei dati personali già nel 2022 con cinque Provvedimenti[43], aveva comminato in via d'urgenza a Facebook, Instagram e Google di adottare immediatamente tutte le misure necessarie ad impedire la diffusione sulle relative piattaforme del materiale (video, foto) segnalato all'Ufficio del Garante da alcune persone che ne temevano la messa *online*.

[39] Si tratta di un fenomeno della pornografia non consensuale, consiste nella diffusione di immagini pornografiche o sessualmente esplicite a scopo vendicativo (ad esempio per "punire" l'ex partner che ha deciso di porre fine ad una relazione) o per denigrare pubblicamente, bullizzare e molestare la persona cui si riferiscono.

[40] G. NATALE, *Intelligenza artificiale, neuroscienze, algoritmi. aggiornato al nuovo Regolamento Europeo AI Act*, Pisa, 2024, 243.

[41] Cass. pen., Sez. V, sent. n. 14927 del 22 febbraio 2023, https://www.sistemapenale.it/pdf_contenuti/1696488184_sentenza-612-ter-oscurata.pdf

[42] https://www.theguardian.com/technology/2023/feb/08/biased-ai-algorithms-racy-women-bodies

[43] https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9775414; https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9775327; https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9775401; https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9775948; https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9775932.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

6. Prevenzione delle discriminazioni algoritmiche

L'intelligenza artificiale si sta rivelando uno strumento diretto a rendere più acuta le vulnerabilità, ai fini della trattazione delle donne, e di conseguenza del divario di genere, producendo questi effetti su scala globale. Eppure, il ricorso a procedimenti automatizzati o semi automatizzati potrebbe di per sé migliorare sia la prevenzione che la repressione delle discriminazioni in ambito lavorativo e personale. Tuttavia, ciò è possibile solo in presenza di regole utili a far emergere e arginare il rischio discriminazione. È auspicabile, infatti, grazie alle prescrizioni contenute all'interno dell'AI Act, che l'utilizzo di questi sistemi venga subordinato al rispetto delle garanzie e dei limiti stabiliti dalle vigenti disposizioni in materia di protezione delle persone fisiche, anche riguardo al trattamento dei dati personali e, più in generale, che venga operato un congruo bilanciamento tra gli interessi coinvolti.

Nel dicembre 2020, il Consiglio d'Europa e il *Committee on Artificial Intelligence* (CAHAI) hanno adottato il documento *Feasibility study on a legal framework on AI design, development and application based on CoE standards*[44], al fine di dare seguito alle criticità emerse circa le specificità della *AI-derived discrimination*. Invero, l'art. 4 lett. m) del Regolamento allegato al Quadro relativo agli aspetti etici dell'intelligenza artificiale, della robotica e delle tecnologie correlate di cui alla Risoluzione del Parlamento europeo del 20 ottobre 2020 definiva già la discriminazione, come «qualsiasi trattamento differenziato di una persona o di un gruppo di persone per un motivo privo di giustificazione obiettiva o ragionevole e, pertanto, vietato dal diritto dell'Unione». Impostazione ripresa anche dall'AI Act, il quale affronta il problema del possibile utilizzo degli algoritmi con finalità discriminatorie prevedendo una differenziazione degli obblighi basata sul criterio del rischio, come sopra illustrato. L'obiettivo è minimizzare il rischio di discriminazione algoritmica, nel rispetto di quanto previsto all'art. 21 della Carta dei diritti fondamentali dell'Unione Europea. Il Regolamento fa proprio, quindi, il principio di equità e non discriminazione, sostenendo che le organizzazioni sono tenute a garantire che i loro sistemi non pregiudichino o perpetuino discriminazioni basate su caratteristiche personali quali sesso, genere, razza e/o origine etnica. Invero, al fine di evitare simili distorsioni, l'AI Act riprende, in parte, quel *risk approach* tipico del Regolamento Europeo n. 679/2016 in materia di protezione dei dati personali (d'ora in avanti GDPR). L'attenzione al rischio di discriminazione algoritmica emerge già dal Considerando 71 GDPR, ove si legge che il titolare del trattamento dei dati deve garantirne la sicurezza e impedire effetti discriminatori nei confronti di persone fisiche sulla base della razza o dell'origine etnica, delle opinioni politiche, della religione o delle convinzioni personali, dell'appartenenza sindacale, dello status genetico, dello stato di salute o dell'orientamento sessuale. Sempre al fine di proteggere le persone dalla discriminazione, l'art. 22 GDPR vieta determinate decisioni completamente automatizzate con effetti significativi. Si legge, infatti, al primo comma, che «l'interessato ha il diritto di non essere sottoposto a una decisione basata unicamente sul trattamento automatizzato, compresa la profilazione, che produca effetti giuridici che lo riguardano o che incida in modo analogo significativamente sulla sua persona». Nei casi nei quali il trattamento automatizzato è, invece, consentito in base al secondo comma dell'art. 22[45], il titolare del trattamento è tenuto ad attuare misure appro-

[44] Il testo integrale dello studio può essere letto al link: https://rm.coe.int/cahai-2020-23-final-engfeasibility-study-/1680a0c6da.

[45] Art. 22 comma 2 – GDPR - «Il paragrafo 1 non si applica nel caso in cui la decisione: a) sia necessaria per la conclusione o l'esecuzione di un contratto tra l'interessato e un titolare del trattamento; b) sia autorizzata dal

Downloaded from www.biodiritto.org.
ISSN 2284-4503

priate per tutelare i diritti, le libertà e i legittimi interessi dell'interessato, garantendo almeno il diritto di ottenere l'intervento umano da parte del titolare del trattamento, il diritto di esprimere la propria opinione insieme al diritto di contestarne la decisione. Non a caso, il GDPR reca in sé i principi fondamentali di legalità algoritmica, quali principio di non esclusività della decisione algoritmica, di conoscibilità e di non discriminazione algoritmica.

Il principio di non esclusività della decisione algoritmica stabilisce che nelle ipotesi in cui una decisione algoritmica produca effetti giuridici o incida significativamente sulla persona, l'interessato ha il diritto che questa non sia basata unicamente su un trattamento automatizzato, ivi compresa la profilazione, ma deve comunque sempre essere garantito un intervento umano. In tal modo, il soggetto titolare della decisione, pur potendo avvantaggiarsi dello strumento informatico idoneo a fornirgli la soluzione apparentemente migliore, mantiene il controllo della decisione.

Il principio di conoscibilità prevede che ognuno ha il diritto di conoscere l'esistenza di processi decisionali automatizzati, che lo riguardino, ai sensi dell'art. 15, comma 1, lett. h). Il principio di conoscibilità è strettamente correlato al principio di comprensibilità, secondo cui l'interessato ha anche il diritto di ottenere «informazioni significative sulla logica utilizzata». Nell'attribuire un rilievo centrale al principio di conoscibilità dell'algoritmo, la giurisprudenza assume come riferimento normativo gli artt. 13, comma 2, lett. f), e 14, comma 2, lett. g), del GDPR, i quali, impongono al titolare del trattamento l'obbligo di fornire indicazioni circa «l'esistenza di un processo decisionale automatizzato», nonché di procurare «informazioni significative sulla logica utilizzata[46] ».

Infine, il principio di non discriminazione, in virtù del quale «la legittimità dell'azione non è garantita dalla sola presenza di un algoritmo conoscibile e comprensibile, oggetto di controllo e validazione da parte di un funzionario, ma occorre che lo stesso non assuma carattere intrinsecamente discriminatorio[47]». L'attuazione del principio di non discriminazione algoritmica richiede dunque, come evidenziato dalla dottrina[48] e giurisprudenza[49], da un lato, la verifica della correttezza, dell'affidabilità e della qualità dei dati di input, al fine di evitare che gli eventuali profili di errore influenzino il risultato decisionale e producano un effetto discriminatorio; dall'altro, coinvolge la responsabilità organizzativa e preventiva nella fase iniziale di configurazione dei procedimenti automatizzati e delle regole algoritmiche da utilizzare.

Sulla scia del GDPR e allo scopo di arginare le preoccupanti dimensioni della diffusione di *deep fake*, come sopra accennato, il Titolo IV dell'AI Act[50] introduce precisi obblighi di trasparenza. Questi dovranno applicarsi ai sistemi che interagiscono con gli esseri umani, che rilevano emozioni, che stabili-

diritto dell'Unione o dello Stato membro cui è soggetto il titolare del trattamento, che precisa altresì misure adeguate a tutela dei diritti, delle libertà e dei legittimi interessi dell'interessato; c) si basi sul consenso esplicito dell'interessato».

[46] E. Carloni, *Algoritmi su carta. Politiche di digitalizzazione e trasformazione digitale delle amministrazioni*, in Dir. pubbl., 2, 2019, 363 ss.

[47] E. Carloni, *AI, algoritmi e pubblica amministrazione in Italia, in IDP. Revista de Internet, Derecho y Política*, 1, 2020.

[48] E. Carloni, *I principi della legalità algoritmica. Le decisioni automatizzate di fronte al giudice amministrativo*, in *Dir. amm.,* 2, 2020, 298-299.

[49] Cons. St., sez. VI, 13 dicembre 2019, n. 8472.

[50] M. Colonna, *Sezione I – I sistemi ad alto rischio,* in AIRIA Associazione per la Regolazione dell'Intelligenza Artificiale (a cura di), *Navigare l'European AI Act*, Milano, 2024, 71-82.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

scono associazioni con categorie sociali sulla base di dati biometrici, o che, appunto, generano o manipolano contenuti. Il considerando 134[51] precisa che i fornitori dovranno adottare soluzioni tecniche per i *deep fake* e *deloyer*, in modo tale da render chiaro che il contenuto è stato manipolato artificialmente. Nel contesto dell'IA Act, la trasparenza dei provider presuppone che vi sia a monte una fiducia da parte degli utenti finali che essi effettivamente mettano a disposizione le informazioni utili per conoscere e comprendere i sistemi immessi sul mercato. Laddove è impossibile avere contezza dei dati di addestramento, dei modelli, delle infrastrutture informatiche dei sistemi di IA, viene imposta al provider l'obbligo di una comunicazione chiara e comprensibile, secondo standard di ragionevolezza che, però, lascia ampi spazi di valutazione a chi deve raccontare e quindi selezionare le informazioni. L'art. 50 che apre il Capo IV dedicato agli obblighi di trasparenza prevede che i fornitori devono garantire che le soluzioni tecniche adottate per la marcatura siano «efficaci, interoperabili, solide e affidabili nella misura in cui ciò sia tecnicamente possibile, tenendo conto delle specificità e dei limiti dei vari tipi di contenuti, dei costi di attuazione e dello stato dell'arte generalmente riconosciuto, come eventualmente indicato nelle pertinenti norme tecniche». Analoga disposizione – art.50 paragrafo 4 - è prevista anche per i sistemi che generano *deep fake* con video, immagini, musica; o testi linguistici su questioni di pubblico interesse. In particolare, il Considerando 120[52] sancisce l'obbligo di dichiarare l'origine artificiale di un contenuto allo scopo di individuare i rischi sistemici che possono derivare dalla diffusione di contenuti manipolati e causare forme di discriminazione verso soggetti maggiormente vulnerabili. La sfida più importante, quindi, si svolge sul terreno della trasparenza "co-

[51] Cons. 134 – AI Act: «Oltre alle soluzioni tecniche utilizzate dai fornitori del sistema di IA, i *deployer* che utilizzano un sistema di IA per generare o manipolare immagini o contenuti audio o video che assomigliano notevolmente a persone, oggetti, luoghi, entità o eventi esistenti e che potrebbero apparire falsamente autentici o veritieri a una persona (*deep fake*), dovrebbero anche rendere noto in modo chiaro e distinto che il contenuto è stato creato o manipolato artificialmente etichettando di conseguenza gli output dell'IA e rivelandone l'origine artificiale. L'adempimento di tale obbligo di trasparenza non dovrebbe essere interpretato nel senso che l'uso del sistema di IA o dei suoi output ostacola il diritto alla libertà di espressione e il diritto alla libertà delle arti e delle scienze garantito dalla Carta, in particolare quando il contenuto fa parte di un'opera o di un programma manifestamente creativo, satirico, artistico, fittizio, o analogo fatte salve le tutele adeguate per i diritti e le libertà dei terzi. In tali casi, l'obbligo di trasparenza per i *deep fake* di cui al presente regolamento si limita alla rivelazione dell'esistenza di tali contenuti generati o manipolati in modo adeguato che non ostacoli l'esposizione o il godimento dell'opera, compresi il suo normale sfruttamento e utilizzo, mantenendo nel contempo l'utilità e la qualità dell'opera. È inoltre opportuno prevedere un obbligo di divulgazione analogo in relazione al testo generato o manipolato dall'IA nella misura in cui è pubblicato allo scopo di informare il pubblico su questioni di interesse pubblico, a meno che il contenuto generato dall'IA sia stato sottoposto a un processo di revisione umana o di controllo editoriale e una persona fisica o giuridica abbia la responsabilità editoriale della pubblicazione del contenuto».

[52] Cons. 120 – AI Act: «Inoltre, gli obblighi imposti dal presente regolamento ai fornitori e ai deployer di taluni sistemi di IA, volti a consentire il rilevamento e la divulgazione del fatto che gli output di tali sistemi siano generati o manipolati artificialmente, sono molto importanti per contribuire all'efficace attuazione del regolamento (UE) 2022/2065. Ciò si applica specialmente agli obblighi per i fornitori di piattaforme online di dimensioni molto grandi o motori di ricerca online di dimensioni molto grandi di individuare e attenuare i rischi sistemici che possono derivare dalla diffusione di contenuti generati o manipolati artificialmente, in particolare il rischio di impatti negativi effettivi o prevedibili sui processi democratici, sul dibattito civico e sui processi elettorali, anche mediante la disinformazione».

Downloaded from www.biodiritto.org.
ISSN 2284-4503

municata" e sull'auspicio che questa percorra senza troppi ostacoli tutta la catena dai fornitori di sistemi di AI, a chi li utilizza a chi ne subisce l'impatto per effetto di singole decisioni[53].

7. Il dovere delle Istituzioni

Alla luce delle considerazioni svolte, è possibile affermare che l'impiego delle nuove tecnologie sta determinando, e continuerà a determinare, un significativo cambiamento nella nostra vita. Nelle pagine che precedono, si è cercato di illustrare come l'impiego delle nuove forme tecnologiche, in particolare dei sistemi di AI, ponga gli interpreti del diritto dinanzi a nuove sfide e a nuovi interrogativi, ancora in attesa di un'unanime e puntuale risposta. Il punto più problematico pare rintracciarsi nell'esigenza di evitare che la complessità e l'incertezza, insieme con le strumentazioni e la progettazione di natura strettamente tecnica, si qualifichino come fattori di giustificazione delle discriminazioni.

Occorre, invece, intervenire, innanzitutto, a livello di c.d. morale soggettiva, ossia delle conoscenze, della cultura e dei codici di comportamento degli individui. Si rivela determinante l'azione sul piano educativo, mediante quella che gli studiosi definiscono tecnica di tutela by *education*[54]*,* al fine di avere un'IA in grado di rispettare i valori umani fondamentali che promuovono l'inclusione, l'equità, l'uguaglianza di genere e le diversità linguistiche e culturali, nonché di rispettare opinioni ed espressioni plurali. L'UNESCO ha già invitato, infatti, la comunità internazionale a riflettere sulle implicazioni di questa tecnologia sul lungo periodo in termini di conoscenza, insegnamento, apprendimento e valutazione, e ha offerto raccomandazioni concrete ai decisori politici e alle istituzioni educative su come l'uso degli strumenti di IA possa essere progettato per proteggere l'azione umana.

È compito del legislatore e delle autorità di regolamentazione, quindi, non sottovalutare l'impatto che la discriminazione ha sulle persone vulnerabili, perché un tale errore di calcolo potrebbe danneggiare lo spazio di libertà e di diritto che è alla base dell'Unione europea. Le istituzioni, pertanto, a parere di chi scrive, dovranno sempre più assumere un ruolo guida nello sviluppo di standard internazionali etici e di linee guida in grado di proteggere i diritti e le libertà degli interessati, soprattutto dei più vulnerabili. Questo richiederà sicuramente lo sviluppo di meccanismi di sorveglianza e audit che consentano di identificare tempestivamente eventuali pregiudizi insiti nei sistemi di AI. La cooperazione tra le istituzioni dei diversi Paesi risulta essere fondamentale per sviluppare standard comuni e garantire un utilizzo e una tutela uniforme. Inoltre, è necessario che vengano reperiti e fornite risorse umane capaci di regolamentare e sorvegliare sull'AI, quindi, in possesso non solo di competenze tecnico-informatiche, ma altresì di competenza legali ed etiche per identificare il rischio di discriminazione.

In particolare, compito primario delle istituzioni- alla luce degli obblighi imposti dal Regolamento – sarà quello di promuovere la trasparenza, quale architrave della progettazione e conseguente uso dei sistemi di AI. Solo così, attraverso la cultura della trasparenza, si potranno scongiurare i rischi di di-

[53] G. Lo Sapio, *L'Artificial Intelligence Act e la prova di resistenza per la legalità algoritmica*, in *Federalismi.it*, 16, 2024.

[54] A. Simoncini, S. Suweis, *Il cambio di paradigma nell'intelligenza artificiale e il suo impatto sul diritto costituzionale*, in *Rivista di filosofia del diritto*, 1, 2019, 87 ss.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

scriminazione algoritmica. A tal proposito, però, sarà necessario che il pubblico di utenti sia educato ai rischi della discriminazione che potrebbero essere insiti nel sistema stesso.

In conclusione, quindi, si ritiene che minare la discriminazione prodotta dall'impiego di questi strumenti di AI rappresenti la vera sfida attuale. L'AI Act, quale complesso normativo di regolamentazione dell'intelligenza artificiale rappresenta solo un primo passo verso la prevenzione dei rischi. Il suo successo dipenderà, infatti, dalla capacità dei governi, delle istituzioni e della società di individuare e sfruttare gli enormi benefici dell'AI, senza sacrificare i diritti fondamentali della persona.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Alla ricerca degli "anticorpi" contro le discriminazioni di genere nell'*AI Act*

Paolo Gambatesa*

LOOKING FOR THE "ANTIBODIES" AGAINST GENDER DISCRIMINATION IN THE AI ACT

ABSTRACT: The technological evolution, increasingly developing in a globalized world with the creation of «intelligent» systems, inevitably produces results that often escape human control. Among these are the often-unintentional distorting effects of algorithms on women. With the AI Act, the EU aims to limit such distortions through rigorous regulation to protect fundamental rights. This paper explores the different tools, implicit and explicit, offered by the new regulations to tackle gender discrimination.

KEYWORDS: AI Act; impact assessment; fundamental rights; gender discrimination; vulnerability.

ABSTRACT: L'evoluzione tecnologica, sviluppandosi sempre più in un mondo globalizzato con la creazione di sistemi «intelligenti», produce inevitabilmente risultati che spesso sfuggono al controllo umano. Tra questi, emergono gli effetti distorsivi degli algoritmi sulle donne, che agiscono spesso anche in maniera inconsapevole. Con l'AI Act, l'UE mira a limitare tali distorsioni attraverso una rigorosa disciplina a tutela dei diritti fondamentali. Questo contributo esplora i diversi strumenti, impliciti ed espliciti, offerti dalla nuova regolamentazione per contrastare le discriminazioni di genere.

PAROLE CHIAVE: AI Act; valutazione di impatto; diritti fondamentali; discriminazioni di genere; vulnerabilità.

1. Introduzione

È opinione diffusa e largamente condivisa che l'Intelligenza Artificiale (IA) sta rapidamente trasformando interi settori della società, anche attraverso la scoperta e la diffusione di nuove e migliori opportunità di vita. A questi indubbi elementi di novità, però, conseguono anche i signifi-

* *Assegnista di ricerca in Diritto costituzionale, Università di Milano. Mail: paolo.gambatesa@unimi.it.* *Il presente contributo costituisce un prodotto realizzato nell'ambito del progetto PRIN AiGeDi (Artificial Intelligence between generating and tackling gender-based discriminations), P.I. Prof.ssa Marilisa D'Amico. Contributo sottoposto a doppio referaggio anonimo.*

Downloaded from www.biodiritto.org.
ISSN 2284-4503

cativi rischi derivanti dalla sempre maggiore diffusione di sistemi di IA che finiscono perlopiù per impattare sulla vita delle persone vulnerabili, acuendo le "vecchie" discriminazioni e, allo stesso tempo, creandone delle "nuove".

In questo contesto, sono sempre più frequenti pregiudizi e discriminazioni sistemiche amplificati dalle tecnologie di IA a danno delle donne[1]. Il noto studio *gender shade*[2] ha messo in guardia dal trattamento altamente discriminatorio dei sistemi di riconoscimento facciale; così come anche gli algoritmi di *recruiting*[3], attingendo da modelli storici, hanno dimostrato la tendenza all'esclusione dei curricula femminili; e ancora, nell'ambito dei sistemi di IA generativa, l'UNESCO ha recentemente osservato come sia particolarmente elevato il tasso di *bias* di genere nei modelli linguistici di grandi dimensioni (LLM)[4].

Le radici dei pregiudizi di genere si possono rintracciare nella circostanza per cui le donne sono state storicamente escluse e sottorappresentate nel mondo delle nuove tecnologie. Nonostante i progressi tecnologici, il mondo dell'IA rimane dominato dagli uomini, con poche donne coinvolte nello sviluppo, nella ricerca e nelle decisioni strategiche. Questa esclusione non solo perpetua disuguaglianze di genere, ma limita anche la diversità di prospettive necessarie per creare un'IA più inclusiva e giusta.

---

[1] I riferimenti in letteratura sull'argomento sono molteplici, per tutti, si v. M. D'Amico, *Una parità ambigua: costituzione e diritti delle donne*, Milano, 2020, 313 ss.; F. Balaguer Calléjon, *La trasformazione dei diritti nella società digitale e il suo impatto sulla parità*, in M. D'Amico, B. Liberali (a cura di), *I diritti delle donne. Problematiche attuali e prospettive future*, Torino, 2024, 201 ss. e, nella medesima raccolta si v. anche E.C. Raffiotta, *Intelligenza artificiale e tutela dell'uguaglianza di genere*, 169 ss.; inoltre, v. E. Stradella, *Stereotipi e discriminazioni: dall'intelligenza umana all'intelligenza artificiale*, in *Liber amicorum per Paquale Costanzo*, 2020, in *Consultaonline.org*; C. Nardocci, *Intelligenza artificiale e discriminazioni*, in *Rivista del Gruppo di Pisa*, 3, 2021, 9 ss. (spec. 35 ss.).

[2] J. Buolamwini, T. Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, in *Proceedings of Machine Learning Research*, 81, 2018, 1 ss. Lo studio, in particolare, ha rivelato disparità nei sistemi di riconoscimento facciale, attraverso l'analisi delle tecniche di IA adoperate da tre aziende leader per la classificazione di genere. È stato stimato un tasso di errore nel riconoscimento di volti maschili bianchi inferiore all'1%, mentre per le donne nere superava il 34,7%. Similmente, un altro studio ha evidenziato gli effetti discriminatori di sistemi di riconoscimento facciale sulla base dell'età, specificatamente, a danno delle generazioni più anziane che più difficilmente vengono riconosciute correttamente dai sistemi di IA (cfr. J. Sung Park et al., *Understanding the Representation and Representativeness of Age in AI Data Sets*, in *AIES '21: Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, New York, 2021, 834 ss.).

[3] Cfr. G. Elisabeth Birkelund et al., *Gender Discrimination in Hiring: Evidence from a Cross-National Harmonized Field Experiment,* in *European Sociological Review,* 38, 2022, 337 ss. In aggiunta, sulle *policies* da implementare negli Stati europei in questo ambito, v. EIGE, *Artificial intelligence, platform work and gender equality*, report pubblicato nel dicembre 2021 e disponibile al seguente indirizzo https://eige.europa.eu/publications-resources/publications/artificial-intelligence-platform-work-and-gender-equality?language_content_entity=en.

[4] Unesco, *Challenging systematic prejudices: an investigation into bias against women and girls in large language models*, studio realizzato dall'*International Research Centre on Artificial Intelligence,* consultabile al seguente indirizzo: https://www.unesco.org/en/articles/generative-ai-unesco-study-reveals-alarming-evidence-regressive-gender-stereotypes (ultima consultazione 26/07/2024). Sulle problematiche sottese ai sistemi di *Natural Language Processing* (NLP) in chiava costituzionalistica si v. M. D'Amico, *Parole che separano. Linguaggio, Costituzione e diritti*, Milano, 2023, 128 ss.; per un ulteriore approfondimento si v. F. Mohammadi et al., *Identifying Gender Stereotypes and Biases in Automated Translation from English to Italian using Similarity Networks*, in *EWAF Conference Proceedings* (in corso di pubblicazione).

Downloaded from www.biodiritto.org.
ISSN 2284-4503

L'*AI Act*, recentemente pubblicato nella Gazzetta Ufficiale dell'Unione europea[5], attraverso un approccio basato sul rischio, intende arginare gli effetti distorsivi che i sistemi di IA possono generare in relazione alla tutela dei diritti fondamentali.

Benché la nuova regolamentazione europea risulti imperniata sulla tutela dei diritti fondamentali, sono esigue le disposizioni volte a rimuovere le discriminazioni che possono prodursi sulla base del genere[6].

Il tema delle discriminazioni di genere resta perlopiù "sotto traccia"[7], così da gravare sull'interprete il compito di ricostruire le coordinate "paritarie" entro cui poter considerare opportuno l'inserimento di limiti alla commercializzazione, all'uso e all'immissione sul mercato di determinati sistemi di IA.

L'obiettivo della presente analisi è quello di esplorare quali strumenti mette a disposizione il nuovo regolamento per la rimozione delle discriminazioni perpetrate ai danni delle donne.

Sulla scorta di questo presupposto, il percorso argomentativo prenderà le mosse dalle diverse sfumature del concetto di vulnerabilità che si rinvengono sia nelle disposizioni sui sistemi di IA vietati (*infra* §2) sia in quelle dedicate ai sistemi ad alto rischio (*infra* §3). Successivamente, l'analisi indugerà sul nuovo meccanismo della valutazione di impatto (*infra* §4), i cui risultati potranno agevolare sia l'individuazione di ulteriori effetti distorsivi sia nuove tecniche di rimozione di vizi di "genere" a base algoritmica. Nelle battute conclusive, l'attenzione sarà focalizzata sull'implementazione dell'alfabetizzazione di genere dell'IA (*infra* §5), quale profilo che completa e rafforza gli strumenti giuridici volti a contrastare le discriminazioni di genere.

[5] Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio, del 13 giugno 2024, che stabilisce regole armonizzate sull'intelligenza artificiale e modifica i regolamenti (CE) n, 300/2008, (UE) n, 167/2013, (UE) n, 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 (regolamento sull'intelligenza artificiale), GU, L, 2024/1689.

[6] Approfondisce i lavori preparatori in relazione ai profili di diritto antidiscriminatorio, C. NARDOCCI, *IA e Unione europea: primi (timidi) passi verso la tutela dei diritti,* in *Quaderni costituzionali*, 2, 2022, 385 ss.

[7] A parte quanto si dirà nel §5, si rinvengono alcuni espliciti riferimenti in tre considerando. Il considerando n. 27 precisa che «Con "diversità, non discriminazione ed equità" si intende che i sistemi di IA sono sviluppati e utilizzati in modo da includere soggetti diversi e promuovere la parità di accesso, l'uguaglianza di genere e la diversità culturale, evitando nel contempo effetti discriminatori e pregiudizi ingiusti vietati dal diritto dell'Unione o nazionale»; il considerando n. 48 afferma che «La portata dell'impatto negativo del sistema di IA sui diritti fondamentali protetti dalla Carta è di particolare rilevanza ai fini della classificazione di un sistema di IA tra quelli ad alto rischio. Tali diritti comprendono il diritto alla dignità umana, il rispetto della vita privata e della vita familiare, la protezione dei dati personali, la libertà di espressione e di informazione, la libertà di riunione e di associazione e il diritto alla non discriminazione, il diritto all'istruzione, la protezione dei consumatori, i diritti dei lavoratori, i diritti delle persone con disabilità, l'uguaglianza di genere, i diritti di proprietà intellettuale, il diritto a un ricorso effettivo e a un giudice imparziale, i diritti della difesa e la presunzione di innocenza e il diritto a una buona amministrazione»; ed infine, il considerando n. 58 «È inoltre opportuno classificare i sistemi di IA utilizzati per valutare il merito di credito o l'affidabilità creditizia delle persone fisiche come sistemi di IA ad alto rischio, in quanto determinano l'accesso di tali persone alle risorse finanziarie o a servizi essenziali quali l'alloggio, l'elettricità e i servizi di telecomunicazione. I sistemi di IA utilizzati a tali fini possono portare alla discriminazione fra persone o gruppi e possono perpetuare modelli storici di discriminazione, come quella basata sull'origine razziale o etnica, sul genere, sulle disabilità, sull'età o sull'orientamento sessuale, o possono dar vita a nuove forme di impatti discriminatori».

Downloaded from www.biodiritto.org.
ISSN 2284-4503

2. I sistemi di IA vietati e le ampie maglie del concetto di vulnerabilità

L'art. 1 del *corpus* normativo europeo sull'IA, nel definire il suo oggetto precisa che esso mira sia a «migliorare il funzionamento del mercato interno», sia a «promuovere la diffusione di un'intelligenza artificiale (IA) antropocentrica e affidabile», attraverso la garanzia di un «livello elevato di protezione della salute, della sicurezza e dei diritti fondamentali sanciti dalla Carta dei diritti fondamentali dell'Unione europea, compresi la democrazia, lo Stato di diritto e la protezione dell'ambiente [...]»[8].
L'espresso richiamo ai diritti della Carta di Nizza consente, in via implicita, di richiamare gli articoli 21 e 23 di quest'ultima, che rispettivamente impongono, da un lato, il divieto di forme di discriminazioni fondate sul sesso e, dall'altro, la parità di trattamento tra uomini e donne «in tutti i campi, compreso in materia di occupazione, di lavoro e di retribuzione». Ciò sino al ritenere, nel secondo periodo del par. 1 dell'art. 23, del tutto legittime azioni positive volte ad attribuire «vantaggi specifici a favore del sesso sottorappresentato»[9].
L'assonanza giuridica più vicina di queste disposizioni della Carta la si rinviene nel concetto di vulnerabilità[10], che emerge in relazione alla necessità di tenere indenni persone singole o gruppi[11] da effetti distorsivi dei sistemi di IA.
Il regolamento, però, non definisce compiutamente il concetto di vulnerabilità e il suo impego presenta diverse sfumature a seconda del sistema di IA cui la disciplina rivolga attenzione.
L'art. 5, in materia di sistemi di IA vietati, al par. 1, lett. b) dispone che ad essere vietato è tanto l'immissione sul mercato quanto la messa in servizio e l'uso di sistemi che deliberatamente sfruttino la vulnerabilità di una persona fisica o di uno specifico gruppo, in relazione «all'età, alla disabilità o a una specifica situazione sociale o economica».
Da una prima lettura della disposizione pare si possa desumere che sistemi di IA che sortiscano effetti discriminatori su determinati soggetti a motivo del loro genere non siano da considerarsi vietati. Tuttavia, si potrebbe dare una lettura estensiva dell'art. 5, par. 1, lett. b), al fine di ricomprendere anche il genere nel novero dei fattori che concretizzano una situazione di vulnerabilità.

[8] Sulle implicazioni, i possibili rischi e le prospettive derivanti da un approccio di regolamentazione dell'IA basato sulla tutela dei diritti, si v. M. ALMADA, N. PETIT, *The EU AI act: a medley of product safety and fundamental rights?,* in *EUI, RSC, Working Paper*, 59, 2023, disponibile al seguente indirizzo https://hdl.handle.net/1814/75982 .

[9] Al fine di concretizzare l'attuazione dei principi di non discriminazione e di parità, la Commissione europea ha adottato nel marzo 2020, la strategia «Un'Unione dell'uguaglianza: la strategia per la parità di genere 2020-2025» (COM(2020) 152 final), ove si precisa la necessita di un intervento da parte dei maggiori attori istituzionali, nazionali e sovranazionali, per affrontare efficacemente l'incidenza negativa dell'IA sulle donne.

[10] Sul concetto di «vulnerabilità» e le sue ricadute nell'Era digitale si rinvia alle più ampie riflessioni di G. MALGIERI, *Vulnerability and Data Protection Law*, Oxford, 2023.

[11] Come osserva, G. MALGIERI, *Vulnerability and Data Protection Law*, cit., 49-51, intorno alla vulnerabilità sono stati sviluppati due distinti approcci: il primo enfatizza la connotazione particolaristica della vulnerabilità, dal momento che essa caratterizzerebbe determinate persone o gruppi sulla base di specifiche situazioni socio-economiche; la seconda, invece, ritiene la vulnerabilità quale condizione universale che accomuna tutti gli esseri umani, pur potendo essa variare a seconda del tempo e del luogo in cui emerge. Un tentativo di composizione delle due teorie è quello di F. Luna (spec. nt. 22, 50) che propone una concezione «stratificata» della vulnerabilità, secondo cui «layers of vulnerability are not fixed attributes of specific individuals or groups but are features constructed by an individual's status, time, and location. In this sense, the concept of layering provides an opening to a more intersectional approach and stresses its cumulative and transitory potential» (pp. 50-51).

Downloaded from www.biodiritto.org.
ISSN 2284-4503

A suffragio di una simile interpretazione possono considerarsi due argomenti.

Il primo è legato alle parole «specifica situazione sociale ed economica», che seguono i concetti di più facile determinazione, quale «disabilità» ed «età» e che pare rinviino a precise situazioni la cui definizione spetta all'interprete. In quest'ottica, potrebbe rilevarsi come le discriminazioni di genere radicandosi in una società patriarcale, ove proliferano stereotipi contro le donne, possano ritenersi fonte di una specifica situazione sociale[12].

In altri termini, in una società ancora fortemente permeata da una cultura insensibile alla parità devono ritenersi necessarie tutte quelle misure volte a rimuovere gli ostacoli che sollecitano le disuguaglianze di genere.

Il secondo argomento si potrebbe rintracciare nella stessa sovraordinazione nel sistema delle fonti dell'Unione delle norme della Carta di Nizza. L'idea di limitare la tutela non discriminatoria solo a fattori come la «disabilità» e l'«età» genererebbe una ingiustificata (se non irragionevole) differenziazione rispetto agli altri elementi[13] che a norma dell'art. 21 della Carta posso fondare un trattamento discriminatorio. Proseguendo in questo solco, in un futuro rinvio pregiudiziale potrebbe dubitarsi della stessa validità dell'art. 5, par. 1, lett. b) in relazione all'art. 21 della Carta dei diritti fondamentali dell'UE.

Vi è un ulteriore profilo da prendere in considerazione in relazione all'art. 5, par. 1, lett. b). Il sistema di IA per essere vietato non dovrebbe solo «sfruttare» una delle condizioni riconducibile alla vulnerabilità, ma esso deve essere posto «con l'obiettivo o l'effetto» di arrecare un danno significativo a carico del soggetto o del gruppo di soggetti.

L'utilizzo dei termini «obiettivo» ed «effetto», legati da una preposizione disgiuntiva, lasciano intendere che l'elemento soggettivo sotteso alla realizzazione della lesione dei diritti da parte del sistema di IA sia da ricomprendersi tanto nella sfera del dolo quanto della colpa. Il sistema di IA, infatti, potrebbe essere stato volutamente creato con l'obiettivo di arrecare il c.d. danno significativo, ma quest'ultimo potrebbe anche semplicemente conseguire come effetto.

Emerge, così, la connessione tra il tema danno e la responsabilità degli attori che concorrono alla realizzazione e diffusione del sistema di IA.

Il tema della giustiziabilità del "danno" non trova approfondimenti nella regolamentazione europea, ma è stata premura della Commissione, già nel corso della precedente legislatura, avanzare la proposta di una direttiva *ad hoc* sulla responsabilità derivante dai sistemi di IA[14]. In quest'ultima, emerge la necessità di risarcire i danni derivanti da sistemi di IA (non solo ad alto rischio) e con essa l'idea che la tutela effettiva dei diritti costituisca un tassello necessario della regolamentazione comune europea.

---

[12] In questi termini, *mutatis mutandis*, è possibile leggere il fenomeno della violenza contro le donne. Per un approfondimento del tema in questi termini si v. M. D'Amico, C. Nardocci, S. Bissaro, *Le violenze contro la donna. Origini, forme, strumenti di prevenzione e repressione della violenza di genere*, Milano, 2023, e i saggi ivi contenuti.

[13] In particolare, gli ulteriori fattori di discriminazione individuati dall'art. 21 della Carta dei diritti fondamentali sono: il sesso, la razza, il colore della pelle o l'origine etnica o sociale, le caratteristiche genetiche, la lingua, la religione o le convinzioni personali, le opinioni politiche o di qualsiasi altra natura, l'appartenenza ad una minoranza nazionale, il patrimonio, la nascita e l'orientamento sessuale.

[14] Commissione europea, *Proposta di Direttiva del Parlamento europeo e del Consiglio relativa all'adeguamento delle norme in materia di responsabilità civile extracontrattuale all'intelligenza artificiale (direttiva sulla responsabilità da intelligenza artificiale)*, avanzata in data 8 settembre 2022, (COM/2022/496 final).

Downloaded from www.biodiritto.org.
ISSN 2284-4503

3. I sistemi ad alto rischio e la violazione in concreto del principio di parità

La nozione di vulnerabilità emerge anche in relazione ai sistemi ad alto rischio, che costituiscono il cuore della regolamentazione europea. Ciò in quanto tali sistemi risultano caratterizzati da una intrinseca propensione a ledere i diritti fondamentali e per tale ragione possono essere utilizzati solo nel caso in cui essi rispettino le condizioni previste dal regolamento (specificamente le norme contenute nel capo III).

Nella determinazione dei sistemi ad alto rischio l'art. 6, parr. 1 e 2, opera un rinvio agli allegati I e III. Quest'ultimo, in particolare, individua le otto macro materie in cui può aversi un sistema ad alto rischio.

Vi è modo di ritenere che la maggior parte dei sistemi di IA che potenzialmente potranno generare una discriminazione di genere saranno collocati proprio in questo novero.

Si pensi, ad esempio, agli algoritmi di selezione del personale che, basandosi su dati storici, possono perpetuare *bias* di genere nell'*output*, escludendo candidate che non rientrino nei modelli precedentemente assunti come standard[15].

Inoltre, le applicazioni di riconoscimento facciale, come già detto, hanno dimostrato tassi di errore più elevati per le donne, in particolare per le donne di colore, rispetto agli uomini bianchi, mettendo in evidenza la necessità di affrontare le disuguaglianze intrinseche nei dati di addestramento[16].

L'art. 7 descrive la procedura per emendare l'allegato III e, a tal fine, viene conferito alla Commissione il potere di adottare atti delegati, attraverso i quali si possono apportare modifiche o aggiunte in relazione ai sistemi di IA individuati nell'allegato in virtù del loro impatto negativo sui diritti fondamentali (art. 7, par. 1, lett. b)). In aggiunta, la medesima disposizione, al par. 2, elenca una serie di criteri che devono orientare la scelta emendativa. Tra questi, alla lett. h), viene dato rilievo all'esistenza di «uno squilibrio di potere» o alla circostanza per cui «le persone che potrebbero subire il danno o l'impatto negativo si trovino in una situazione vulnerabile rispetto al *deployer* di un sistema di IA, in particolare a causa della condizione, dell'autorità, della conoscenza, della situazione economica o sociale o dell'età».

Il riferimento alla situazione vulnerabile darebbe credito all'idea di una vulnerabilità potenzialmente transitoria, perlopiù legata al contesto in cui emerge piuttosto che ad elementi fissi predeterminabili

[15] Cfr. Allegato III, par. 4, lett. a), («i sistemi di IA destinati a essere utilizzati per l'assunzione o la selezione di persone fisiche, in particolare per pubblicare annunci di lavoro mirati, analizzare o filtrare le candidature e valutare i candidati»).

[16] A norma dell'Allegato III, par. 2 rientrerebbero sistemi che fanno leva sui dati biometrici, ed in particolare:
«a) i sistemi di identificazione biometrica remota. Non vi rientrano i sistemi di IA destinati a essere utilizzati per la verifica biometrica la cui unica finalità è confermare che una determinata persona fisica è la persona che dice di essere;
b) i sistemi di IA destinati a essere utilizzati per la categorizzazione biometrica in base ad attributi o caratteristiche sensibili protetti basati sulla deduzione di tali attributi o caratteristiche;
c) i sistemi di IA destinati a essere utilizzati per il riconoscimento delle emozioni».
Fanno eccezione quei sistemi di identificazione biometrica remota in tempo reale che, invece, sono annoverati tra quelli vietati (art. 5, par. 1, lett. h) e ss.).

Downloaded from www.biodiritto.org.
ISSN 2284-4503

ex ante[17]. Ad essere determinante è, infatti, la specifica posizione di squilibrio che emerga tra il *deployer* e «le persone».

Se, da un lato, questa specifica circostanza rende più agevole considerare le discriminazioni di genere nel novero dei sistemi di IA non ancora ricompresi in quelli ad alto rischio, dall'altro, però, il ruolo dell'interprete risulta sempre più discrezionale e per questo suscettibile di interpretazioni eterogenee in spregio ad una più certa tutela dei diritti su tutto il suolo europeo.

La vulnerabilità, così considerata, si lega all'eventualità che emerga l'impatto negativo o il danno, ora non più considerato come «significativo» (cfr. *supra* §2), nei confronti delle persone.

Un ulteriore profilo di interesse per la nostra analisi è costituito dall'articolato sistema di *compliance* in relazione ai sistemi di IA ad alto rischio.

La sezione II del capo III è interamente dedicata ai requisiti di questo sistema che si compone di un'analisi dei rischi (art. 9), di un vaglio sulla qualità dei *data set* (art. 10), della documentazione tecnica necessaria prima dell'immissione di un nuovo sistema (art. 11), di un meccanismo di tracciabilità e conservazione delle registrazioni (art. 12), di un apparato di istruzioni per l'uso volte a solidificare e concretizzare il principio di trasparenza (art. 13), dell'indefettibile procedimento di supervisione umana nell'ambito dello sviluppo e dell'implementazione di tali sistemi di IA (art. 14), e infine di una serie di regole in grado di assicurare la robustezza, l'accuratezza e la cybersicurezza (art. 15).

In particolare, l'art. 9 prescrive un sistema di gestione dei rischi continuo, ovverosia «come un processo iterativo continuo pianificato ed eseguito nel corso dell'intero ciclo di vita di un sistema di IA ad alto rischio [...]» (par. 2, primo periodo)[18]. E l'analisi del rischio si sviluppa sul binomio: uso improprio ragionevolmente prevedibile, da un lato, e accettabilità del rischio residuo, dall'altro.

Nell'effettuare tale analisi, precisa il par. 9, «i fornitori prestano attenzione [...] all'eventualità che il sistema di IA ad alto rischio possa avere un impatto negativo sulle persone di età inferiore a 18 anni o, a seconda dei casi, su altri gruppi vulnerabili».

Pertanto, il contrasto a precise situazioni discriminatorie assume importanza anche nell'ambito della *compliance*, cementificando così gli obblighi di controllo da parte degli attori principali che intervengono nel processo sia di emissione sia di utilizzazione dei sistemi ad alto rischio.

4. (segue...) La valutazione di impatto ai sensi dell'art. 27 del Regolamento sull'IA

Una delle principali novità che introduce il regolamento in relazione ai sistemi ad alto rischio è la valutazione dell'impatto che tali sistemi possono generare sui diritti fondamentali.

Il meccanismo in questione viene disciplinato all'art. 27 dell'*AI Act* che indirizza l'obbligo di realizzare una simile valutazione in capo ai *deployers*, siano essi «organismi di diritto pubblico, enti privati che forniscono servizi pubblici» o enti che agiscono in campo creditizio e assicurativo[19].

La procedura può essere suddivisa in diverse fasi chiave.

[17] In questi termini, G. MALGIERI, *Human vulnerability in the EU Artificial Intelligence Act,* in *Oxford University Press Blog,* 27 maggio 2024.

[18] Sui punti di forza e sulle problematiche che potranno sorgere in ambito applicativo in relazione al sistema basato sul rischio, si v. di C. NOVELLI, *L'Artificial Intelligence Act Europeo: alcune questioni di implementazione*, in *Federalismi.it,* 2, 2024, 95 ss.

[19] Cfr. allegato III, pt. 5, lett. b) e c).

Downloaded from www.biodiritto.org.
ISSN 2284-4503

La prima concerne l'identificazione dei rischi, in cui gli sviluppatori devono identificare i processi sottesi al processo di realizzazione del sistema di IA (art. 2, par. 1, lett. a)), il periodo di tempo in cui si prevede che debba essere utilizzato (art. 2, par. 1, lett. b)), le categorie di persone fisiche, ma non solo, che potenzialmente potranno essere destinatarie del sistema di IA (art. 2, par. 1, lett. c)) e i rischi che tali categorie potranno riscontrare nell'utilizzo dello stesso (art. 2, par. 1, lett. d)). Già in questa fase l'analisi potranno emergere potenziali effetti discriminatori, violazioni della privacy e altri impatti negativi.

Successivamente, gli sviluppatori dovranno indicare, da un lato, le misure di sorveglianza umana, ovverosia con quali modalità l'essere umano vigilerà sul «comportamento» del sistema di IA (art. 2, par. 1, lett. e)) e, dall'altro, le misure che verranno impiegate per mitigare i rischi che, con buona dose di probabilità verranno in essere (art. 2, par. 1, lett. f)).

Una volta raccolte tutte le informazioni, graverà sul *deployer* l'obbligo di notificarle all'autorità competente attraverso uno specifico modello elaborato dall'Ufficio per l'IA (art. 2, parr. 3 e 5).

In aggiunta, si noti come la norma qui in esame contribuisce a configurare questo meccanismo come dinamico e non già statico.

Ciò almeno per due ragioni.

La prima risiede nel costante aggiornamento a cui è sottoposta la valutazione *ex* art. 27. A norma del par. 2, infatti, se lo stesso sviluppatore, nel corso dell'utilizzo di tale sistema, avverte la presenza di modifiche deve adottare «le misure necessarie per aggiornare le informazioni».

La seconda investe i rapporti tra la valutazione di impatto e le altre valutazioni già operate in passato, le quali concorrono ad integrare quella disposta all'art. 27 dell'*AI Act*.

In quest'ottica, assume un particolare rilievo la valutazione di impatto relativa alla protezione dei dati (DPIA) a carico dei titolari dei trattamenti dei dati, prevista all'art. 35 del GDPR[20].

La DPIA e la valutazione di impatto dell'*AI Act* condividono l'obiettivo di identificare e mitigare i rischi, ma differiscono in alcuni aspetti chiave, come l'ambito di applicazione che risulta maggiormente circoscritto nella prima rispetto alla seconda. In aggiunta, la valutazione di impatto sull'IA, una volta effettuata deve essere trasmessa all'autorità competente, mentre la DPIA va esibita solo a richiesta, non essendo necessaria la sua trasmissione al Garante della Privacy, benché anch'essa debba essere predisposta preventivamente.

In definitiva, il meccanismo della valutazione di impatto agevolerà sia l'identificazione sia la gestione dei rischi, prima che possano causare danni significativi. E in aggiunta, la richiesta di fornire documentazione dettagliata renderà più chiari i processi di trasparenza e la responsabilità degli/lle attori/rici coinvolti/e.

Questi elementi potranno svolgere un ruolo strumentale fondamentale alla prevenzione e rimozione delle discriminazioni di genere.

Tuttavia, affianco ai pregi, è possibile individuare anche degli elementi potenzialmente negativi che potrebbero ostacolare suo fine antidiscriminatorio. In particolare, l'alto tasso di discrezionalità di cui gode il soggetto che effettuerà la valutazione, che potrebbe sottostimare o evitare di prendere in

[20] Sulle assonanze e differenze tra le due valutazioni di impatto si v. D. FULCO, *AI Act e Gdpr, come si rapportano: "valutazione d'impatto" e DPIA,* in *AgendaDigitale*, disponibile al seguente indirizzo https://www.agendadigitale.eu/cultura-digitale/ai-act-analogie-e-differenze-tra-la-valutazione-dimpatto-sui-diritti-fondamentali-fria-e-la-dpia/ (ultima consultazione 26/07/2024).

Downloaded from www.biodiritto.org.
ISSN 2284-4503

considerazione determinati effetti discriminatori generati dai sistemi di IA ad alto rischio. Sotto questo profilo, saranno particolarmente determinanti le linee guida che predisporrà la Commissione europea, con l'auspicio che essa fornisca strumenti adeguati per il riconoscimento e la mitigazione di tutte le forme di discriminazioni[21].

Inoltre, non è possibile stabilire sin da ora se tale strumento extragiudiziale avrà una qualche valenza probatoria nell'ambito delle future controversie che sorgeranno in relazione all'accertamento degli effetti discriminatori di determinati sistemi di IA. In questo contesto, il carattere discrezionale della valutazione di impatto potrebbe essere d'ostacolo al riconoscimento di tale funzione probatoria.

Da ultimo, per le piccole e medie imprese, la valutazione di impatto può essere un processo complesso e costoso, specie in termini di formazione e acquisizione di competenze specialistiche.

5. L'altra faccia della medaglia: l'*alfabetizzazione* paritaria dell'IA

Le sfide poste dall'IA sono sempre di maggiore impatto nella definizione della società del futuro e a queste si affiancano inevitabilmente anche quelle legate all'esigenza, più specifica, di una sempre più effettiva tutela antidiscriminatoria per le donne.

In questo senso sembra si possano leggere le uniche due disposizioni dell'*AI Act* dedicate al genere.

La prima è l'art. 68 che disciplina composizione e funzioni del gruppo di "esperti" che sarà chiamato a coadiuvare la Commissione nell'attuazione del regolamento. L'ultimo periodo del par. 2 esplicita che nella istituzione di tale gruppo viene garantita "un'equa rappresentanza di genere".

La seconda, invece, riguarda i codici di condotta che gli Stati sono chiamati ad implementare per i sistemi di IA non ad alto rischio. A norma dell'art. 95, tali codici, tra le altre cose, dovranno tener conto de "la valutazione e la prevenzione dell'impatto negativo dei sistemi di IA sulle persone vulnerabili o sui gruppi di persone vulnerabili, anche per quanto riguarda l'accessibilità per le persone con disabilità, nonché sulla parità di genere" (lett. e)).

Entrambe le norme sottolineano come sia essenziale la diffusione un'alfabetizzazione dell'IA in chiave paritaria.

Alfabetizzazione che non può limitarsi a quote o codici di condotta, ma dovrà necessariamente investire, in una prospettiva più ampia, anche nell'educazione e nella formazione nel campo dell'IA a tutti i livelli, dalle scuole primarie alle università, con un'attenzione particolare all'inclusione delle donne. Iniziative educative devono mirare a smantellare gli stereotipi di genere e a incoraggiare le ragazze e le giovani donne a intraprendere carriere nel campo dell'IA.

In conclusione, per affrontare efficacemente le situazioni di vulnerabilità nell'IA, è necessario un approccio integrato che combini regolamentazioni giuridiche solide e una diffusa alfabetizzazione paritaria. Solo così sarà possibile creare un ambiente tecnologico più inclusivo e giusto, dove le donne possano contribuire pienamente e beneficiare delle opportunità offerte dall'IA. Una IA sviluppata da una comunità eterogenea è una IA più robusta, creativa e capace di rispondere alle sfide globali in modo più efficace.

[21] Cfr. C. NOVELLI, *op. cit.*, 110 ss.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Special issue

(Trans)gender shades. I pericoli dell'intelligenza artificiale per il diritto all'identità delle persone trans

Sara Di Giovanni*

(TRANS)GENDER SHADES. THE DANGERS OF ARTIFICIAL INTELLIGENCE FOR THE RIGHT TO IDENTITY FOR TRANSGENDER PEOPLE

ABSTRACT: This contribution evaluates the impact of artificial intelligence (AI) on transgender rights, focusing on gender identity protection. While society tries to acknowledge gender complexity, AI operates with biased binary models, marginalizing transgender people. AI's foundation often misinterprets gender as a fixed, physiological binary, excluding transgender input. This disparity between AI construction and individual gender identity rights presents constitutional issues.

KEYWORDS: Artificial Intelligence (AI); Discriminatory Algorithms; Transgender People's Rights; Right to Gender Identity; Constitutional Challenges.

ABSTRACT: Questo contributo intende esaminare l'impatto discriminatorio dell'intelligenza artificiale (IA) sui diritti delle persone trans, concentrandosi sulla protezione dell'identità di genere. Mentre la società cerca di riconoscere la complessità di genere, l'IA opera con modelli binari, marginalizzando ulteriormente le persone trans. Le fondamenta dell'IA spesso interpretano il genere come un binomio fisico e immutabile, escludendo il contributo delle persone trans. Tale disparità tra la costruzione dell'IA e il diritto individuali all'identità di genere solleva problematiche di carattere costituzionale.

PAROLE CHIAVE: Intelligenza Artificiale (IA); algoritmi discriminatori; diritti delle persone trans; diritto all'identità di genere; sfide costituzionali.

* *Dottoranda di ricerca in Diritto costituzionale, Università di Milano. Mail: sara.digiovanni@unimi.it.* *Contributo sottoposto a doppio referaggio anonimo. Il titolo del contributo prende ispirazione dal progetto Gender Shades, un'iniziativa di ricerca promossa dalla ricercatrice Joy Buolamwini del MIT Media Lab nel 2017, volta a condurre uno studio sui sistemi di riconoscimento facciale automatizzato. I risultati hanno dimostrato che molti di questi sistemi sono imprecisi (o erronei) nel riconoscimento delle donne e delle persone nere rispetto agli uomini bianchi, determinando importanti conseguenze di carattere discriminatorio.*

Downloaded from www.biodiritto.org.
ISSN 2284-4503

1. Introduzione

Il rapido – e talvolta incontrollabile – sviluppo dell'intelligenza artificiale ha sollevato importanti interrogativi in ordine all'incidenza di tale tecnologia sui diritti fondamentali della persona.

Se, da un lato, la dottrina è concorde nel riconoscere la natura spesso discriminatoria dei sistemi di intelligenza artificiale[1], dall'altro lato rimangono ancora irrisolti gli interrogativi in merito alla precisa individuazione delle categorie di persone maggiormente colpite dall'impatto discriminatorio di tali tecnologie[2].

Sotto questo punto di vista, le evidenze statistiche dimostrano che gli effetti pregiudizievoli e discriminatori derivanti dall'uso dell'intelligenza artificiale colpiscono in misura maggiormente accentuata gli appartenenti a minoranze[3], intese queste ultime non solo come comunità numericamente più ridotte e limitate rispetto alla popolazione in generale, ma soprattutto come gruppi che occupano una posizione di subordinazione all'interno di una struttura gerarchica[4] che, quindi, li opprime.

Sarebbe possibile ritenere, dunque, che nella categoria da ultimo richiamata rientrino anche i membri della comunità LGBTQ+, e, in particolare le persone *trans*[5].

La dottrina ha evidenziato come gli studi condotti sui sistemi di intelligenza artificiale, che si traducono generalmente in forme di discriminazione nei confronti delle donne, sono in realtà limitati[6]. Molto spesso, la creazione dei sistemi di intelligenza artificiale[7] ha avuto alla base un'idea distorta dello stesso *genere*, venendo questo trattato come «un concetto binario, immutabile e discernibile fisiologicamente»[8]. Ne è esempio il c.d. *automatic gender recognition*[9] una sotto-categoria del riconoscimento facciale che cerca, avvalendosi di un software, di identificare il genere delle persone attraverso una molteplicità di elementi di carattere biometrico. Sistemi di tale portata, basati sul presupposto che l'appartenenza ad un genere sia definibile – oltre che in termini binari – anche aprioristicamente attraverso elementi universali, come possono essere tratti fisici, comportamentali o di altra natura, come il riferimento al nome proprio, fondano un pericolo importante: la tecnologia che si basa su concezioni errate del genere porterà a errori di precisione nel riconoscimento e nella identificazione delle

[1] Sul tema, si v. M. D'Amico, C. Nardocci, *Intelligenza artificiale e discriminazione di genere: rischi e possibili soluzioni*, in G. Cerrina Feroni, C. Fontana, E.C. Raffiotta (a cura di*), AI Anthology. Profili giuridici, economici e sociali dell'intelligenza artificiale*, Milano, 2022; C. Colapietro, *Intelligenza artificiale e discriminazioni, in Studi parlamentari e di politica costituzionale*, 211, 2022, 9-17.

[2] C. Nardocci, *Intelligenza artificiale e discriminazioni*, in *Rivista del Gruppo di Pisa*, 3, 2021, 15.

[3] S.U. Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism*, New York, 2018; C. Nardocci, *Minorities and Minority Rights in the Era of Artificial Intelligence, in European yearbook on minority issues brill* 2025, in corso di pubblicazione.

[4] C. Nardocci, *Intelligenza artificiale e discriminazioni*, cit., 4.

[5] La distinzione tra i termini transessuali, transgender e, ad oggi, trans, è di estremo rilievo. Per una definizione delle prime due categorie terminologiche, si v. B. Pezzini, *Transgenere in Italia: le regole del dualismo di genere e l'uguaglianza*, in G. Vidal Marcìlio Pompeu, F. Facury Scaff (a cura di), *Discriminação por orientação sexual. A homossexualidade e a transexualidade diante da experiência constitucional*, Brazil, 2012, 327 ss.

[6] In questo senso, O. Keyes, *The Misgendering Machines: Trans/HCI Implications of Automatic Gender Recognition, Proceedings of the ACM on Human-Computer Interaction*, Vol. 2, No. CSCW, Article 88, 2018.

[7] Da ora, IA.

[8] O. Keyes, *The Misgendering Machines*, cit.

[9] Da ora, AGR.

Downloaded from www.biodiritto.org.

ISSN 2284-4503

persone, nonché al rafforzamento di stereotipi già esistenti nella realtà con conseguenze psicologiche negative[10] e, da ultimo, ad un uso malintenzionato da parte di governi[11] e attori privati.
L'obiettivo del presente contributo è, allora, quello di valutare l'impatto che i sistemi di intelligenza artificiale dispiegano nei confronti dei diritti delle persone *trans*, con particolare riferimento alla tutela del loro diritto all'identità personale[12].

2. Il diritto costituzionale all'identità sessuale (prima) e di genere (poi)

Per poter affrontare il tema dell'impatto dell'intelligenza artificiale sui diritti delle persone *trans* è necessario muovere da una ricostruzione dei loro diritti nella prospettiva del diritto costituzionale.
Nel 1968, il *transessualismo* viene definito dalla scienza medica come una «sindrome caratterizzata dal fatto che un individuo, genotipicamente e fenotipicamente di un sesso determinato, ha la consapevolezza di appartenere al sesso, o meglio, al genere opposto»[13]. Una simile definizione ha conosciuto, nel corso del tempo, un importante ampliamento – seppur mai abbandonando la componente "patologizzante"[14] – in quanto le categorie di sesso e di genere, sulle quali si gioca il riconoscimento delle soggettività *trans* sono categorie dinamiche e, inevitabilmente, suscettibili di profonde trasformazioni.
Tuttavia, è la divergenza tra sesso attribuito alla nascita e identità nella quale ci si riconosce – e la eventuale e successiva necessità di raggiungere un riequilibrio tra tali componenti – a interessare il diritto.
Come noto, il «problema del transessualismo»[15] si pone all'attenzione delle scienze giuridiche a partire dagli anni Settanta e, in particolare all'attenzione della Corte costituzionale nel 1979.
In occasione della sentenza n. 98 del 1979, la Corte costituzionale fu chiamata a pronunciarsi sulla legittimità costituzionale degli artt. 165 e 167 del r.d.l. 9 luglio 1939, n. 1238 (*Ordinamento dello stato*

[10] In generale, occorre sottolineare come lo stereotipo (e, nello specifico, lo stereotipo di genere) sia funzionale al rafforzamento di classificazioni che ridondano in trattamenti discriminatori riservati ad un gruppo sociale. Cfr. C. NARDOCCI, *La generalizzazione irragionevolmente discriminatoria: lo stereotipo di genere tra diritto e corti*, in *Genius*, 2018, 4.

[11] Basti soltanto ricordare che, ad oggi, sono nove i paesi nel mondo dove il transessualismo è considerato un reato, cui si aggiungono altri paesi dove, nonostante l'assenza di leggi specifiche, i governi attuano politiche di repressione o di limitazione della libertà di espressione. Per saperne di più, si v. *Human Rights Watch*, #OUTLAWED, *The love that dare not speak its name*, consultabile al link https://features.hrw.org/features/features/lgbt_laws/ (ultima consultazione 2/12/2024)

[12] Fondamentale, in tal senso, è la sentenza n. 13 del 1994, con la quale la Corte costituzionale ha riconosciuto, per la prima volta, il diritto all'identità personale come parte del «patrimonio irretrattabile della persona umana», riconducendolo nell'alveo dell'art. 2 Cost. Sul tema, si v. A. PACE, *Nome, soggettività giuridica e identità personale*, in *Giurisprudenza costituzionale*, 1, 1994, 103.

[13] A. MURATORIO, U. PALAGI, *Aspetti psichiatrici e medico legali del transessualismo*, in *Giorn. med. leg., infortun. e tossicol.*, 1968, 259.

[14] Sul tema della patologizzazione del transessualismo, si v. A. LORENZETTI, *Diritti in transito. La condizione giuridica delle persone transessuali*, Milano, 2013, 88 ss.; F. SACCOMANDI, *Spesso non binarie, sempre non conformi: la "piena depatologizzazione" delle soggettività trans*, in *Genius*, 2, 2020.

[15] Così si esprimeva la dottrina in occasione delle prime riflessioni proposte in seguito alla sentenza n. 98 del 1979, con la quale la Corte costituzionale non ha riconosciuto fondamento costituzionale alle pretese di modifica del sesso da parte delle persone transessuali. In questo senso, S. BARTOLE, *Transessualismo e diritti inviolabili dell'uomo*, in *Giurisprudenza Costituzionale*, 1979, 1179 ss.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

civile), e dell'art. 454 del codice civile, in relazione agli artt. 2 e 24 della Costituzione, nella misura in cui escludevano il diritto alla rettificazione dell'atto di nascita e alla attribuzione del sesso femminile «nell'ipotesi di modificazioni artificiali di un sesso che facciano perdere ad un individuo le caratteristiche peculiari maschili ed acquistare quelle femminili esterne, qualora le modificazioni stesse trovino corrispondenza in una originaria, indiscutibile, personalità psichica di natura femminile».

Il dibattito dottrinale che ha interessato il riconoscimento costituzionale del diritto all'identità si sviluppava, all'epoca, in relazione all'art. 2 Cost. e, più specificatamente, al principio personalista. Si riteneva, cioè, che il richiamo all'art. 2 Cost. fosse destinato ad offrire il fondamento non solo per una ricostruzione delle implicazioni di carattere civilistico, ma anche penalistico. In altri termini, il richiamo ad una capacità estensiva dell'art. 2 Cost.[16] avrebbe impedito alla Corte costituzionale di imbarcarsi in un'opera di «ingegneria giuridica»[17], sopperendo ad un'inerzia legislativa che mancava di riconoscere una tutela normativa a persone che chiedevano una correzione del sesso attribuito alla nascita.

Tuttavia, con la sentenza richiamata, la Corte costituzionale ha deciso di percorrere la strada di una "chiusura" della clausola di cui all'art. 2 Cost., affermando che «dalla Costituzione non è possibile desumere una tutela di quel diritto cui richiamavasi l'attore in giudizio e che il giudice *a quo* – riconoscendolo sprovvisto di tutela nella legge ordinaria – ha ritenuto potesse essere compreso fra i diritti inviolabili dell'uomo»[18], ritenendo si debba parlare non tanto di «identità sessuale», quanto piuttosto di (presunto) diritto a far riconoscere un sesso esterno diverso dall'originario, a seguito di un'operazione chirurgica – comunque non riconducibile ad assi costituzionali.

La dottrina ha criticato fortemente una simile impostazione, ritenendo che per «identità sessuale» deve intendersi piuttosto il diritto a far riconoscere il sesso reale dell'individuo, apparendo essa un aspetto della più generale identità, intesa come «complesso degli elementi che caratterizzano (e distinguono) l'individuo e la sua personalità»[19]. Di conseguenza, la garanzia dello sviluppo della persona umana deve essere piuttosto considerata in relazione alla protezione dell'identità e, nello specifico, dell'identità sessuale[20], necessariamente riconducibile alla portata "aperta" dell'art. 2 Cost.

Nonostante la sua posizione di chiusura di cui si è detto, la Corte costituzionale consegnava un auspicio a che il legislatore italiano, così come stava accadendo in altri ordinamenti giuridici, potesse intervenire e definire il "problema" del transessualismo, soprattutto in considerazione delle implicazioni giuridiche che avrebbe comportato sul fronte dell'istituto del matrimonio[21].

---

[16] Come noto, la dottrina costituzionalista ha impegnato diversi momenti di riflessione attorno alla capacità o meno dell'art. 2 Cost. di ricomprendere nella sua portata diritti di libertà ulteriori rispetto a quelli indicati negli artt. 13 e ss. della Carta costituzionale. Per una lettura dell'art. 2 Cost. come contenente una "clausola aperta" si v. A. BARBERA, Art. 2, in G. BRANCA (a cura di), *Commentario della Costituzione*, Bologna-Roma, 1979, 91-92. Nel senso di interpretare l'art. 2 Cost. come "clausola chiusa", invece, A. PACE, *Problematica delle libertà costituzionali*, Padova, 1992, 4 ss., il quale sostiene che il riconoscimento di diritti ulteriori rispetto a quelli previsti espressamente dalla Costituzione potrebbe comportare delle insanabili antinomie con altre norme costituzionali.

[17] S. BARTOLE, *op. cit*., 1180.

[18] Così, Corte costituzionale, sent. n. 98 del 1979, cons. in dir. 2.

[19] M. DOGLIOTTI, *Identità personale, mutamento del sesso e principi costituzionali*, in *Giurisprudenza italiana*, 26, 1981.

[20] *Ibidem*.

[21] La Corte costituzionale, in diverse occasioni, è stata chiamata a pronunciarsi sulla disciplina del matrimonio e dell'unione civile a fronte di un percorso di riaffermazione di genere intrapreso da uno dei partner. In tal senso,

Downloaded from www.biodiritto.org.
ISSN 2284-4503

In questo senso, la legge n. 164 del 1982 (Norme in materia di rettificazione di attribuzione di sesso)[22] rappresenta un punto di svolta, in quanto non solo sancisce e riconosce, per la prima volta, il diritto ad ottenere una rettifica del sesso e dei dati anagrafici per i c.d. "diversi"[23], ma alimenta altresì un cambiamento giurisprudenziale della stessa Corte costituzionale. Non sorprende, così, che, nel 1985 il Giudice costituzionale sia arrivato a riconoscere il fondamento costituzionale di un diritto all'identità sessuale, inteso come un «aspetto e fattore di svolgimento della personalità»[24]. Il diritto all'identità personale deve essere inteso, infatti, come «esigenza di "essere sé stessi", nella prospettiva di una compiuta rappresentazione della personalità individuale in tutti i suoi aspetti ed implicazioni, nelle sue qualità ed attribuzioni»[25]. In questo modo, esso viene ricondotto nell'alveo dei c.dd. diritti di identità personale, i quali, insieme al diritto al nome, all'immagine e ai segni distintivi, contraddistinguono la persona[26].

Per ciò che concerneva le persone transessuali, originariamente l'identità sessuale veniva ricondotta alla tutela dell'integrità fisica e alla problematica degli atti di disposizione del proprio corpo, attraverso il criterio esclusivo degli organi genitali esterni per l'individuazione del sesso[27]. La procedura di rettifica del sesso, in virtù della legge n. 164 del 1982, prevede che la persona interessata presenti ricorso al Tribunale ordinario, il quale valuta documenti, testimonianze e una perizia medico-legale; inoltre, se necessario, autorizza l'intervento chirurgico e pronuncia una sentenza di rettifica del sesso negli atti di stato civile. Così costruita, la disciplina della riaffermazione di genere è apparsa nel tempo frammentaria, tanto da essere stata interessata da diversi interventi della Corte costituzionale. Significative, in questa prospettiva, sono le sentenze n. 221 del 2015[28] e n. 180 del 2017[29], con le quali la Corte costituzionale ha interpretato la legge n. 164 del 1982 in modo da "eliminare" la rigidità del requisito essenziale della modifica dei caratteri sessuali sia primari sia secondari per accedere alla rettifica definitiva dei dati anagrafici – riconoscendo questo momento piuttosto come possibile ed eventuale – e garantendo, di contro, il diritto alla salute psico-fisica che si lega inscindibilmente ai percorsi di

si richiamano le sentenze n. 170 del 2014 e la sentenza n. 66 del 2024 della Corte costituzionale, con le quali la Corte si è pronunciata in punto di effetti della rettifica nei confronti dell'istituto del matrimonio e del legame da unione civile. Per saperne di più, si v. B. PEZZINI, *La Corte costituzionale applica una condizione risolutiva al matrimonio del transessuale*, in *www.confronticostituzionali.eu*, 2014; A.LORENZETTI, *Corte costituzionale e Corte europea dei diritti umani: l'astratto paradigma eterosessuale del matrimonio può prevalere sulla tutela concreta del matrimonio della persona trans*, in *La nuova giurisprudenza civile commentata*, 12(I), 2014, 1152; B. LIBERALI, *Sulla trasformazione del rapporto di coppia a seguito di rettificazione di sesso dieci anni dopo: la parola (ancora) alla Corte costituzionale*, in *Diritti comparati*, 2024.

[22] Sul tema, si v. C. LA FARINA, *Alcune osservazioni riguardo alla legge sul cambiamento di sesso*, in *Rivista italiana di medicina legale*., 1983, 815-939; S. PATTI, M. R. WILL, *Commento alla legge 14 aprile 1982, n. 164*, in *Nuove leggi civ. comm*., 1983, 38-46.

[23] Così, Corte cost., sent. 161 del 1985.

[24] Corte cost., sent. 161 del 1985. Per un commento alla sentenza, si v. M. DOGLIOTTI**,** *La Corte costituzionale riconosce il diritto all'identità sessuale*, in *Giurisprudenza italiana*, parte I, sez. I, 1987, 235.

[25] *Ibidem*.

[26] Così, F. MODUGNO, *I nuovi diritti nella giurisprudenza costituzionale*, Torino, 1995, 12 ss.

[27] M. DOGLIOTTI, *La Corte costituzionale riconosce il diritto all'identità sessuale*, cit., 235 ss.

[28] Per un commento, I. RIVERA, *Le suggestioni del diritto all'autodeterminazione personale tra identità e diversità di genere. Note a margine di Corte cost. n. 221 del 2015*, in *Consulta Online*, 1, 2016;

[29] Per un commento, si v. A. LORENZETTI*, Il cambiamento di sesso secondo la Corte costituzionale: due nuove pronunce (nn. 180 e 185 del 2017)*, in *Studium iuris*, 4, 2018, 446.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

transizione[30]. Proprio in queste pronunce, si realizza un cambio terminologico e concettuale, non parlandosi più solo di "identità sessuale", ma, piuttosto, di «diritto all'identità di genere quale elemento costitutivo del diritto all'identità personale, rientrante a pieno titolo nell'ambito dei diritti fondamentali della persona (art. 2 Cost. e art. 8 della CEDU)». Una affermazione che costituisce un approdo senza dubbio centrale della evoluzione culturale dell'ordinamento giuridico[31].

Una simile impostazione ha permesso di rafforzare (e riformare) la nuova concezione di "identità sessuale" introdotta con la legge n. 164 del 1982 e confermata con la sentenza n. 161 del 1985, in favore di una tutela ad ampio raggio del diritto all'identità di genere. Essa non deve più essere considerata solo in relazione agli organi genitali esterni, ma anche ad elementi di carattere psicologico e sociale, comunque essenziali al processo di sviluppo della personalità della persona umana. In questo modo, le richieste giuridiche di riaffermazione di genere si sono spostate da un piano prettamente medico e chirurgico ad un piano di autodeterminazione di genere, fondato sul principio costituzionale di autodeterminazione di cui all'art. 2 Cost.[32].

3. Le sfide del genere nella società: oltre il sistema binario dell'identità

Il tema del transgenerismo impone, inoltre, di riflettere su che cosa si intende per sesso e genere e su che cosa si dovrebbe intendere per sesso e genere affinché a tutte le soggettività *trans* venga riconosciuta una particolare tutela giuridica, nell'ottica del dispiegamento del diritto costituzionale all'identità di genere.

Il dibattito dottrinale e scientifico che ha interessato la discussione attorno ai concetti di sesso e genere è molto ampio e non è questa la sede per ridurre a poche battute una discussione accademica che ha visto l'intersezione di una molteplicità di voci e anche di discipline[33].

Basti qui soltanto ricordare che la rilevanza dei termini sesso e genere *f*onda le proprie origini nella scienza medica, dove il sesso era inizialmente inteso come «una delle due forme principali di individui che si presentano in molte specie e che si distinguono rispettivamente come femmine o maschi, soprattutto in base agli organi e alla struttura riproduttiva»[34].

Accanto ad un primo approccio di carattere strettamente biologico, si è sviluppato un secondo approccio, di matrice psichiatrica, con la quale appare per la prima volta nel panorama medico, politico e dottrinale anche il termine *genere*. Gli psichiatri e gli psicologi, che, nell'ambito dello studio delle

[30] Sul rapporto tra transessualismo, rettificazione di sesso ed essenzialità dell'intervento chirurgico, si v. N. POSTERARO, *Transessualismo, rettificazione anagrafica del sesso e necessità dell'intervento chirurgico sui caratteri sessuali primari: riflessioni sui problemi irrisolti alla luce della recente giurisprudenza nazionale*, in *Rivista italiana di medicina legale*, 4, 2017, spec. 1359 ss.

[31] Corte cost., sent. n. 221 del 2015, cons. in dir. 4.

[32] Per una ricostruzione giurisprudenziale delle tappe che hanno interessato il riconoscimento del diritto al cambiamento di sesso, si v. M. D'AMICO, *I diritti dei "diversi". Saggio sull'omosessualità*, in *Osservatorio AIC*, 6, 2021, 163 ss.

[33] La letteratura, come anticipato, è molto vasta. Tra le voci più autorevoli, si v. T. MOI, *What is a Woman? And Other Essays*, Oxford, 1999, A. OAKLEY, *Sex, Gender and Society*,Farnham, 1985: J. BUTLER, *Gender trouble: Feminism and the Subversion of Identity*, London, 1999. Per una ricognizione più puntuale del dibattito dottrinale sul tema, si v. E. SCHIAPPA, *The transgender exigency. Defining Sex and Gender in the 21st Century*, London, 2022.

[34] E. SCHIAPPA, *The transgender exigency*, cit., 15.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

persone intersessuali e transessuali, per primi introdussero il concetto di genere, tentarono di descrivere la condizione transessuale come quella di una persona che si sente "intrappolata" nel corpo sbagliato: ciò ha permesso di distinguere tra sesso biologico e orientamento psicologico di un individuo. Nel 1955, lo psicologo John Money accolse nel dibattito scientifico la nozione di "*gender role*", basata sulla storia delle persone ermafrodite, per differenziare le attitudini e i comportamenti distinti dalle caratteristiche connesse al sesso biologico.

È in questo panorama scientifico che viene proposta per la prima volta la definizione di sesso come categoria biologica e di genere come categoria sociale e psicologica, funzionale ad una classificazione sociale tra mascolinità e femminilità in quanto bacino delle ricostruzioni sociali di quei tratti e comportamenti considerati appropriati per ciascuno dei due sessi[35].

Infatti, come noto, la differenziazione nei ruoli e negli *status* giuridici riconosciuti in capo a ciascun individuo rappresenta una pratica ricorrente nell'organizzazione pubblica[36], dove la categorizzazione sessuale binaria era (e, in un certo modo, è ancora) necessaria in una società patriarcale basata su un sistema di separazione delle sfere. In questo modo, il sistema giuridico – attraverso un contesto di «discriminazione strutturale»[37] – assegna alle donne il compito di gestire la "sfera privata" della cura e della produzione, mentre all'uomo affida la "sfera pubblica" della politica e del lavoro retribuito[38]. In un mondo così costruito, il sesso anagrafico non rappresenta una realtà pre-giuridica[39], quanto piuttosto una categoria funzionale all'attribuzione di compiti e responsabilità.

La sopravvivenza di una identità binaria immediatamente intellegibile, esclusiva e stabile è motivata dalla necessità di mantenere un sistema istituzionale caratterizzato e alimentato da eteronormatività[40]: di conseguenza, le categorie binarie sessuali mantengono un ruolo di mantenimento degli status individuali e di divisione dei ruoli. In quest'ottica, la transizione e il cambio di identità rappresentano un elemento di incertezza e al tempo stesso di "minaccia" ad un sistema così definito.

Tuttavia, l'evoluzione della scienza medica e biologica[41], ma anche del dibattito accademico e dottrinale dimostrano come la prevalenza del binarismo rappresenta il residuo di una dominazione culturale

[35] *Ibidem*, 18 ss.

[36] P. STARR, *Social categories and claims in the liberal state*, in *Social research*, 1992, 263 ss.

[37] S. OSELLA, *Come evolve il diritto all'identità di genere? Fattori strutturali, culturali e dogmatici nella giurisprudenza costituzionale italiana e colombiana. Un'analisi comparata*, in *Rivista di Diritti Comparati*, 4, 2023, 2.

[38] *Ibidem*.

[39] *Ibidem*, dove l'A. sostiene che neanche il principio di eguaglianza avrebbe ridimensionato l'importanza del sesso anagrafico, ma la sua dimensione sostanziale (attraverso, ad esempio, le azioni positive) avrebbe piuttosto rafforzato l'utilità di questa categoria. L'A. porta come esempio la preservazione di un diritto di famiglia esclusivamente eterosessuale: in questo caso, la scriminante per accedere all'istituto matrimoniale continua a rimanere il sesso anagrafico.

[40] J. HALLEY, *Split decisions. How and why to take a break from feminism*, Princeton and Oxford, 2006, 136; J. BUTLER, *Gender trouble*, cit., 22-23, così come richiamati da S. OSELLA, *Come evolve il diritto all'identità di genere?,* cit., 22.

[41] Si pensi al famoso studio condotto dalla biologa femminista Anne Fausto-Sterling, che nel 1993 ha sostenuto che nella natura umana è possibile individuare almeno cinque sessi o comunque un numero indefinito, essendo impossibile ridurre ad un numero la complessità del sesso. Si veda, a tal proposito, A. FAUSTO-STERLING, *The five sexes. Why male and female are not enough*, in *The Sciences*, 1993, 20.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

e politica europea[42] che non riesce a rispondere più alle esigenze delle nuove minoranze sessuali[43] e, in particolare, delle minoranze *trans non binarie*[44].

Anche la legge n. 164 del 1982 dispiega una struttura binaria e "binarizzante", escludendo – almeno per ora[45] – la possibilità di un riconoscimento di un «terzo sesso» [*rectius*: terzo genere], soddisfacendo le pretese di riconoscimento di persone e identità (che comunque esistono), in un'ottica non tanto (e non più) di diritto alla rettifica del sesso, quanto di diritto all'autodeterminazione di genere[46].

Il tema centrale che qui interessa sottolineare riguarda la difficoltà della società e del diritto di attenersi ad un'ondata di cambiamenti nelle rivendicazioni di riconoscimento, nei ruoli di genere, nelle capacità di autodeterminazione degli individui che porteranno inevitabilmente ad un ripensamento delle categorie culturali e politiche, verso una de-binarizzazione degli *status* giuridici.

Al tempo stesso occorre chiedersi come, se la società e il diritto faticano ad abbandonare i pregiudizi e gli stereotipi che animano la società patriarcale e il sistema di divisione dei ruoli su cui da sempre si fonda, questi possano essere abbandonati (o non trasferiti) nelle nuove frontiere della società.

---

[42] M. LUGONES, Heterosexualism and the colonial/modern gender system, in Hypatia, 2007, 186 ss.

[43] Intese cioè come «minoranze che [incarnano] forme di vita non conformi alla norma e che le rende destinatarie di un qualche tipo di stigma» (Cfr. M. MONTALTI, *Orientamento sessuale e Costituzione decostruita*, Bologna, 2007, 24 ss.). Si v. inoltre E. HEINZE, *Sexual Orientation: a Human Right. An Essay on International Human Rights Law*, Dordrecht-Boston-London, 1995, 243, in cui l'A. sottolinea come il concetto di minoranza nasce per alludere ad uno stato d'oppressione e viene mantenuto per la trasformazione in stato di liberazione, ad un tempo conservando e sorpassando le sue stesse origini storiche. Seguendo questa logica, adottata anche da M. MONTALTI, *Orientamento sessuale e Costituzione decostruita*, cit., 20 ss., e affermando che gli individui possono tendere a sentirsi parte di una minoranza in virtù del loro comportamento sessuale, che può essere considerato come espressivo di una deviazione inaccettabile, si potrebbe giungere ad affermare che nella più ampia nozione di "minoranza sessuale" possano rientrare anche le persone trans, in quanto distinte per comportamenti devianti dai paradigmi di "normalità".

[44] Intese cioè come quelle identità che negano, superano o rifiutano il genere e il sistema dicotomico di maschile e femminile. Sul tema, si v. C. P. GUARINI, *Appunti su "terzo sesso" e identità di genere*, in *Dirittifondamentali.it*, 1, 2019.

[45] In realtà, il diritto sta iniziando a percepire la tensione del sistema sessuale binario. Si pensi alla sentenza n. 143 del 2024, con la quale la Corte costituzionale si è pronunciata sulla questione di legittimità costituzionale sollevata dal Tribunale di Bolzano anche sull'art. 1 della legge n. 164 del 1982, nella parte in cui non permette – al momento della rettifica del sesso anagrafico – di indicare un sesso diverso da quello maschile e da quello femminile. In altri termini, quindi, alla Corte costituzionale è stato chiesto indirettamente di pronunciarsi sul legittimo riconoscimento delle identità non binarie. Tuttavia, nella sentenza richiamata, nonostante la Corte riconosca il «tono costituzionale» della problematica, ritiene che il riconoscimento delle identità non binarie sia materia riservata al legislatore.

[46] Sul tema, si v. R. RUBIO MARIN, S. OSELLA, *La autodeterminación de género: Gender Critical Radfems a la prueba de la proporcionalidad*, in *IberICONnect Blog*, 2021. Interessante è anche il panorama comparato, dove gli ordinamenti giuridici stanno introducendo progressivamente forme di tutela per il riconoscimento delle identità non binarie. Si pensi, ad esempio, alla Colombia, dove la Corte colombiana si è pronunciata sul tema in occasione della sentenza T-033/22. Per saperne di più si v. S. OSELLA, *Come evolve il diritto all'identità di genere?*, cit.; S. OSELLA, R. RUBIO MARIN, *Gender recognition at the crossroads: Four models and the compass of comparative law, in International Journal of Constitutional Law*, 21, 2, 2023, 574-602.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

4. Come risponde l'algoritmo?

L'evoluzione scientifica, dottrinale e politica della categoria del genere, nonché delle rivendicazioni di riconoscimento che ad essa sono connesse, sollecita alcune riflessioni sulle modalità con cui la società e il diritto possono rispondere a evoluzioni e cambiamenti che necessitano di una puntuale regolamentazione normativa.

In altre parole, al legislatore non è richiesto di "disciplinare" l'esistenza delle persone *trans* e/o non binarie *prima*, quanto piuttosto di garantire il dispiegamento costituzionale della loro tutela dopo[47], tracciando dei confini che possano rendersi "sentinella" anche di effetti discriminatori derivanti dalla stessa società che, ad oggi, inevitabilmente si intreccia con gli sviluppi tecnologici dei sistemi di intelligenza artificiale.

L'esistenza di persone che negano, mettono in discussione, modificano e incidono sulla struttura binaria della propria identità si scontra con sistemi che fanno da eco ad una impostazione, viceversa, strettamente binaria della società e dei rapporti tra individui.

Il tema dell'interazione tra le persone *trans* e i sistemi di intelligenza artificiale si inserisce a pieno titolo nel dibattito dottrinale attorno alla presunta neutralità dell'algoritmo e, soprattutto, alla categorizzazione dei nuovi dogmi di diritto antidiscriminatorio[48] che possano al meglio definire e ricondurre a sistema i trattamenti differenziati derivanti dall'intelligenza artificiale.

Il superamento della dimensione prettamente binaria della categoria del genere pone sfide nuove al funzionamento dei sistemi di intelligenza artificiale, che sempre più dovranno trovare spazio nel dibattito dottrinale e non solo.

Nello specifico, ciò che si verifica per le persone *trans* che entrano in contatto con determinati sistemi di IA è un disallineamento tra come l'algoritmo costruisce l'identità binaria di genere e come la persona si autodetermina rispetto alla propria identità di genere.

Il già citato *automatic gender recognition* rappresenta un esempio importante di come i nuovi sistemi di IA siano costruiti su idee standardizzate della nozione di genere, ponendosi alla base di conseguenze pregiudizievoli per il diritto alla propria identità di genere.

Come noto, il riconoscimento automatico è un sistema di intelligenza artificiale ampiamente utilizzato in diversi ambiti, affinché possa velocizzare e semplificare processi di identificazione delle persone[49] che richiederebbero altrimenti maggior tempo e maggiori elementi[50].

Tuttavia, se nella maggior parte dei casi la tecnologia richiede che la persona inserisca i dati relativi alla propria identità, diverso è ciò che accade con l'AGR, in quanto esso rimuove l'opportunità per la persona di auto-identificarsi, deducendo, invece, il genere di appartenenza attraverso altri dati raccolti: questa tecnologia utilizza informazioni come il nome legale, la fisionomia facciale, il modo di parlare,

[47] Corte cost., sent. 161 del 1985, in cui si afferma che il legislatore deve limitarsi a disciplinare gli effetti di una situazione preesistente [*rectius*, la «sindrome transessuale»].

[48] In tal senso, C. NARDOCCI, *Intelligenza artificiale e discriminazioni*, cit.

[49] Sebbene, come si dirà in seguito, l'*AI Act* adottato dall'Unione europea abbia indicato i sistemi di riconoscimento facciale come sistemi "ad alto rischio", essi sono legittimi laddove usati per ragioni di sicurezza e sorveglianza pubblica.

[50] Sono molteplici i casi di sistemi di AI che "discriminano". Sul tema, si v. M. D'AMICO, *Una parità ambigua. Costituzione e diritti delle donne*, Milano, 2020, 316 ss.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

la scelta di indossare o meno *make up* per poter ridurre l'identità di genere della persona ad una mera dicotomia semplicistica[51].

Un sistema così impostato rischia potenzialmente di compromettere e ledere il diritto all'identità di genere delle persone, soprattutto *trans*, nella misura in cui i sistemi di IA riflettono schemi binari su cui è costruito il concetto di identità di genere.

In realtà, le potenzialità discriminatorie dei sistemi di intelligenza artificiale si sono già trasformate in azione. Si pensi, per fare qualche esempio, all'iniziativa promossa dalla metropolitana di Berlino che, in occasione dell'8 marzo, ha permesso alle donne di viaggiare gratis. In concreto, però, ciò ha permesso solo alle *donne* identificate come tali dall'intelligenza artificiale di viaggiare gratis[52]: l'utilizzo di sistemi di intelligenza artificiale costruiti su struttura binarie del genere ha determinato l'apertura dei tornelli soltanto a coloro che, in conformità a parametri stereotipati, venivano riconosciute come *donne*.

Ciò accade in quanto gli AGR sono generalmente costruiti attraverso una consapevolezza stringente del genere: ad esempio, attraverso i sistemi di riconoscimento facciale, coloro che hanno i capelli corti saranno più probabilmente identificati come uomini, mentre coloro che indossano *make up* saranno identificate come donne. Tecnologie di tale portata fanno infatti affidamento su presunzioni del genere inteso in termini strettamente biometrici e dicotomici, come può essere la struttura ossea oppure la forma del viso.

L'esempio riportato aiuta a comprendere come quanto viene "leso" dai sistemi di IA non è tanto (e solo) l'accesso ad un servizio o ad una prestazione, quanto, a monte, il diritto delle persone di autodeterminarsi – e soprattutto di vedersi riconosciute in base alla propria identità. L'esaurimento dell'identità di genere a categorie biometriche binarie comporta delle problematicità rilevanti in termini di autodeterminazione e di diritto all'identità personale per tutte e tutti ma, in particolar modo, per le persone *trans* che, in sistemi di intelligenza artificiale così costruiti, non solo vengono *misgendered*[53], ma negati nella loro esistenza[54].

In generale, come noto, gli effetti potenzialmente lesivi e discriminatori dei sistemi di intelligenza artificiale deriverebbero dalla loro natura interamente umana[55] e, nello specifico, dalla loro stretta connessione con il mondo maschile. Di conseguenza, non sarebbe l'intelligenza artificiale per sua natura a ledere i diritti fondamentali, quanto piuttosto il singolo individuo che, attraverso l'intelligenza artificiale e il bagaglio culturale che ha portato alla sua programmazione, discrimina[56].

Tuttavia, se, da un lato, i *gendered data* sono funzionali alla comprensione di disparità, diversificazioni e discriminazioni basate sul *genere*, soprattutto al fine di costruire strategie e politiche di promozione

[51] Per saperne di più, O. KEYES, *op. cit*., 4.

[52] La notizia di cronaca richiamata è consultabile al https://www.wired.it/attualita/tech/2021/04/26/lgbt-algoritmi-genere-orientamento-sessuale/ (ultima consultazione 2/12/2024).

[53] Termine inglese utilizzato per indicare le situazioni in cui viene attribuito o utilizzato un genere errato.

[54] Diversi studiosi hanno iniziato a domandarsi come gli algoritmi opprimono e discriminano categorie di individui già suscettibili di discriminazione nella società. Tra questi, O. KEYES, op. cit.; S.U. MOBLE, *op. cit*., la quale sostiene che gli algoritmi discriminano in quanto riflettono negativamente i *bias* già esistenti all'interno della società.

[55] K. CRAWFORD, *The Hidden Biases in Big Data*, in *Harvard Business Review*, 1 aprile, consultabile al link https://hbr.org/2013/04/the-hidden-biases-in-big-data, 2013 (ultima consultazione 2/12/2024).

[56] M. D'AMICO, *Parole che separano. Linguaggio, Costituzione, Diritti*, Milano, 2023, 122.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

dell'inclusione, dall'altro lato appare diverso laddove tali sistemi siano indirizzati alle persone *trans*, in quanto gli AGR «operazionalizzano costantemente il genere in modo *trans*-escludente»[57].

Dal punto di vista del diritto antidiscriminatorio, è interessante provare a sistematizzare la discriminazione operata dai sistemi di AGR nei confronti delle persone *trans*. Accogliendo la nozione di *AI-derived discrimination*[58] e l'idea che gli effetti discriminatori che discendono dalla costruzione ovvero dall'utilizzo dei sistemi di IA sono propri di una nuova categoria antidiscriminatoria, non potendo – per le sue peculiarità – essere ricondotta ed esaurita nelle categorie "classiche" della discriminazione diretta o indiretta[59] – occorre domandarsi dove si colloca la discriminazione derivante dai sistemi di AGR.

Si parla, a tal proposito, di *proxy discrimination*[60]. Nel quadro del concetto di *proxy discrimination*, occorre soffermarsi anzitutto sulla nozione di *proxy*. In questo senso, si argomenta in dottrina che il *proxy* può essere definito come «un elemento, come ad esempio una qualità che definisce gli esseri umani, che è utilizzato dai sistemi di intelligenza artificiale per distinguere gli individui e/o i gruppi sociali»[61]. La *proxy discrimination* nasce, così, dalla presenza dei *data-set* di "*redundant encodings*", cioè da ipotesi in cui l'appartenenza ad un determinato gruppo o categoria protetta risulta «codificata da altri dati, che risultano però associati alla medesima categoria protetta»[62]. Il *proxy* lavora attraverso meccanismi di associazione e correlazione creati tra i dati forniti alla macchina e la caratteristica che il sistema ricerca[63]: più dati vengono forniti, più verranno creati *proxy* utili a identificare caratteristiche predittive dell'appartenenza della persona ad un determinato gruppo[64].

Così come accade per le categorie del diritto antidiscriminatorio "tradizionale", è possibile distinguere tra *proxy discrimination* intenzionale o diretta e *proxy discrimination* non-intenzionale o indiretta[65].

Per quanto qui riguarda la forma specifica di discriminazione che nasce dalla frizione tra identità biometricamente costruita e identità autodeterminata dalla persona interessata, è indubbio che essa possa essere ricondotta alla categoria della *proxy discrimination* non-intenzionale. L'elemento da cui deriva la discriminazione non è immediatamente predittivo dell'appartenenza ad un gruppo, in quanto è l'elemento stesso (ad esempio, avere i capelli corti significa essere uomini) a escludere *indirettamente* tutte quelle persone che non lo presentano. In altri termini, il *proxy* e il fattore di discriminazione

[57] O. KEYES, *op. cit*., 14.

[58] Sul tema, si v. C. NARDOCCI, *Intelligenza artificiale e discriminazioni*, cit.; N. SCHMID, B. STEPHENS, *An Introduction to Artificial Intelligence and Solutions to the Problems of Algorithmic Discrimination*, in *ArXiv*, 2019, 130 ss.

[59] C. NARDOCCI, *Intelligenza artificiale e discriminazioni*, cit., 20 ss.

[60] In letteratura, si v. B. A. WILLIAMS, C. F. BROOKS, Y. SHMARGAD, *How Algorithms Discriminate Based on Data They Lack: Challenges, Solutions and Policy Implications*, in *Journal of Information Policy*, 2018, 78 ss.; C. NARDOCCI, *Proxy Discrimination in Artificial Intelligence: What We Know and What We Should Be Concerned About*, in *Chaire De Recherche Du Canada sur la culture collaborative en droit et politiques de la santé*, 2024.

[61] C. NARDOCCI, *Proxy Discrimination in Artificial Intelligence*, cit.

[62] Così, S. BAROCAS, A. D. SELBST, *Big data disparate impact*, in *California Law Review,* 104, 2016, 13.

[63] Per poter identificare una *proxy discrimination* è essenziale interrogarsi sull'esistenza o meno di forme di associazione o correlazione tra gli elementi di cui la macchina si nutre per operare distinzioni e uno o più fattori di discriminazione. Per saperne di più, C. NARDOCCI, *Proxy Discrimination in Artificial Intelligence*, cit. in cui l'A. sottolinea l'importanza di tracciare una correlazione anche in riferimento ai concetti di discriminazione diretta e indiretta, in quanto la proxy potrebbe essere il risultato di una correlazione diretta o indiretta.

[64] A. E. R. PRINCE, D. SCHWARCZ, *Proxy discrimination in the age of artificial intelligence and big data*, in *Iowa Law Review*, 2020, 1275.

[65] C. NARDOCCI, *Intelligenza artificiale e discriminazioni*, cit., 20 ss., 28 ss.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

non risultano immediatamente correlati, in quanto il primo non è immediatamente (o apparentemente) predittivo del secondo. Quello che accade è che la macchina costruisce una correlazione e un collegamento tra una variabile e un dato e l'associazione che ne risulta – ancorché formulata in termini "neutri" – identifica una categoria protetta sulla base di un elemento non presente, in modo esplicito, nel *data-set*[66].

Ancora più significativa, perché idonea a rappresentare in modo ancora più evidente la forma di discriminazione sofferta dalle persone *trans* da parte dei sistemi di AGR è la sotto-categoria delle c.d. *omitted variables bias*, in quanto l'algoritmo , pur non riferendosi a dati sensibili esclusi e indisponibili nel *data-set*, produce effetti discriminatori in ragione delle associazioni create a partire da altri dati che sono indirettamente predittivi dell'appartenenza del singolo categorie protette[67].

E, per quanto riguarda le persone *trans*, occorre allora domandarsi se le associazioni tra i dati riguardino le persone *trans* in quanto tali, oppure se la discriminazione che ne deriva sia l'effetto di un'associazione tra dati che riguardano altre categorie di minoranze. In altri termini, si potrebbe affermare che la discriminazione che ne deriverebbe non sarebbe fondata sull'identità *trans* in sé, ma sarebbe piuttosto il "riflesso" della discriminazione di *genere* che – attribuendo all'essere *donna o uomo* determinate caratteristiche – "esclude" e "categorizza" tutti coloro che in tali qualità o caratteristiche non sono riconducibili, finendo con incidere negativamente sul diritto di ciascuna e di ciascuno di autodeterminarsi in relazione alla propria identità.

5. Conclusioni

La crescente diffusione dei sistemi di IA e le difficoltà di "governarla" pongono problemi significativi per la tutela dei diritti fondamentali e per la tutela dei diritti delle persone *trans*, soprattutto per quanto riguarda il diritto all'identità personale e di genere.

Gli algoritmi sui quali sono costruiti i sistemi di IA si basano, come si è visto, su modelli binari, spesso densi di pregiudizi e stereotipi. Questi modelli non solo non riescono a cogliere la complessità e la fluidità del concetto di genere, ma tendono anche ad amplificare le discriminazioni già presenti nella società, tanto che in un futuro non così remoto, potrebbero inoltre fornire agli stati intolleranti strumenti di identificazione che rischiano di alimentare climi di oppressione e violenza[68].

Gli *Automatic gender recognition* sono un chiaro e significativo esempio di come tali sistemi, se modellati su concezioni e idee frutto di stereotipi, possano marginalizzare e stigmatizzare ulteriormente una categoria di individui già oggetto di oppressione e violazione di diritti, tanto da ridurre la loro identità ad una mera categoria biometrica.

Da un punto di vista normativo, l'*Artificial Intelligence Act* adottato dall'Unione europea sembra recepire le problematicità di sistemi di IA così strutturati, proponendone una distinzione in virtù del loro grado di rischio. In particolare, proprio i sistemi di riconoscimento facciale, basati su un utilizzo dei dati

[66] *Ibidem*, 29.
[67] *Ibidem*, 30.
[68] *Infra* 11.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

biometrici raccolti e creatori di distinzioni irragionevoli, sono qualificati come sistemi ad alto rischio (*unacceptable risk*) e, per questo motivo, vietati[69].

Di conseguenza, è importante e urgente garantire un corretto utilizzo, nonché una corretta programmazione delle tecnologie di IA che rispettino e promuovano i diritti fondamentali di tutte le persone, compresi coloro che appartengono a minoranze.

Ciò comporta al tempo stesso una revisione critica e una regolamentazione rigorosa dei sistemi di intelligenza artificiale, cui si aggiunge anche un impegno culturale nella costruzione degli algoritmi, così che possano riflettere le diversità e le molteplicità dell'essere umano in tutta la sua complessità.

Soltanto attraverso un approccio di inclusione e di rispetto delle diversità, nonché di comprensione dei rischi e dei pericoli dell'IA, si potrà garantire che i primi sistemi di intelligenza artificiale diventino strumenti di progresso e non di accentuazione di oppressione e di marginalizzazione.

[69] Regolamento UE 2024/1689, Annex III.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Il ruolo dell'IA nella costruzione di una società rispettosa dei diritti fondamentali. Il caso di studio del filtro Bold Glamour di TikTok

Fabiana Ciccarella, Lucrezia Fortuna, Elisabetta Lambiase, Mattia Mogetti*

AI'S ROLE IN BUILDING A SOCIETY THAT RESPECTS FUNDAMENTAL RIGHTS. THE CASE STUDY OF THE TIKTOK BOLD GLAMOUR FILTER

ABSTRACT: This article explores the role of Artificial Intelligence (AI) in building a society that respects fundamental rights, using TikTok's "Bold Glamour" filter as a case study. It analyses the ethical and social implications of AI-driven beauty filters, highlighting how they impact self-perception and reinforce unrealistic beauty standards. The research emphasises the need for a balanced approach in regulating AI, protecting consumer rights, and promoting ethical use. It discusses the European Commission guidelines and the importance of transparency, diversity, and non-discrimination in AI systems. The study concludes by proposing legislative, technological, and social interventions to mitigate potential harms

KEYWORDS: Artificial Intelligence; Ethical implications; Beauty filters; User vulnerability; Regulatory framework.

ABSTRACT: L'articolo esplora il ruolo dell'Intelligenza Artificiale (IA) nella costruzione di una società che rispetti i diritti fondamentali, utilizzando il filtro "Bold Glamour" di TikTok come caso di studio. L'articolo analizza le implicazioni etiche e sociali dei filtri di bellezza basati sull'IA, evidenziando il loro impatto sulla percezione di sé e come rafforzino standard di bellezza irrealistici. La ricerca sottolinea la necessità di un approccio equilibrato nella regolamentazione dell'IA, che protegga i diritti dei consumatori e promuova un uso etico. Esamina le linee guida della Commissione europea e l'importanza della trasparenza, della diversità e della non discriminazione nei sistemi di IA. Lo studio si conclude proponendo interventi legislativi, tecnologici e sociali per mitigare i danni potenziali.

PAROLE CHIAVE: Intelligenza artificiale; Implicazioni etiche; Filtri di bellezza; Vulnerabilità dell'utilizzatore; Quadro normativo.

* *Dottorand* di ricerca in Gender Studies, Università di Bari. Mail: fabiana.ciccarella@uniba.it, lucrezia.fortuna@uniba.it, elisabetta.lambiase@uniba.it, mattia.mogetti@uniba.it.* *L. Fortuna ha curato il paragrafo I; M. Mogetti ha curato il paragrafo II; E. Lambiase ha curato il paragrafo III; F. Ciccarella ha curato il paragrafo IV. Contributo sottoposto a doppio referaggio anonimo.*

Downloaded from www.biodiritto.org.
ISSN 2284-4503

SOMMARIO: 1. Il filtro *Bold Glamour*: un primo inquadramento alla ricerca di una socialità etica – 1.1. Etica e diritto per la costruzione di un'intelligenza artificiale umano centrica - 2. Relazioni scisse e iperrealtà di genere - 3. Implicazioni etiche e tutela dei diritti fondamentali nell'utilizzo del filtro *Bold Glamour* - 4. Alcuni cenni conclusivi

1. Il filtro *Bold Glamour*: un primo inquadramento alla ricerca di una socialità etica

La rivoluzione digitale ha profondamente trasformato la vita umana: la parola «*onlife*», coniata dal filosofo Floridi, coglie pienamente un tratto essenziale del nostro vivere, l'ubiquità: per la prima volta nella storia abbiamo la possibilità di essere e agire in due dimensioni contemporaneamente, *online* e *offline*.[1] Su questo *mondo nuovo*, l'Intelligenza artificiale ha – e sempre più avrà – un impatto enorme, con ciò sollevando una moltitudine di questioni dal punto di vista etico, anche con riferimento alle grandi piattaforme digitali come *Meta* e *TikTok*: i *social media* e il crescente spazio che hanno guadagnato nella vita quotidiana plasmano le nostre interazioni sociali e la nostra percezione di noi stessi e degli altri. All'interno di questa nuova esperienza di sé e di ciò che è altro da sé, i filtri rappresentano un fenomeno sempre più diffuso e pervasivo. Si tratta di *maschere* sviluppate in realtà aumentata[2] che si applicano sul proprio volto o su quello di altre persone ritratte nella foto o nel video. Le piattaforme forniscono una vasta gamma di filtri, tra i quali quelli *bellezza*, che modificano l'aspetto fisico della persona *migliorandolo*. Questa capacità di manipolare l'immagine, non nuova certamente, pone oggi questioni del tutto inedite in ragione della facilità di accesso, della velocità di diffusione e, non da ultimo, del sempre maggiore *realismo* dei filtri: ne è un esempio *Bold Glamour*, oggetto del presente lavoro, un filtro iperrealistico che ha suscitato accese polemiche, dettate anche dal massivo utilizzo fattone dagli utenti. Esso tende ad assottigliare, definire e scolpire i tratti del volto, enfatizzandoli attraverso una pelle più levigata e un trucco professionale ma non marcato e un'illuminazione sofisticata: effetti che si traducono nella realizzazione di un'immagine estremamente verosimile, anche perché restano *attaccati* al volto, non scomparendo neppure se la persona si muove o frappone un oggetto tra sé e lo schermo. L'estremo realismo del filtro, e quindi la difficoltà ad accorgersi del ricorso allo stesso, non sono il solo dato preoccupante: le principali critiche muovono dal fatto che questo tende a schiarire il colore della pelle, ad adattarsi ai lineamenti al punto da rimuovere il trucco laddove riconosce tratti più mascolini e, più in generale, a promuovere un modello estetico occidentale binario, escludente e inarrivabile[3].

Sebbene l'uso dei filtri possa apparire innocuo, è importante non sottostimarne i potenziali effetti negativi, *in primis* sulla percezione di sé e sulla salute mentale degli utenti, profondamente frustrati da canoni estetici resi desiderabili dalla continua esposizione, immediatamente accessibili *online* eppure irraggiungibili *offline*. Ciò appare ancor più vero laddove si consideri che i sistemi di IA sono ad oggi largamente addestrati sulla base di dati che riflettono pregiudizi e disuguaglianze sociali, e dunque in grado di perpetuare stereotipi e discriminazioni accrescendo il rischio di esclusione di tutte quelle

[1] Si v. L. FLORIDI, *Etica dell'intelligenza artificiale, Sviluppi opportunità, sfide*, Milano, 2022.

[2] Per una individuazione più puntuale della definizione si rinvia al II paragrafo.

[3] M. MANFREDI, *ClioMakeUp e le altre imprenditrici beauty contro l'uso dei filtri Instagram*, in *La Repubblica*, disponibile online al https://www.repubblica.it/moda-e-beauty/2021/04/21/news/cliomakeup_contro_filtri_instagram_star_senza_filtri_social-342223446/ (ultima consultazione 10/07/2024).

Downloaded from www.biodiritto.org.
ISSN 2284-4503

persone che divergono dal modello *fintamente* universale proposto. Si pone con evidenza un problema di equità e giustizia sociale che invece potrebbe trovare proprio nella dimensione digitale un'opportunità di riaffermazione, anche alla luce della natura embrionale della normazione in materia.[4] In tal senso rappresentano uno snodo fondamentale la consapevolezza circa il ruolo trasformativo dei Sistemi di IA e l'autonomia che residua in seno alla persona nei cambiamenti così generati (paragrafo 2).

Sebbene di tutto questo si abbia sempre maggiore contezza – ne sono prova le linee guida implementate dalle piattaforme[5]– non sembra ancora adeguatamente considerato l'impatto della manipolazione *di massa* delle immagini e le complesse e multifacetiche questioni etiche che esso solleva. Come più diffusamente si dirà in seguito, le prospettive da cui è possibile guardarvi sono innumerevoli e così le soluzioni approntabili, non esauribili in una disciplina incentrata sul riconoscimento in capo alle piattaforme di doveri, e dunque sulla sola responsabilità di queste ultime. L'uso diffuso dei filtri sui *social,* infatti, porta con sé una serie di problematiche etiche e sociali che coinvolgono gli stessi utenti: diventa essenziale adottare un approccio olistico nello sviluppo e nell'uso di queste tecnologie, assicurando che contribuiscano a ridurre, piuttosto che aumentare, le disuguaglianze sociali, a partire dalla promozione di un ambiente *online* sano, inclusivo e rispettoso. Un primo piano di analisi è quello della eticità e affidabilità del filtro e dei sistemi di IA in genere: come si illustrerà nel paragrafo terzo, esistono una serie di criteri alla luce dei quali valutare se un sistema di IA possa dirsi rispondente a quei criteri, così permettendo di elaborare le risposte più opportune.

1.1 Etica e diritto per la costruzione di un'intelligenza artificiale umano centrica

Sebbene possa dirsi che l'etica venga prima della legge – e durante e dopo di questa – e che dunque anche dal punto di vista logico il discorso etico precede quello più strettamente normativo, illustrare – sia pure per brevi cenni – alcune delle soluzioni adottate e gli obiettivi cui esse mirano intervenendo sulla materia, appare utile per coglierne portata e limiti al fine di orientare lo stesso dibattito etico che ne è alla base. Come si vedrà più compiutamente nel prosieguo, gli interventi sembrano limitarsi a *tamponare* gli usi distorti o discriminatori che possono essere fatti dell'IA, senza cogliere il quadro d'insieme: definirne i tratti consente al piano etico di svilupparsi anche laddove la legge non arriva. Non è sufficiente che gli algoritmi seguano le regole, è necessario che la persona e il suo benessere

[4] A tal proposito si rinvia a S. Ferilli, E. Girardi, C. Musto, M. Paolini, P. Poccianti, S. Pochettino, G. Semaforo, *L'Intelligenza Artificiale per lo Sviluppo Sostenibile*, Roma, 2021 disponibile al link: https://www.cnr.it/sites/default/files/public/media/attivita/editoria/VOLUME%20FULL%2014%20digital%20LIGHT.pdf.

[5] A esempio: https://www.facebook.com/help/instagram/477434105621119 (ultima consultazione 10/07/2024). E in specie le regole e i programmi per la creazione di una *comunità solidale*, in cui si fa riferimento esplicito alla lotta contro il bullismo e la pressione per la perfezione, alla promozione dell'accettazione del corpo; colpisce il riferimento alla «promozione dell'empatia per uno spazio Internet migliore e più inclusivo». Ancora, come si illustrerà meglio in seguito, le linee guida di TikTok: «TikTok ha otto pilastri della *community* che scaturiscono dal nostro impegno al rispetto di sicurezza e diritti umani. A tali principi si ispirano le nostre attività giornaliere e il nostro approccio alle decisioni difficili, inerenti all'applicazione di regole. I principi sono incentrati su questi temi: bilanciare l'espressione creativa e la prevenzione dei danni; sostenere la dignità umana; garantire che le nostre azioni siano eque», https://www.tiktok.com/community-guidelines/it/(ultima consultazione 10/07/2024).

Downloaded from www.biodiritto.org.
ISSN 2284-4503

restino il fine, mentre ad oggi questi sembrano piuttosto relegati a mezzi.[6] La sfida diventa non solo l'elaborazione di una normativa rispettosa dei diritti della persona, si pensi alle disposizioni sull'utilizzo dei dati e sulla *privacy*, ma anche e soprattutto di una legislazione capace di orientare l'IA verso la promozione attiva di quei diritti.[7] Si deve, in sostanza, rivendicare la centralità di un approccio all'intelligenza artificiale che sappia rispondere a finalità ulteriori rispetto a quelle meramente economiche e scientifiche, e segnatamente sicurezza, trasparenza e garanzia dei diritti e delle minoranze, in una prospettiva più ampia di tutela della persona che passa anche attraverso la protezione della sua salute mentale per così dire *umano centrica*. Emblematica in tal senso l'approvazione della legge europea sull'intelligenza artificiale (*AI Act*),[8] volta a «proteggere i diritti fondamentali, la democrazia, lo Stato di diritto e la sostenibilità ambientale dai sistemi di IA ad alto rischio, promuovendo nel contempo l'innovazione e assicurando all'Europa un ruolo guida nel settore». La normativa, non ancora in vigore, si candida a compiere un passo storico nella disciplina dell'IA: è la prima ad affrontare specificamente il tema.[9] Elemento rilevante è, tra gli altri, la previsione dell'obbligo di etichettare espressamente e chiaramente ogni immagine e contenuto audio o video artificiale o manipolato (*deep fake*) come tale, coerentemente con gli obiettivi di trasparenza che le norme pongono. Il fine ultimo della legge è tenere insieme intelligenza artificiale e valori fondanti dell'Unione e degli stati membri, rispondendo ad una esplicita richiesta della cittadinanza europea: alla base vi sono, infatti, anche le proposte elaborate dalla Conferenza sul futuro dell'Europa (COFE)[10]. Non a caso tra i suoi obiettivi c'è

[6] In questo senso, A. D'Aloia a proposito della direzione intrapresa dai primi atti paranormativi in tema di IA: «Si lavora essenzialmente in chiave di adattamento interpretativo dei principi consolidati del costituzionalismo alle applicazioni di AI: l'obiettivo è quello di configurare un'AI «*human-centric*» and «t*rustworthy*», appunto perché connotata da un "*ethical purpose*", e da un obbligo di conformità ai valori fondamentali della convivenza civile (rispetto della dignità umana e dei diritti, eguaglianza e non discriminazione, non interferenza rispetto ai processi democratici, sicurezza, rispetto per la privacy), e al benessere delle persone», in *Il diritto verso "il mondo nuovo".Le sfide dell'Intelligenza Artificiale,* in *BioLaw Journal – Rivista di BioDiritto*, 1, 2019, 6.

[7] Emblematico in questo senso anche G. Mobilio, *La co-regolazione delle nuove tecnologie, tra rischi e tutela dei diritti fondamentali*, in *Osservatorio sulle fonti*, 1, 2024, 260, disponibile in: http://www.osservatoriosullefonti.it.

[8] Si tratta del Regolamento UE 2024/1689 del Parlamento Europeo e del Consiglio del 13 giugno 2024 che stabilisce regole armonizzate sull'intelligenza artificiale. Si rinvia al paragrafo 3 e https://www.europarl.europa.eu/news/it/press-room/20240308IPR19015/il-parlamento-europeo-approva-la-legge-sull-intelligenza-artificiale (ultima consultazione 10/07/2024).

[9] Sebbene già molti paesi siano intervenuti in materia, come sottolineato da T. E. Frosini: «Vale la pena ricordare come più di 30 Paesi nel mondo – fin dal 2017 il Canada, il Giappone, la Cina e la Finlandia – hanno previsto e poi adottato una strategia nazionale per lo sviluppo dei sistemi di IA: a conferma di come la maggior parte delle economie sviluppate attribuisca alla IA un significato e un valore davvero rivoluzionario, che incide significativamente sulla crescita economica, sociale, occupazionale e culturale del Paese», T. E. Frosini, *L'orizzonte giuridico dell'intelligenza artificiale*, in *BioLaw Journal – Rivista di BioDiritto*, 1, 2022, 158.

[10] La Conferenza sul futuro dell'Europa ha rappresentato un importante rafforzamento del processo democratico europeo: hanno potuto prendere parola e discutere proponendo al termine del percorso apposite raccomandazioni. La conferenza si è svolta dall'aprile 2021 al maggio 2022, quando è stata presentata la relazione finale contenente 49 proposte e 326 misure rivolte alle istituzioni. https://www.consilium.europa.eu/it/policies/conference-on-the-future-of-europe/ (ultima consultazione 10/07/2024). Si segnalano, riguardo l'AI, la raccomandazione sul rafforzamento della competitività dell'UE nei settori strategici e la promozione dell'innovazione digitale e su una società sicura e affidabile, tra cui la lotta alla disinformazione e la garanzia di un controllo umano di ultima istanza e sull'uso affidabile e responsabile dell'IA, stabilendo salvaguardie e garantendo la trasparenza, e alla proposta sull'utilizzo dell'IA e degli strumenti digitali per migliorare l'accesso dei cittadini

Downloaded from www.biodiritto.org.
ISSN 2284-4503

anche la lotta alla discriminazione. Mentre nell'AI Act appare più pacifica la soluzione dell'equazione tecnologia e diritto, la tensione tra dimensione economica e tutela della persona sembra ancora irrisolta nelle prime risposte date ad altre questioni poste dall'utilizzo dei filtri, a partire dalla specifica regolamentazione di cui si sono dotate le piattaforme su cui questi ultimi vengono impiegati, come Meta[11]e Tiktok. Un primo piano di intervento ha riguardato il loro impiego in ambito commerciale da parte di *influencer, brand* e aziende: si parla della promozione e pubblicizzazione di prodotti e/o servizi. Qui l'alterazione prodotta dall'impiego dei filtri rileva per la potenziale distorsione della concorrenza mediante condotte ingannevoli, in quanto capace di promuovere presso i consumatori prodotti la cui efficacia è potenziata dall'uso dei filtri stessi e, dunque, di trarli in inganno. Ciò ha portato nel Regno Unito all'intervento dell'*Advertising Standard Authority* (*ASA*), autorità di autodisciplina pubblicitaria: l'ente ha affermato che «l'uso dei filtri nella comunicazione commerciale inerente a determinati prodotti potrebbe aumentare l'effetto del prodotto cosmetico risultando così fuorviante per i consumatori, configurandosi in tal modo una condotta di pubblicità ingannevole» e per questo *influencer* e *advertiser*, che promuovano prodotti cosmetici, sono tenuti a segnalare l'utilizzo di filtri su foto e video direttamente inerenti al prodotto o servizio pubblicizzato e a non applicare a tali contenuti quei filtri laddove questi possano esagerare l'effetto *naturale* del prodotto[12]. La Norvegia è invece intervenuta sulla propria legislazione[13] rendendo illegale la pubblicazione di foto e video ritoccati. La sanzione può persino arrivare alla pena carceraria: ciascun contenuto pubblicitario che contenga alterazioni alla forma o al peso del corpo ovvero al colore della pelle dovrà obbligatoriamente essere segnalato attraverso apposite etichette predisposte dallo stesso legislatore[14].

Alla luce di quanto sinora detto, appare chiaro come gli interventi più puntuali sembrino concentrarsi sulla portata del fenomeno dei filtri in termini di concorrenza e influenza sulle scelte dei consumatori, oltreché sul solo momento dell'uso: è necessario, tuttavia, riflettere altresì sulla dimensione sociale dell'utilizzo dei filtri e sulla loro fase di elaborazione. Pur se importante, non appare decisivo rendere

alle informazioni, comprese le persone con disabilità.

[11] Cfr. nota numero 6.

[12] Si v. per i dettagli: The (mis)use of social media beauty filters when advertising cosmetic products - ASA | CAP. Una simile presa di posizione non è stata assunta in Italia da parte dell'Istituto di Autodisciplina Pubblicitaria (IAP), d'altra parte resta applicabile la normativa in materia di pubblicità ingannevole, tra le pratiche commerciali scorrette sanzionate ai sensi dei d.lgs. 145 e 146 del 2007. Per approfondire si rinvia a: https://www.linkedin.com/pulse/le-foto-ritoccate-su-instagram-sono-legali-norvegia-regno-alba?originalSubdomain=it, Filtri Instagram e ingannevolezza: l'autodisciplina inglese mette un freno all'utilizzo dei filtri su Instagram per la promozione di prodotti cosmetici - DGRS - Studio Legale Milano (ultima consultazione 10/07/2024). Si v. però le linee guida emanate dall'Autorità per le Garanzie nelle Comunicazioni (AGCOM) volte ad individuare le disposizioni del Testo Unico sui Servizi di Media Audiovisivi ("TUSMA") che gli *influencer* saranno tenuti a rispettare, «in vista della crescente diffusione delle attività degli influencer e della necessità di garantire una pubblicità trasparente» di recente pubblicazione: Nuove Linee Guida AGCOM per Influencer Marketing: Trasparenza e Regolamentazione (dejalex.com).

[13] Si tratta di emendamenti al *Marketing Act* del 2009.

[14] Per leggere la normativa in commento si rinvia a: https://www.stortinget.no/no/Saker-og-publikasjoner/Saker/Sak/?p=84478 (ultima consultazione 10/07/2024) di cui si riporta il passaggio che esplicita le finalità con cui il legislatore ha adottato gli emendamenti: «gli emendamenti alla legge sul marketing mirano a ridurre la pressione del corpo nella società a causa delle persone idealizzate nella pubblicità. Tra le altre cose, viene introdotto l'obbligo di etichettare la pubblicità ritoccata o altrimenti manipolata quando ciò significa che il corpo della persona nelle pubblicità si discosta dalla realtà in termini di forma del corpo, dimensioni e pelle».

Downloaded from www.biodiritto.org.
ISSN 2284-4503

conoscibile l'uso del filtro, in quanto se ne trascurano tanto gli effetti quanto i presupposti. Sul primo piano rischiano di essere perpetrati e diffusi stereotipi e *standard* di bellezza dannosi e discriminatori, frutto, del resto, del secondo piano, ossia i pregiudizi, i valori e le finalità di chi crea e diffonde simili filtri.

2. Relazioni scisse e iperrealtà di genere

Per inquadrare teoricamente le questioni etiche sollevate dal sistema di IA in esame, è necessario partire dal funzionamento della tecnologia che ne è alla base: *Generative Adversarial Network* (GAN)[15], tecnologia di apprendimento automatico che mette in competizione tra loro diverse reti neurali. *Bold Glamour* è un processo bifasico, in cui gli elementi del volto ripreso dalla fotocamera sono confrontati con un set di immagini e con esso successivamente ricombinati. Due sono le questioni principali sollevate: la sua estrema efficacia e l'adeguamento delle più diverse fisionomie a un unico *standard*[16], che trovano fondamento nell'innovazione apportata dalla tecnologia all'opera. Non abbiamo più infatti elementi sovrapposti più o meno bene al dato reale, come accadeva con i filtri precedenti, ma la produzione di un'immagine completamente nuova. Il filtro tende a ricondurre i volti reali a un unico modello, non solo sotto l'aspetto fisiognomico, ma anche culturale: applica un certo stile di trucco e lo fa soltanto con i volti che decodifica e poi ri-codifica – in questo doppio movimento di confronto e rielaborazione – come femminili. La realtà viene così parametrata, rapportandola a dei termini di riferimento che contengono un giudizio di valore – il filtro è inteso come migliorativo – e con essi armonizzata.

Prendendo le mosse da queste premesse, è ora possibile individuare alcuni punti di riferimento concettuali utili per tracciare un fondamento teorico all'interpretazione e valutazione del caso di studio. Si parla di GAN[17] anche in relazione al fenomeno dei *deep fake*, quel sistema di IA che consiste nel sintetizzare immagini verosimili al punto da essere difficilmente distinguibili dalla realtà[18], spesso associate alla diffusione di *fake news*. E queste tecnologie, come *Bold Glamour*, possono essere lette a partire dalla nozione di iperrealtà[19]. Il concetto, teorizzato nel 1981 dal sociologo e filosofo francese Baudrillard in *Simulacres et simulation*[20], fa riferimento alla creazione di modelli del reale la cui origine non si può più rintracciare nello stesso. Il contenitore non sostituisce, ma prevale sul contenuto, il significante sul significato, la rappresentazione sul suo oggetto.

Le tecnologie di IA sono addestrate a partire da dati che nel reale hanno di fatto origine, se non altro perché prodotti a partire da un esercizio dell'intelligenza umana e perché tesi a essere riconoscibili e intelligibili dalla stessa: come nel caso di *Bold Glamour*, la persona nell'atto di rappresentarsi attra-

[15] W. Pendergrass, *Artificial intelligence and its potential harm through the use of generative adversarial network imagine filters on Tiktok*, in *Issues in Information Systems*, 24, 2023, 113-127.

[16] *Infra.*

[17] W. Pendergrass, *op.cit*.

[18] Cfr. M. Westerlund, *The Emergence of Deepfake Technology: A Review,* in *Technology Innovation Management Review*, 9, 2019, 39-52.

[19] Cfr. J. Nosta, *The 'Scientific Simulacra': When AI and hyperreality Collide*, 2023 https://johnnosta.medium.com/the scientific-simulacra-when-ai-and-hyperreality-collide-b676eb265045 (ultima consultazione 07/07/2024).

[20] J. Baudrillard, *Simulacres et simulation*, Paris, 1981.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

verso fotografie e video, insieme al set di dati di riferimento, prodotto in un momento precedente e ora disponibile per essere trasformato. Tuttavia, l'oggetto iperreale è caratterizzato da un'autonomia dai suoi *input* senza precedenti, finendo per imporsi sul reale stesso: la simulazione si colloca, cioè, sullo stesso piano delle informazioni che elabora, modificando così l'intero sistema con cui l'utente non solo si interfaccia, ma di cui fa parte. Pertanto, abbiamo un soggetto umano che cede a una macchina la sua *agency*, la sua capacità di agire rispetto alle proprie forme di partecipazione al mondo sociale – pratica che richiama ciò che Floridi[21] definisce «potere di scissione» dell'IA e che descrive come un «divorzio senza precedenti tra l'intelligenza e la capacità di agire».

E in tal senso la dimensione di genere, nei filtri *social*, costituisce un ampio riferimento per la capacità di agire. Poiché il genere è qualcosa che «si fa», che si performa, in un infinito processo di negoziazione e di *feedback* tra l'individuo e il contesto, tra il genere come categoria di assegnazione e il genere come identificazione. I sociologi West e Zimmerman, con lo storico articolo *Doing gender*[22], hanno posto una pietra miliare negli studi sul genere, interpretato attraverso la lente della pratica e della reiterazione, contestando l'assunto per cui «*doing gender merely involves making use of discrete, well-defined bundles of behavior that can simply be plugged into interactional situations to produce recognizable enactments of masculinity and femininity*» e affermando che «*doing gender is not so easily regimented [and] to be successful, marking or displaying gender must be finely fitted to situations and modified or transformed as the occasion demands*» [23].

Si potrebbe, dunque, sostenere che l'utilizzo di *Bold Glamour* definisca le condizioni stesse della partecipazione dell'utente, che non ha più bisogno di padroneggiare – e persino incarnare – alcuni codici necessari a inserirsi nella rete sociale di riferimento. Legandoci alla riflessione dei biologi ed epistemologi Maturana e Varela, in relazione alla nozione di sistemi autopoietici[24], nonché all'elaborazione portata avanti nell'ambito dell'*Actor-Network Theory* dal filosofo Latour[25], potremmo affermare che l'IA favorisca l'accoppiamento dell'utente a un determinato gruppo legato da linguaggi e valori comuni, contribuendo alla riproduzione del sistema sociale stesso.

La nozione di sistema autopoietico, infatti, pone l'accento sui *feedback* che continuamente riproducono e, allo stesso tempo, ridefiniscono il sistema. L'IA, così come trova applicazione nei GAN, vede in azione il medesimo principio: ciò che desideriamo evidenziare con questo contributo è che il soggetto umano, nel suo interfacciarsi con l'IA, entra a sua volta a far parte del sistema e del processo trasformativo che ne coinvolge tutte le componenti, contestualmente quale attore e oggetto del mutamento.

La riflessione su *Bold Glamour* permette così di evidenziare la dimensione oggettuale della soggettività che può, dunque, essere trattata come un codice linguistico: l'IA come la conosciamo oggi, infatti, è stata resa possibile dalla svolta statistica di Jelinek[26], innovazione che, laddove applicata alla fisicità umana, con tutti i suoi marcatori sociali, può dare adito a un meccanismo di *feedback* normativo e

[21] L. Floridi, *op. cit.*, 26 e ss.

[22] C. West, D. H. Zimmerman, *Doing gender*, in *Gender & Society*, 2, 1987.

[23] C. West, D. H. Zimmerman, *op.cit*, 135.

[24] H. R. Maturana, F. J. Varela, *Autopoiesi e cognizione. La realizzazione del vivente*, Venezia, 2022.

[25] B. Latour, *Reassembling the social: An introduction to actor-network-theory*, Oxford, 2007.

[26] J. McMahon, F.J. Smith, *A Review of Statistical Language Processing Techniques*, in *Artificial Intelligence Review*, 12, 1998, 347–391.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

dettare i criteri dell'inclusione e dell'esclusione. In questo emerge la natura di sistema socio-tecnico[27] dell'IA, che non solo assume un ruolo attoriale, ma potrebbe persino giungere a declassare l'utente nel suo fare e partecipare. Gli effetti di questo tipo di pratiche attengono quindi alla sfera dell'inclusione sociale: resta da determinare il ruolo dei sistemi di IA in relazione allo sviluppo della persona.

Appare però evidente come, tornando a Baudrillard, il simulacro - che qui possiamo intendere come un *avatar* - e i suoi *standard* di leggibilità abbiano un impatto ampio e profondo sulla realtà sociale e sui processi di soggettivazione che vi intervengono, in particolare, in questo contesto, rispetto al sistema di genere, che trova nelle applicazioni di IA un nuovo e potente dispositivo disciplinare.

3. Implicazioni etiche e tutela dei diritti fondamentali nell'utilizzo del filtro Bold Glamour

Le tecnologie di IA, dunque, potrebbero certamente paragonarsi a prodotti *prêt-à-porter*, essendo riproducibili, utilizzabili e trasportabili da chiunque. Tali caratteristiche, sintomatiche di un avanzamento tecnologico inarrestabile[28] e imprevedibile[29], rendono necessario analizzare i risvolti etici e i risultati che questi sono idonei a conseguire. Non è un caso, come già anticipato *supra*, che diversi attori siano coinvolti affinché i sistemi di IA siano puntualmente disciplinati, si faccia riferimento alla rilevante attività dell'Unione europea, intenzionata a ripetere quanto avvenuto con il GDPR[30], auspicandosi l'effetto Bruxelles[31]. In tal senso, si può richiamare l'*AI Act*[32], il cui articolato, rispetto al contenuto proposto dalla Commissione europea, risulta maggiormente incentrato sulla tutela dei diritti fondamentali[33], apparendo candidarsi, ad ogni buon conto, alla concezione di un IA *human-centric* e *trustworthy*[34], seppure sembri piuttosto avventato pronunciarsi sull'efficacia delle disposizioni[35]. In estrema sintesi, l'assetto dell'articolato è basato sul rischio (RBR), che risulta il fattore fondante dello sviluppo della normativa[36], che differisce in larga misura in base al livello di rischio riscontrabile nell'impiego dell'IA[37]: a seconda della pericolosità e della predittività dei risultati dell'utilizzo di tali sistemi aumenta o diminuisce la forbice di discrezionalità dell'ideatore, nonché la libertà nel contenuto della co-regolamentazione[38]. Il Regolamento, seppur trovi la propria base giuridica negli articoli

[27] Si v. AI Act, nota 8.

[28] V. P. Contucci, *Rivoluzione intelligenza artificiale. Sfide, rischi e opportunità*, Bari, 2023.

[29] G. Mobilio, *op. cit.*

[30] Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione dei dati e che abroga la direttiva 95/46/CE, del 27 aprile 2016, in GUUE L 119, del 4 maggio 2016, 1.

[31] Da intendere l'idoneità dell'Unione europea di candidarsi come legislatore globale. Si v. A. Bradford, *The Bruxelles Effect: How the European Union Rules the World*, Oxford, 2020; E. Cirone, *op. cit.*

[32] Regolamento UE 2024/1689 del Parlamento Europeo e del Consiglio del 13 giugno 2024 che stabilisce regole armonizzate sull'intelligenza artificiale.

[33] Considerando 7 e 8 del Regolamento, cfr, E. Cirone, *op. cit.*; G. Mobilio, *op. cit.*

[34] A. D'Aloia, *op. cit*.

[35] E. Cirone, *op. cit*.

[36] Cfr. G. De Gregorio, P. Dunn, *The European risk-based approaches: Connecting constitutional dots in the digital age*, in *Common Market Law Review*, 2, 2022, p. 475.

[37] Tanto emerge chiaramente dalla suddivisione in Capi del Regolamento.

[38] Cfr. G. Mobilio, *op. cit*.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

16 e 114 TFUE – vale la pena sottolineare che i sistemi di IA sono considerati dei servizi [39] – risponde, perlomeno astrattamente, a quella concezione per cui l'IA debba essere portatrice di un *ethical purpose*[40]. Concezione che emerge parimenti da quanto evidenziato negli Orientamenti etici per un'IA affidabile[41], documento in cui la Commissione europea fornisce le linee guida, nonché la definizione di un sistema IA etico ed affidabile, il quale deve rispondere a determinate componenti sintetizzabili in *legalità*, *eticità* e *robustezza*, operanti armonicamente e capaci di sovrapporsi[42]. La guida può essere applicata a qualsiasi macchina, tra cui *Bold Glamour*: l'accessibilità del filtro è pressoché universale ed è proprio l'universalità a rendere fortemente indispensabile che questo sistema rispecchi e risponda a *standard* di eticità e affidabilità elevati, considerando che l'utenza è composta in larga misura da minori che potrebbero mal beneficiare di soluzioni volte alla modifica del proprio viso per adattarlo a *standard* artificiali e irraggiungibili di bellezza. Posti i primi ostacoli verso un'effettiva e strutturale analisi della corrispondenza dei requisiti della *check-list* del documento della Commissione[43], l'indagine proverà a comprendere se *Bold Glamour* costituisca o meno un esempio di IA affidabile ed etica. È oltremodo chiaro, come già sottolineato, che l'applicazione del filtro al volto dell'utenza comporti delle implicazioni etiche specialmente dal punto visuale dei diritti fondamentali, della affidabilità, della trasparenza[44], nonché della diversità, non discriminazione ed equità, valori che, stando al contenuto della linee guida di Tiktok, si ascrivono tra i principi della *community*. Si pone, anzitutto, il dubbio circa l'effettiva analisi dell'impatto negativo che l'effetto può sortire: quali sono le tutele poste allo scopo di evitare l'eccessiva sessualizzazione[45] degli utenti – soprattutto minori – derivante dalla modifica del loro aspetto in soggetti più maturi e piacenti? Esistono strumenti idonei a limitare la sottoposizione passiva alla visione delle immagini standardizzate e *perfezionate* dal filtro? Ad oggi, non risulta essere presente alcuna opzione che consenta all'utente di richiedere la disattivazione di tali contenuti, non esiste, in altre parole, un «pulsante d'arresto»[46]. Invero, tra le forme di tutela previste per i minori, nella sezione loro dedicata, "sicurezza e benessere dei più giovani", il rimando a filtri costruiti su sistemi di IA idonei a modificare i caratteri fisionomici della persona è assente, mancando nella lista di contenuti ascrivibili come pericolosi per la salute mentale e il benessere dei giovani, tra cui si annoverano, tra altri, i contenuti sui DCA, salvo siano di promozione verso un percorso riabilitativo, e quelli di incentivazione di pratiche di chirurgia estetica[47]. Pertanto, il filtro non sembrerebbe soddisfare i requisiti relativi ai diritti fondamentali richiesti dalla *check-list* della Commissione[48]. Ciò, tuttavia, può dirsi anche in riferimento alla sua affidabilità e alla sua riproducibilità, non appare possibile selezionare e segnalare eventuali errori commessi dall'IA in sede di applicazione dell'effetto,

[39] Cfr. M. INGLESE, *Il regolamento sull'intelligenza artificiale come atto per il completamento e il buon funzionamento del mercato interno*, in *Quaderni ASIDUE*, 2, 2024.

[40] Cfr. A.D'ALOIA, *op. cit*.

[41] Commissione europea, *Orientamenti etici per un'IA affidabile*, Bruxelles, 2019.

[42] *Ibidem*, 5 e ss.

[43] Si segnala, infatti, che nei Termini e politiche di Tiktok non sono specificate le modalità di raccolta dei dati per ciò che concerne l'impiego di filtri che si applicano al volto.

[44] Sulla spiegabilità T.E. FROSINI, *op.cit.*

[45] Cfr. W. PENDERGRASS, *op. cit.*

[46] Commissione europea, *op. cit.*, 34.

[47] Sezione "Salute mentale e comportamentale" in "Salute e benessere dei più giovani".

[48] Commissione europea, *op. cit.*, 34.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

es. la diversificazione del filtro tra donne e uomini, oltre ad apparire estremamente binaria e discriminatoria, è risultata fallace anche nel non discernere chi si definisce donna e chi uomo a causa dei propri tratti più *femminili* o *maschili*. Il sistema è, dunque, poco preciso e trasparente – requisiti per l'ottenimento di un'IA affidabile –: nel valutare l'opportunità di applicare o meno il *make-up*, il sistema non è affatto spiegabile – appare ascrivibile tra i sistemi noti come *black box*[49] – non vengono ricostruite le ragioni poste alla base della decisione. La carenza di spiegabilità è naturale portatrice di problematiche, essa non solo risulta fondamentale per rispondere ai requisiti di trasparenza[50], bensì è posta alla base dell'eventuale sindacabilità del sistema di IA[51], nonché degli strumenti di educazione dell'algoritmo[52].

Quanto finora premesso non risulta affatto incoraggiante se si considera che non sono stati posti, ancor'oggi, dei correttivi ai diversi *bias* ivi presenti. *Bold Glamour*, invero, fallisce appieno il test sulla diversità, non discriminazione ed equità richiesti dalla *check-list*. È chiaro – o perlomeno, sperabile, laddove non si voglia accettare l'intenzionalità della proposizione di *standard* di bellezza occidentali – che alcuna strategia sia stata prevista per evitare di creare, rafforzare e/o perpetrare i c.d. *unfair bias*, basti far riferimento alle più palesi discriminazioni di cui si è reso protagonista il filtro: persone nere a cui la pelle e gli occhi vengono schiariti, donne a cui viene automaticamente applicato del *make-up*, uomini a cui, al contrario, viene automaticamente eliminato, con un costante assente: le persone *queer*. *Bold Glamour*, pertanto, sembra rispondere ad una società ormai obsoleta, attanagliata dal binarismo e dalla concezione escludente e discriminatoria della bellezza. Il disegno restituito è in grado di rafforzare i cd. *confirmation bias*, pregiudizi di conferma, sulla psicologia umana: la bellezza è e dev'essere solo occidentale[53].

Peraltro, stando alla *check-list* in esame, al fine di succedere negli *standard* di trasparenza rileva finanche lo scopo dell'IA, la precisazione dello stesso e il vantaggio che l'utente può trarre dall'utilizzo, è forse d'uopo domandarsi se un effetto che propone *standard* di bellezza inarrivabili sia etico e restituisca un valore aggiunto alla società? Inoltre, di quanta conoscenza godono effettivamente gli utenti dell'utilizzo del filtro da parte di altri *users*? Seppure la sua applicazione sia rintracciabile qualora il realizzatore del video lo posti sulla piattaforma originale, il video sarà salvabile e ripubblicabile e il *disclaimer* non sarà presente, rendendo quasi impossibile a chi osserva discernere se quello è l'aspetto effettivo della persona: gli utenti non avranno la consapevolezza di interagire «con un sistema di IA»[54]. Tanto può essere contenitore di molteplici conseguenze quali attaccamento, influenza o svilimento dell'essere umano[55]. Lo scenario risulta più scoraggiante considerando che nella *policy* di Tik-

[49] Sul punto cfr. A. D'Aloia, *op. cit.*

[50] La spiegabilità dei sistemi di IA è dedicato il Capo IV dell'*AI Act*.

[51] Cfr. T.E. Frosini, *op. cit.*

[52] Cfr. A. D'Aloia, *op. cit.*

[53] Commissione europea, *op. cit*. punto 44, 15. Sulla polarizzazione dei sistemi di intelligenza artificiale si rimanda a Commissione europea, *Draft Ethics Guidelines for Trusthworthy AI*, 2018; Cfr. A. D'Aloia, *op. cit.*

[54] *Ibidem*, punto 45, 131.

[55] M. Madary, T. K. Metzinger, *Real Virtuality: A Code of Ethical Conduct. Recommendations for Good Scientific Practice and the Consumers of VR-Technology*, in *Fronters in Robotics and IA*, 3, 2016.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Tok relativa alla segnalazione di contenuti creati con l'IA[56], l'*invito* all'apposizione dell'etichetta non è richiesto se le «modifiche non alterano il significato principale del contenuto»[57].
Quanto detto è idoneo ad esacerbare le complesse vulnerabilità sofferte dagli utenti, un recente studio ha rilevato che il 94% delle donne e delle persone non binarie partecipanti hanno dichiarato di sentirsi sotto pressione e di dover assumere un determinato aspetto fisico a causa di contenuti simili[58]. Sulla base di questa analisi, risulta inevitabile escludere il *Bold Glamour* tra i sistemi di IA affidabili ed etici. Sarebbe, forse, opportuno domandarsi se esperimenti del genere siano necessari in una società già sofferente ed altamente performativa e costituiscano un valore aggiunto alla società e perseguano l'*ethical purpose* di cui appare doversi connotare qualsiasi sistema di IA?[59]

4. Alcuni cenni conclusivi

Alla luce dell'analisi sin qui condotta, preme da ultimo ragionare sulla mitigazione dei potenziali danni derivanti dall'impiego dei filtri: le prospettive di intervento sono molteplici – tecnologiche, legislative e sociali – e richiedono tanto il coinvolgimento di attori esterni quanto una consapevolezza interna da parte degli utenti.
Dal punto di vista normativo, è emerso come l'intervento legislativo spesso si dimostri inefficace se non controproducente: si veda ad esempio l'infruttuoso tentativo di alcuni paesi di contrastare l'insoddisfazione dell'immagine corporea mediante leggi che vietano la pubblicità di immagini potenzialmente fuorvianti[60] o ancora l'introduzione di *disclaimer,* che tuttavia sembrerebbero non solo non ridurre l'insoddisfazione, ma addirittura promuovere confronti sociali nocivi[61]. Dunque, il rischio è che l'azione legislativa, sebbene appaia un approccio diretto, finisca per risultare troppo ampia e lenta per affrontare il problema in modo strutturale.
Da un punto di vista tecnologico, l'eliminazione dei filtri incriminati, da parte del *provider* sembrerebbe una soluzione immediata, ma in concreto risulta di difficile attuazione, sollevando anche questioni legate al ruolo degli Internet Service Providers nella moderazione dei contenuti. Data la crescente diffusione della tecnologia IA, inclusa quella basata su GAN, la sua regolamentazione risulta poi complessa: anche se esistono meccanismi per identificare e segnalare tali filtri, come l'auto-identificazione o l'installazione di marcatori digitali, questi potrebbero essere facilmente aggirati[62]. Attualmente,

[56] Cfr. le linee guida della *community* di Tiktok, Integrità e autenticità, Contenuti multimediali modificati e contenuti generati dall'intelligenza artificiale (AIGC).
[57] *Ibidem*.
[58] R. Gill, *Changing the perfect picture: Smartphones, social media and appearance pressures*, 2021, https://www.city.ac.uk/__data/assets/pdf_file/0005/597209/Parliament-Report-web.pdf (ultima consultazione 10/07/2024).
[59] Cfr. D'Aloia, *op. cit.*
[60] G. Sharp, Y. Gerrard, *The body image "problem" on social media: Novel directions for the field*, in *Body Image,* 41, 2022.
[61] E.S. Danthinne, F.E. Giorgianni, R.F. Rodgers, *Labels to prevent the detrimental effects of media on body image: A systematic review and meta-analysis*, in *International Journal on Eating Disorder*, 53, 2024.
[62] F. Shan, A.R. Cillo, C. Cardello, D.Y. Yuan, S.R. Kunning, J. Cui, C. Lampenfeld, A.M. Williams, A.P. McDonough, A. Pennathur, J.D. Luketich, J.M. Kirkwood, R.L. Ferris, T.C. Bruno, C.J. Workman, P.V. Benos, D.A.A. Vignali, *Integrated BATF transcriptional network regulates suppressive intratumoral regulatory T cells*, in *Science Immunology*, 7, 76, 2022.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

la limitazione del tempo di condivisione dei video su TikTok contribuisce a mitigare l'impatto dei filtri, ma resta dubbia la sua utilità a lungo termine. Considerazioni analoghe riguardano la percezione delle immagini manipolate dalla tecnologia, che solleva interrogativi sull'equivalenza tra IA e l'uomo. Dunque, una soluzione tecnologica, sebbene possa sembrare praticabile, potrebbe anche in questo caso rivelarsi inefficace nel lungo periodo.

Da un punto di vista sociale esterno, l'attivismo per la consapevolezza rappresenta un approccio ampio per affrontare il problema, in questo senso vanno letti movimenti come il *body positivity* che sfidano gli ideali estetici convenzionali e promuovono l'accettazione del corpo in tutte le sue forme; del resto, l'impatto dei filtri dannosi è stato riconosciuto dalle piattaforme stesse allorché hanno provveduto alla loro rimozione. Come visto, problematico è l'uso dei filtri da parte degli *influencer* e la promozione di ideali estetici irreali, rispetto ai quali le sfide rimangono significative. Ciò che risulta cruciale è la consapevolezza individuale del contesto tecnologico e sociale circostante, ossia la capacità di riconoscere le distorsioni generate dalle tecnologie digitali e comprendere la propria relazione con le stesse. L'educazione riguardo alle alterazioni nelle rappresentazioni digitali e il sostegno reciproco tra pari possono, pertanto, contribuire a mitigare le conseguenze dannose delle immagini manipolate[63].

Risulta ancora una volta chiaro che l'approccio in tal senso debba essere sfaccettato su diversi livelli: mentre l'azione normativa e tecnologica possono fornire alcuni strumenti per mitigare il problema, la consapevolezza sociale e individuale resta fondamentale per contrastare le potenziali lesioni dei diritti fondamentali[64]. In tal senso è auspicabile un ulteriore sviluppo dell'analisi circa l'impatto dell'uso dei filtri, in specie con riferimento a quei gruppi di cui mancano ad oggi dati chiari, come persone trans* e *queer*.

Da tanto premesso, è evidente come sia ancora lontana una soluzione a lungo termine circa l'utilizzo di filtri bellezza come *Bold Glamour*, che tuttavia non può non partire dalla responsabilità del *provider* di questi strumenti, che hanno un ruolo centrale nell'immissione e gestione dei contenuti realizzati con tale filtro: la piattaforma, detenendo il controllo sui filtri disponibili e sulle politiche di utilizzo, può essere decisiva nella mitigazione dei danni[65]. Questo implica non solo la necessità di implementare sistemi di monitoraggio efficaci e di reagire prontamente alle segnalazioni di abusi, ma anche l'adozione di un approccio proattivo nella valutazione, nella limitazione e anche nell'opportunità dell'uso di filtri potenzialmente dannosi[66], implementando le *policy* della *community*. In tal senso, TikTok potrebbe collaborare con esperti di salute mentale e *stakeholders* per sviluppare linee guida più rigorose e, contestualmente, sensibilizzare gli utenti sugli effetti psicologici derivanti dall'uso eccessivo di tali filtri.

[63] M. TIGGEMANN, V.G. VELISSARIS, *The effect of viewing challenging "reality check" Instagram comments on women's body image*, in *Body Image*, 33, 2022.

[64] W. PENDERGRASS, *op. cit.*

[65] J.M. TRAMMEL, *Artificial Intelligence for Social Evil: Exploring How AI and Beauty Filters Perpetuate Colorism—Lessons Learned from a Colorism Giant*, in K. LANGMIA (a cura di), *Black Communication in the Age of Disinformation*, Londra, 2023.

[66] *Ibidem.*

Downloaded from www.biodiritto.org.
ISSN 2284-4503

La vulnerabilità del migrante nell'era delle *smart-borders* e delle tecnologie *lie-detecting*

Roberta Nobile*

THE VULNERABILITY OF THE MIGRANT IN THE AGE OF SMART-BORDERS AND LIE-DETECTING TECHNOLOGIES

ABSTRACT: At the borders, biometric recognition is being used to detect suspicious movements and anticipate the screening of migrants as they enter European borders. To this end, the EU has funded iBorderCtrl, an AI liedetecting system operated by a virtual border guard to interrogate travelers trying to cross borders by assessing facial microexpressions. Facial recognition raises serious concerns, among them the danger of surveillance and discriminatory profiling, which draw the edges of a group vulnerability. In addition, the creation of biometric records within the interoperability dimension contributes to exacerbating the vulnerability of the migrant.

KEYWORDS: Migrants; lie-detecting; biometrics; vulnerability; border control.

ABSTRACT: Alle frontiere il riconoscimento biometrico viene utilizzato per individuare movimenti sospetti e anticipare lo screening dei migranti che entrano nei confini europei. L'UE ha finanziato *iBorderCtrl*, un sistema di rilevamento delle menzogne gestito da una guardia di frontiera virtuale per interrogare i viaggiatori che cercano di attraversare le frontiere mediante la valutazione delle microespressioni facciali. Il riconoscimento facciale solleva preoccupazioni, tra cui il pericolo di sorveglianza e di profilazione discriminatoria, che disegnano i margini di una forma di vulnerabilità di gruppo. A questo si aggiunga la creazione di registri biometrici nella dimensione dell'interoperabilità, contribuendo ad aggravare la fragilità del migrante.

PAROLE CHIAVE: Migranti; lie-detecting; biometria; vulnerabilità; sorveglianza.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

* *Dottoranda di ricerca, Università Campus Bio-Medico di Roma. Mail: robertanobile110@gmail.com.* *Contributo sottoposto a doppio referaggio anonimo.*

1. Introduzione: il confine della paura o la paura del confine?

La digitalizzazione delle frontiere statali ha determinato la progressiva creazione delle frontiere intelligenti o *smart-borders*, che implementano i processi tecnologici di biometria, sorveglianza dei dati e automazione in tandem, all'interno di una esasperazione del binomio corpo-confine, decantato dalle tecnologie di *lie-detecting*, che condannano gli immigranti ad interrogazioni a cui rispondere con il corpo e non con la parola, assolvendo al ruolo di macchina della verità in veste di giudice del confine.

L'espansione della biometria nel controllo strategico militare dei flussi di persone è stata parallela alla sorveglianza biometrica delle frontiere volta a prevenire movimenti migratori indesiderati[1]. Diversi autori hanno posto in evidenza i cambiamenti indotti da nuove forme di sicurezza delle frontiere: Johnson ha osservato in che modo la militarizzazione dei confini, che va di pari passo con il crescente uso di tecnologie biometriche, abbia innescato una riarticolazione ed espansione della sovranità statale[2]. Per Longo «i confini zonali più ampi sono diventati frontiere, mentre ugualmente la sovranità assomiglia ad un imperium crescente», vale a dire un'autorità politica territorialmente illimitata[3]. Per Lyon il confine «ora è ovunque»[4].

Si parla, pertanto, di confini biometrici per descrivere il modo in cui la biometria riconfigura i margini della società e i corpi delle persone al suo interno. Studiando la sorveglianza dei dati nella guerra al terrore, è possibile mostrare come le tecniche biometriche implichino processi di oggettivazione, pratiche, cioè, che dividono e scompongono l'individuo in fattori di rischio calcolabili, trasformando, in tal modo, il soggetto in oggetto[5]. Tale oggettivizzazione si traduce in nuove tecnologie di sorveglianza che identificano "popolazioni sospette", "gruppi a rischio", separano i "cittadini" dagli "anti-cittadini" e dai "non-cittadini", disciplinano il corpo indisciplinato[6] riportandolo in una zona di calcolo e gestibilità e recuperandolo all'interno di intervalli normali di accettabilità. Le informazioni biometriche utilizzate per costruire i dati sono, inoltre, catturate dal vortice tecnologico turbinoso della *dataveillance*, ossia il processo silente e continuo di estrazione e analisi algoritmica dei dati degli individui[7]. L'obiettivo della *dataveillance* non è monitorare soggetti specifici[8], ma sorvegliare tutti per creare profili che possono

[1] L. Amoore, *Biometric borders: Governing mobilities in the war on terror*, in *Political Geography,* 25.3/2013, 336-351.
[2] C. Johnson, *et al., Interventions on the state of sovereignty at the border*, in *Political Geography,* 59/2017, 1-10.
[3] M. Longo, *The politics of borders: Sovereignty, security, and the citizen after 9/11*, Cambridge University Press, 2017.
[4] D. Lyon, *The border is everywhere: ID cards, surveillance and the other*, in *Global surveillance and policing*, 2013, 66-82.
[5] L. Amoore, *op.cit*., 340.
[6] C. Johnson, *op.cit*., 32.
[7] H. Pötzsch, *The Emergence of iBorder: Bordering Bodies, Networks, and Machines*, in *Environment and Planning D: Society and Space*, 33.1/2015.
[8] J. Van Dijck, *Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology*, in *Surveillance & society,* 12.2/2014, 197-208; S. Degli Esposti, *When big data meets dataveillance: The hidden side of analytics*, in *Surveillance & Society,* 12.2/2014; L. Everuss, *AI, smart borders and migration*, in A. Elliott (a cura di), *The Routledge Social Science Handbook of AI*, London, 2021; B.O. Martins, M.G. Jumbert, *EU Border technologies and the co-production of security 'problems' and 'solutions'*, in *Journal of Ethnic and Migration Studies*, 2020.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

essere impiegati per valutare la minaccia rappresentata dalle persone, determinando, consequenzialmente, uno spostamento delle pratiche di sorveglianza dall'esame mirato di popolazioni e di individui «alla terribile deriva del monitoraggio di massa[9]».
Il *rebordering* biometrico viene, pertanto, descritto come un esercizio di biopotere[10], mediante il quale i corpi stessi si trasformano in siti di molteplici codificazioni sociali che aprono o chiudono a possibilità e pratiche di *agency*[11].
All'interno di una progressiva depoliticizzazione dei confini, l'emersione di tecnologie di *lie-detecting*, di verdetti pronunciati da macchine biometriche della verità, contribuisce a tracciare confini invisibili e invalicabili, che disegnano e giudicano il volto del migrante, ancor prima condannato da una forma di vulnerabilità di gruppo e da una discriminazione identitaria, prigioniero di uno stigma etnico-sociale insormontabile.

2. Il progetto iBorderCtrl: il nuovo verdetto dei biomarcatori

Una delle principali caratteristiche che compongono il tessuto morfologico dei confini intelligenti è rappresentata dall'automazione, necessaria per utilizzare gli strumenti biometrici e condurre la sorveglianza dei dati, poiché la quantità di informazioni e i livelli di elaborazione richiesti per intraprendere tali processi richiedono forme di analisi guidate da algoritmi[12]. A tal riguardo, nell'ottobre 2018, l'UE aveva annunciato che stava finanziando un nuovo sistema automatizzato di controllo delle frontiere, chiamato *iBorderCtrl*, da sperimentare in Ungheria, Grecia e Lettonia. Il funzionamento di tale sistema si articola in due fasi. La prima fase prevede che al viaggiatore venga chiesto di fornire informazioni sulla sua persona e sui dettagli del viaggio. Queste indicazioni vengono, poi, verificate con vari database per determinare se sono soddisfatte le premesse per poter effettuare l'attraversamento di frontiera. Successivamente, al fine di accertare che le informazioni fornite dal viaggiatore siano corrette, la seconda fase del sistema *iBorderCtrl* prevede l'utilizzo di un sistema di *lie-detecting* di intelligenza artificiale gestito da una guardia di frontiera virtuale per interrogare i migranti che cercano di attraversare i confini, valutando al contempo i minimi dettagli delle loro espressioni facciali, note come microespressioni, utilizzando tecniche facciali e tecnologie di riconoscimento delle emozioni. La voce e il comportamento diventano più severi se il sistema sospetta che il soggetto stia mentendo. Infatti, l'avatar interagirà con il viaggiatore in modo autonomo, assurgendo al ruolo di giudice artificiale, decidendo, pertanto, quali domande porre, come comportarsi (ad esempio, può adattare il suo comportamento a quello dell'individuo interrogato, talvolta assumendo un atteggiamento piuttosto scettico se una risposta fornita sembra non essere corretta) e, infine, formulando la valutazione conclusiva riguardo il rischio complessivo derivante dal viaggiatore in base alle informazioni fornite nella fase iniziale e ai risultati del rilevamento dell'inganno, in modo da indirizzare, successivamente, il migrante

[9] D. LYON, *op. cit*., 10.
[10] L. AMOORE, *op. cit*., 345.
[11] G. KAFER, *Big Data Biopolitics*. in *Digital Culture & Society, 5*.1/2019, 23-42.
[12] B.A. RAJOUB, R. ZWIGGELAAR, *Thermal facial analysis for deception detection*, in *IEEE Transactions on Information Forensics and Security*, 9.6/2014; J.R. SIMPSON, *Functional MRI lie detection: Too good to be true?* in *The Journal of the American Academy of Psychiatry and the Law*, 36.4/2008; C. MOROSAN, *Information disclosure to biometric e-gates: the roles of perceived security, benefits, and emotions*, in *Journal of Travel Research, 57*.5/2018.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

alle guardie di frontiera umane. Tale sistema si adatta a ciò che viene descritto come un «passaggio dei regimi di sicurezza da una modalità reattiva ad una modalità proattiva, collocata al centro delle logiche statali contemporanee incentrate sulla superiorità tecnologica e sulla sorveglianza persistente»[13].

A supportare *iBorderCtrl* vi è un sistema automatizzato di rilevamento dell'inganno chiamato *Automatic Deception Detection System* (ADDS), sviluppato presso la *Manchester Metropolitan University*, con l'obiettivo di classificare i dati biometrici in linea con microespressioni non verbali facciali, considerate biomarcatori di inganno[14], in grado, cioè, di agire come predittori della menzogna. I biomarcatori dell'inganno sono codificati come 38 caratteristiche, quali, ad esempio, battito di ciglia sinistro, aumento del rossore del viso o direzione del movimento della testa. Ciascuna caratteristica è estratta da un segmento video di un secondo in cui il migrante mente o meno nel momento in cui risponde ad una precisa domanda, anche se dalla documentazione disponibile, prodotta dal team di *iBorderCtrl*, non risulta essere chiaro come vengano generate le caratteristiche per il segmento. Il video viene acquisito a 30 fotogrammi al secondo con una risoluzione video di 640x480. Ogni modello di addestramento del dataset è costituito da un vettore di 38 caratteristiche e dall'etichetta che indica la verità o l'inganno. Per creare il set di dati, a 32 partecipanti viene assegnato un ruolo, con due soluzioni possibili, "veritiero" o "ingannevole", da svolgere durante l'intervista, nel corso della quale ogni partecipante dovrà rispondere a 13 domande e ogni risposta prodotta verrà successivamente segmentata in molti vettori. Il team di *iBorderCtrl* afferma che i biomarcatori dell'inganno sono segnali che assunti singolarmente non possono rivelare un comportamento ingannevole, ma che, invece, complessivamente possono essere utilizzati da un metodo ML per rilevare le bugie dei migranti. Ciò significa che, secondo tale modello, i comportamenti giudicati ingannevoli o veritieri costituiscono due categorie, non sovrapposte, che rappresentano un insieme di stati emotivi. Nell'assegnare un'etichetta al segmento video relativo ad una risposta, l'ADDS, pertanto, considera solo tali due possibilità di categorizzazione escludendo, di conseguenza, ulteriori opzioni possibili, nonostante il modello presenti due parametri per filtrare i segmenti non significativi quando non sussiste una categoria definita e strutturata per essi. Sul solco di tali perplessità, il 5 novembre 2018, un deputato del Parlamento europeo, Patrick Breyer, aveva chiesto l'accesso ai documenti relativi all'autorizzazione del progetto *iBorderCtrl* e a quelli elaborati nel corso di tale progetto, detenuti dalla Commissione europea. Tale richiesta era stata rifiutata dall'agenzia europea responsabile di *iBorderCtrl*, l'*European Research Executive Agency* (REA), in quanto avrebbe compromesso gli interessi commerciali del consorzio, compresi i diritti di proprietà intellettuale, costringendo, pertanto, Breyer ad intentare una causa contro essa al fine di ottenere la pubblicazione di documenti riservati sulla giustificabilità etica e la legalità della tecnologia. Il Tribunale

[13] L. SUCHMAN, K. FOLLIS, J. WEBER, *Tracking and targeting: Sociotechnologies of (in) security*, in *SAGE Publications Sage CA: Los Angeles*, CA, 2017.

[14] J.W. CRAMPTON, *Platform biometrics*, in *Surveillance & Society,* 17(1/2), 2019, 54-62; J. SÁNCHEZ MONEDERO, L. DENCIK, *The politics of deceptive borders: biomarkers of deceit'and the case of iBorderCtrl*, in *Information, Communication & Society,* 25.3/2022, 413-430; L. DINGES, *et al*., *Exploring facial cues: automated deception detection using artificial intelligence*, in *Neural Computing and Applications*, 2024, 1-27; D. MINKIN, L.T. BRANDNER, *Borderline decisions? Lack of justification for automatic deception detection at EU borders*, in *TATuP-Journal for Technology Assessment in Theory and Practice,* 33.1/2024.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

dell'Unione Europea aveva emesso la sentenza il 15 dicembre 2021[15], secondo la quale la decisione della REA doveva essere annullata, nella parte in cui questa aveva omesso di statuire sulla domanda del sig. Patrick Breyer di accesso ai documenti riguardanti l'autorizzazione del progetto *iBorderCtrl* e, in secondo luogo, nella parte in cui aveva rifiutato di concedere l'accesso completo ad ulteriori documenti, nella misura in cui tali documenti contenevano informazioni non coperte dall'eccezione prevista all'art. 4, paragrafo 2 del regolamento (CE) n. 1049/2001 del Parlamento europeo e del Consiglio, del 30 maggio 2001, relativo all'accesso del pubblico ai documenti del Parlamento europeo, del Consiglio e della Commissione. Non soddisfatto da questa decisione, il 25 febbraio 2022 Breyer aveva presentato ricorso[16] sostenendo che «l'interesse pubblico alla divulgazione prevalesse sugli interessi commerciali privati[17]», con la necessità di garantire accuratezza e trasparenza durante l'intero iter procedurale della ricerca. Nella successiva decisione del 7 settembre 2023, la CGUE aveva, innanzitutto, confermato la sentenza del Tribunale stabilendo che l'interesse pubblico, che riguardava, in realtà, un'eventuale futura applicazione di sistemi basati su tecniche sviluppate nell'ambito di tale progetto, era stato soddisfatto dalla diffusione dei risultati.

Inoltre, la circostanza che i partecipanti al progetto *iBorderCtrl* siano tenuti a rispettare i diritti fondamentali e i principi riconosciuti, in particolare, dalla Carta dei diritti fondamentali dell'Unione Europea, e che la Commissione sia tenuta a vigilare sul rispetto di detti diritti e di detti principi, non è in grado di far presumere l'assenza di una qualsiasi violazione di tali diritti e principi e di escludere l'esistenza di un interesse pubblico prevalente alla divulgazione dei documenti relativi a tale progetto, a causa del possibile impatto delle tecniche utilizzate sulla protezione dei diritti fondamentali.

Sorgono, infine, domande critiche in relazione al modo in cui i ricercatori abbiano cercato di validare l'ADDS prima di condurre i programmi pilota[18]. L'esperimento di validazione fornito dal team consisteva nel testare le prestazioni del classificatore ML che sarebbe stato incluso nell'ADDS. Ciò suggerisce, però, che non sia stato effettuato un test con nuovi individui (non visti) i cui dati avrebbero dovuto essere preventivamente acquisiti, elaborati e classificati dal modello ML, generando, di conseguenza, perplessità e dubbi sull'effettiva correttezza del test del modulo ADDS. Il dataset di addestramento non soddisfa, inoltre, le ipotesi della maggior parte degli algoritmi di apprendimento automatico, poiché i campioni ottenuti da ciascun partecipante sono correlati e, quindi, i modelli generati risultano essere molto vicini nello spazio delle caratteristiche e, allo stesso tempo, molto distanti dai vettori generati per altre persone. In altre parole, i dati nello spazio delle caratteristiche risultano molto scarsi, il che può produrre diversi problemi procedurali e strutturali[19].

Pertanto, alla luce di tali considerazioni, il sistema ADDS risulta essere stato addestrato e testato non solo su un numero piccolo di soggetti, ma per di più composto prevalentemente da maschi europei, la

[15] Case T-158/19, *Breyer v. REA*, ECLI:EU: T:2021:902.
[16] Case C-135/22P, *Breyer v. REA*, ECLI:EU:C:2022:640.
[17] Case C-135/22P, *Breyer v. REA*, ECLI:EU:C:2023:640, para. 104.
[18] L. Brandner, S. Hirsbrunner, *Algorithmic fairness in police investigative work. Ethical analysis of machine learning methods for facial recognition*, in *TATuP – Journal for Technology Assessment in Theory and Practice*, 32.1/2023, 24–29; A. Selbst, *Disparate impact in big data policing*, in *Georgia Law Review*, 52.1/2017.
[19] In statistica, questo fenomeno è noto come maledizione della dimensionalità ed è particolarmente rilevante quando la dimensione del campione è inferiore al numero di dimensioni dei dati. Si veda, al tal proposito, N. Altman, M. Krzywinski, *The curse(s) of dimensionality*, in *Nat Methods*, *15*.6/2018, 399-400.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

cui conseguente sottorappresentazione di determinati gruppi, come le persone di colore o le donne, nei set di dati di addestramento dell'IA può condurre inevitabilmente a valutazioni inaffidabili riguardo gli individui appartenenti a tali gruppi e, consequenzialmente, a risultati distorti e discriminatori[20].
Oltre a tali perplessità di matrice empirica, emergono considerazioni critiche riguardo il fondamento teorico del sistema ADDS, in quanto privo di una base scientifica fondata e condivisa. Infatti, la critica epistemologica[21] evidenzia la mancanza di un consenso scientifico sull'ipotesi che le intenzioni ingannevoli possano essere dedotte dalle microespressioni, sottolineando, di conseguenza, come l'uso di sistemi di rilevamento dell'inganno non risulti giustificato a meno che non si raggiunga un riconoscimento scientifico fondato e condiviso sulle relative basi teoriche, in grado di fornire linee guida per una regolamentazione omogenea. Esistono, pertanto, disaccordi riguardo i molteplici aspetti e i diversi livelli di astrazione delle emozioni. In primo luogo, l'interpretazione delle microespressioni come indicatori di inganno non risulta essere conclusiva, poiché gli psicologi hanno formulato altre ipotesi similmente ragionevoli su ciò che le microespressioni potrebbero indicare[22], mettendo di conseguenza in discussione il fondamento del funzionamento dell'ADDS, in quanto, ad esempio, se le microespressioni non sono indicative di inganno ma di emozioni represse e soffocate nell'inconscio, il sistema non misurerà ciò che dovrebbe verificare. In secondo luogo, emerge un disaccordo psicologico sul fatto che l'analisi facciale possa fornire una lettura universale delle emozioni come stati fissi, sulla base del risultato di studi che hanno evidenziato come le espressioni emotive dipendono da fattori culturali e sociali[23], impedendo, pertanto, di giungere ad una classificazione appropriata ed omogenea delle emozioni e dei dati da associare.
Alla luce di tali considerazioni, *iBorderCtrl* costituisce sicuramente un prodotto dell'IA emozionale[24], dell'informatica affettiva, che trova espressione nel modo di affrontare l'emozione come qualcosa che può essere osservata attraverso ciò che può essere rilevato, misurato e ricordato, con il supporto e l'utilizzo di metodologie di rilevamento che classificano il comportamento facciale e corporeo. In particolare, si tratta di tecniche di rilevamento che hanno acquisito importanza in un contesto di *datafication*[25], che si esplica nella tendenza a trasformare sempre più aspetti dei fenomeni sociali e del

[20] J. Rothwell, *et al*., *Silent Talker. A new computer-based system for the analysis of facial cues to deception*, in *Applied Cognitive Psychology*, 20.6/2006; L.F. Barrett, *et al*., *Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements*, in *Psychological science in the public interest*, 20.1/2019.
[21] Si rinvia a L. Podoletz, *We have to talk about emotional AI and crime*, in *AI & Society,* 38.3/2023, 1067–1082; F. Bacchini, L. Lorusso, *Race, again. How face recognition technology reinforces racial discrimination*, in *Journal of Information, Communication and Ethics in Society*, 17.3/2019; P. Helm, T. Hagendorff, *Beyond the prediction paradigm. Challenges for AI in the struggle against organized crime*, in *Law and Contemporary Problems*, 84.3/2021.
[22] H. Elfenbein, N. Ambady, *On the universality and cultural specificity of emotion recognition. A meta-analysis*, in *Psychological Bulletin*, 128.2/2002.
[23] L. Feldman Barrett, *et al*., *Emotional expressions reconsidered. Challenges to inferring emotion from human facial movements*, in *Psychological Science in the Public Interest*, 20.1/2019; L. Zhang, O. Arandjelović, *Review of automatic microexpression recognition in the past decade*, in *Machine Learning and Knowledge Extraction*, 3.2/2021.
[24] A. McStay, *Emotional AI, soft biometrics and the surveillance of emotional life: An unusual consensus on privacy*, in *Big Data & Society,* 7.1/2020; T. Gremsl, E. Hödl, *Emotional AI: Legal and ethical challenges*, in *Information Polity,* 27.2/2022.
[25] C. Southerton, *Datafication*, in L.A. Schintler, C.L. McNeely (a cura di), *Encyclopedia of Big Data*, Cham, 2022.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

comportamento umano in formati quantificati che possono essere tabulati e analizzati. Infatti, nel sintetizzare non solo l'ideologia del "dataismo"[26], basata sulla spinosa relazione tra persone e dati, in particolare sul dominio e la supremazia dei dati sulla fragilità della dimensione umana, ma anche il simultaneo vuoto scientifico dei sistemi di riconoscimento affettivo e la cancellazione dei dubbi negli algoritmi di apprendimento automatico, *iBorderCtrl* emerge come un modello per la politica di governance basata sui dati, la manifestazione della tendenza al tecnosoluzionismo[27], una recente propensione che ha visto sia i governi che le aziende tecnologiche ricorrere a soluzioni tecnologiche, digitali, high-tech per diversi ordini di problemi, dai cambiamenti climatici alla carestia fino alla migrazione.

3. Il corpo come "doppio di dati" nel paradigma della vulnerabilità di gruppo

La crescente combinazione e integrazione delle pratiche e delle tecnologie di sorveglianza in un insieme ampio di dati e di informazioni dà luogo ad una forma di assemblaggio di sorveglianza[28], retta da muri invisibili di intelligenza artificiale che sfidano il confine corporeo. Designano, in tal modo, la convergenza di quelli che prima erano sistemi di sorveglianza discreti verso un punto in cui operano come un insieme turbinoso e offuscato, dai margini labili e incerti. Questa convergenza corrisponde all'obiettivo costante delle autorità di polizia di integrare i diversi sistemi informatici e banche dati per realizzare una forma di interoperabilità, basata sulla consultazione, sullo scambio e sulla condivisione ad ampio spettro dei dati, in grado di condurre progressivamente ad una smaterializzazione del corpo e alla costituzione di un "*doppio di dati*"[29]. L'assemblaggio della sorveglianza delle tecnologie di *lie-detecting* opera, infatti, astraendo i corpi umani dal loro contesto territoriale e scomponendoli in una serie di flussi discreti, che verranno, poi, riassemblati in "doppi di dati" con la possibilità, a loro volta, di essere sottoposti a forme di controllo e intervento.

Il corpo viene così scomposto, astratto e riassemblato attraverso i flussi di informazione: il risultato è un corpo disincarnato[30], un doppio informatico di pura virtualità, in cui l'interesse non risiede nei corpi completi ma nei frammenti di informazione che essi emanano. Questo nuovo modo di divenire trascende la corporeità umana e riduce la carne a pura informazione, producendo la moltiplicazione dell'individuo, la costituzione di un "sé" aggiuntivo[31]. Al di là della violenza implicita nella decomposizione e nella riscrittura del corpo in forma digitale, la sfida si traduce nel tentativo di reincarnare l'individuo[32] e di ripristinare la materialità fisica che è alla base e nelle conseguenze di queste reti informatiche. Come mantenere la distinzione tra il corpo e l'informazione su di esso quando il corpo stesso è costituito da informazioni: dov'è esattamente il passaggio tra la materia e l'informazione del corpo?

[26] J. VAN DIJCK, *op. cit*., 206.
[27] Cfr. D. ANDLER, *Il duplice enigma. Intelligenza artificiale e intelligenza umana*, Torino, 2024.
[28] Sul punto si rinvia: P. MOLNAR, *Technology on the margins: AI and global migration management from a human rights perspective*, in *Cambridge International Law Journal, 8*.2/2019, 305-330; G. KAFER, *Big Data Biopolitics*, in *Digital Culture & Society, 5*.1/2019, 23-42; E.L. HSU, *The sociological significance of non-human sleep*, in *Sociology, 51*.4/2017, 865-879.
[29] C. EPSTEIN, *Guilty bodies, productive bodies, destructive bodies: Crossing the biometric borders*, in *International Political Sociology* 1.2/2007, 149-164; B. AJANA, *Biometric citizenship*, in *Citizenship Studies* 16.7/2012, 851-870.
[30] K. Hill, *La tua faccia ci appartiene*, Milano 2024.
[31] *Ivi*, 40.
[32] *Ivi*, 42.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Come vengono definiti i confini del corpo? La distinzione stessa non è più evidente, ma diventa sempre più ambigua: cosa riguarda il corpo e cosa è l'informazione sul corpo?

Alla luce di tali considerazioni, le tecnologie di *lie-detecting* possono essere interpretate come una macchina panottica che effettua esperimenti sul corpo umano, riflettendo la crescente enfasi posta non solo sulla politicizzazione ma anche sull'informatizzazione della vita, attraverso la delineazione di dispositivi di sicurezza che presuppongono l'incertezza delle minacce, impiegando metodi radicali che diventano pratica legale e governando il «radicalmente sconosciuto»[33]. Inoltre, attraverso tali processi, «il futuro guadagna ingiustamente il primato sia sul presente che sul passato»[34], innescando comportamenti che si basano su meri stimoli e risposte senza una riflessione autocosciente. Questa capacità foucaultiana di creare vita e lasciare morire è utile a coloro che detengono il potere e ha, di conseguenza, importanti implicazioni per la vita di gruppi emarginati, come migranti e rifugiati. Tale discorso anti-stranieri profila le minoranze come sgradite, consentendo alla sorveglianza di proteggere la maggioranza da «rischi oscuri e informi»[35]. La giustificazione delle strategie di sorveglianza in nome della sicurezza interna appare come una carta vincente discorsiva che prevale su tutte le altre affermazioni e disposta a sacrificare la vulnerabilità di gruppi discriminati. Avanza progressivamente l'orizzonte del realismo della sorveglianza, l'idea secondo cui nonostante si riconoscano e si temano gli errori di un sistema, che limita le libertà e invade i diritti, non sia possibile ormai immaginare una società senza un controllo onnipresente. In tal modo, il discorso di *iBorderCtrl* rivela una tendenza a ritrarre le forze di polizia digitali impegnate nella prevenzione del crimine e nell'immobilizzazione dei sospetti come qualcosa di neutrale, naturalmente buona, che sembra richiamare l'immagine foucaultiana di un potere sorvegliato «attraverso una figura gerarchica continua che assicura l'obbedienza, comandata da buoni ufficiali e uomini di sostanza[36]». Emerge, sullo sfondo, la cristallizzazione del biopotere, della capacità di far vivere e lasciar morire, dove la biopolitica trae la sua conoscenza dalle disabilità biologiche e il binomio potere e conoscenza condensa la sua azione di intervento.

Alla luce di tali considerazioni, le tecnologie di *lie-detecting* contribuiscono ad evidenziare due volti della vulnerabilità del migrante: la *precariousness* e la *precarity*[37]. La prima identifica quella vulnerabilità che ogni essere umano condivide in ragione della sua condizione corporea, della sua finitezza e limitazione, dell'esposizione al bisogno e alla sofferenza, mentre la seconda dipende dalle forme sociali, politiche, economiche e relazionali che qualificano le vite dei singoli soggetti. Essendo l'ontologia dell'umano costitutivamente relazionale e sociale e poiché l'essere non è mai definitivamente scindibile dall'altro[38], nonché dalle norme sociali o dalle strutture politiche e sociali e storicamente date, il volto della vulnerabilità ontologica del migrante (*precariousness)* si dispiega nelle forme della sua distribuzione differenziale, sociale, ed economica (*precarity*). È l'oppressione sistemica e non occasionale, perpetrata ai danni di un gruppo, quale quello legato allo status di migrante, che crea una forma

[33] R. WICHUM, *Security as Dispositif: Michel Foucault in the Field of Security*, in *Foucault Studies,* 15/2013, 164-171.
[34] *Ibidem*.
[35] Z. BAUMAN, D. LYON, *Liquid Surveillance. A Conversation*, Cambridge, 2013.
[36] M. FOUCALT, *Sorvegliare e punire*, Torino, 1976.
[37] F. MACIOCE, *La vulnerabilità di gruppo*, Torino, 2021.
[38] *Ivi*, 25.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

di identità di gruppo[39]; allo stesso modo, quando la vulnerabilità non è meramente individuale, ma dipende dal funzionamento di strutture di conoscenza e di potere, o da sistemi di distribuzione di risorse, si creano condizioni accresciute di debolezza ai danni di quelli che vengono percepiti come gruppi di individui, determinando, di conseguenza, una definizione conclusiva di "vulnerabilità di gruppo", senza che questo implichi l'assunzione di una prospettiva essenzialista. La dimensione di gruppo in questo senso non è ontologica, ma, almeno in una certa misura, è identitaria: sta nel trovarsi esposti, insieme ad altri, a medesime forme di oppressione in quanto si viene percepiti come parte di un gruppo[40]. L'oppressione non sta nella vittimizzazione diretta che si ha nel singolo caso, ma nella consapevolezza di tutti gli appartenenti al gruppo di essere esposti a questo rischio proprio in ragione di un'appartenenza identitaria che è collettiva. In un secondo senso, si può parlare di gruppo vulnerabile quando la vulnerabilità dipende da un analogo posizionamento di più individui all'interno di un determinato contesto, tale da condizionarne le possibilità d'azione, e tale, soprattutto, da influire sulla loro capacità di far fronte a rischi e incertezze e gestirne le conseguenze. Questo posizionamento, tuttavia, non ha alcun carattere identitario, cioè non può né essere rivendicato come tale dall'interno (in una sorte di *identity politics*), né può essere utilizzato dall'esterno in modo ascrittivo[41]. Le modalità assunte da tali dispositivi di *lie-detecting* contribuiscono a confinare i migranti all'interno del paradigma di una minoranza, non inferiore dal punto di vista numerico, ma costituito da un gruppo le cui possibilità di accesso al potere risultano limitate. L'identità di tale gruppo si riassume, pertanto, nella condizione di non *dominance*, di subalternità al potere, opaco e inspiegabile, della macchina della verità "intelligente", che costruisce il giudizio finale sulla base di dati estratti dal corpo: quale paura per quale volto?

4. I rischi delle banche dati nella dimensione di interoperabilità

Ad accentuare il profilo di debolezza e di vulnerabilità del migrante, all'interno della crescente e progressiva fluidità e disarticolazione del proprio corpo, quale nuovo strumento appetibile per le logiche di potere umano[42] e per le dinamiche al silicio, contribuisce la delicata e complessa gestione dell'utilizzo dei dati biometrici, associata alle ombre di una condivisione interattiva, che supera i margini nazionali e coinvolge una pluralità eterogenea di protagonisti.

I dati biometrici sono, infatti, considerati particolarmente sensibili in quanto consentono l'identificazione di un individuo attraverso la registrazione di caratteristiche personali immutabili[43]. La creazione di registri biometrici permanenti di rifugiati e migranti pone particolari preoccupazioni in materia di diritti umani. Nel caso dei rifugiati esiste il rischio, a causa delle insidie del *function creep*[44], che le loro informazioni possano essere condivise – intenzionalmente (ad esempio, come una forma di politica statale) o inavvertitamente (attraverso violazioni di dati/sistemi non sicuri) – con le autorità del Paese

[39] F. Macioce, *op.cit.*, 37.
[40] *Ibidem*
[41] L. Everuss, *op. cit.*, 35.
[42] N. Farahany, *Difendere il nostro cervello*, Milano, 2024.
[43] N. Farahany, *op. cit.,* 32.
[44] M. Tzanou, *The EU as an emerging 'Surveillance Society: The function creep case study and challenges to privacy and data protection*, in *ICL Journal*, 4.3/2010.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

da cui sono fuggiti, aumentando le possibilità di ulteriori abusi e persecuzioni. In particolare, l'utilizzo di sistemi centralizzati per l'archiviazione delle informazioni biometriche può facilitare la sorveglianza e l'uso improprio delle informazioni e rendere, di conseguenza, più dannose le violazioni dei dati[45]. Nel 2018 erano emerse notizie sulla condivisione da parte del governo del Bangladesh dei dati biometrici dei rifugiati Rohingya raccolti dall'UNHCR con il Myanmar, il Paese da cui erano fuggiti dal terrore e dalle violenze. Tali notizie erano state confermate da Human Rights Watch, che aveva accusato l'UNHCR di aver fornito informazioni personali dei rifugiati al governo del Bangladesh. I dati biometrici, inizialmente raccolti ai fini della registrazione e dell'accesso ai servizi, erano stati condivisi per il rimpatrio in assenza di un consenso libero e informato da parte dei rifugiati, ponendoli, di conseguenza, inesorabilmente a rischio. Un fattore abilitante di tali collegamenti è la crescita dell'interoperabilità che supporta la condivisione dei dati tra organizzazioni umanitarie, governi nazionali e agenzie di sicurezza, allo scopo di creare una solida ed interattiva rete transnazionale di polizia[46].

A tal proposito, è utile sottolineare che dal 2018 la polizia italiana utilizza un sistema di riconoscimento facciale chiamato S.A.R.I.[47] per identificare, durante le indagini, un soggetto ignoto confrontando la foto del volto con quelle collezionate nella banca dati AFIS. Il sistema è in grado di fornire un elenco di immagini ordinato secondo un grado di similarità, i cui risultati vengono, poi, analizzati dagli operatori specializzati della Polizia scientifica. Il sistema SARI presenta due diversi moduli: *SARI Enterprise* e *SARI Real-Time*. Il primo modulo permette di individuare l'immagine di un sospettato, acquisita, ad esempio, dalle videocamere a circuito chiuso, e confrontarla con riproduzioni di volti presenti nelle banche dati in possesso della polizia. Il *SARI Real-Time*, invece, è in grado di analizzare in tempo reale i volti dei soggetti ripresi dalle telecamere installate in un determinato luogo e di confrontarli con una watch-list la cui grandezza è dell'ordine delle decine di migliaia di soggetti. *SARI Enterprise* è già utilizzato durante le indagini mentre SARI Real-Time, ancora non attivo, è pensato a supporto di operazioni di controllo del territorio in occasione di eventi e/o manifestazioni, sfruttando la sua peculiare potenzialità di generare degli *alert* quando nel video appaiono individui presenti nella *watch-list*. Dal punto di vista legale, la Polizia ha ricevuto l'approvazione dal Garante privacy per l'utilizzo di SARI Enterprise nel luglio 2018, riconoscendo che il sistema automatizza semplicemente un'attività che le forze di polizia hanno sempre svolto manualmente, ossia la ricerca dei volti per anagrafica e dettagli in AFIS[48]. Per *SARI Real-*

[45] Cfr. M. LATONERO, *et al*., *Digital identity in the migration and refugee context*, in *Data & Society,* 4/2019; R. THOMAS, *Biometrics, international migrants and human rights*, in *Eur. J. Migration & L.,* 7/2005; C. COSTELLO, I. MANN, *Border justice: migration and accountability for human rights violations,* in *German Law Journal,* 21.3/2020.

[46] Sul punto si rinvia a H. ADEN, *Interoperability between EU policing and migration databases: Risks for privacy*, in *European Public Law,* 26.1/2020; E. BROUWER, *Large-scale databases and interoperability in migration and border policies: The non-discriminatory approach of data protection*, in *European Public Law,* 26.1/2020; N. VAVOULA, *Interoperability of EU information systems: The deathblow to the rights to privacy and personal data protection of third-country nationals*, in *European public law,* 26.1/2020.

[47] R. LOPEZ, *La rappresentazione facciale tramite software*, in A. SCALFATI (a cura di), *Le indagini atipiche*, Torino, 2019; G. MOBILIO, *Tecnologie di riconoscimento facciale. Rischi per i diritti fondamentali e sfide regolative*, Napoli, 2021; E. SACCHETTO, *Face to face: il complesso rapporto tra automated facial recognition technology e processo penale*, in *La legislazione penale web*, 2020, 1-14.

[48] Garante per la protezione dei dati personali, Parere sul sistema Sari Enterprise, 26.7.2018, n. 440, doc. web n. 9040256, in www.garanteprivacy.it.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Time, invece, il Garante Privacy si era espresso nell'aprile 2021 definendo il sistema di riconoscimento facciale, così come progettato, una possibile forma di sorveglianza e identificazione di massa che non poteva essere utilizzata dal Ministero dell'Interno perché non vi era ancora una base legale per il trattamento di dati biometrici da parte delle forze dell'ordine[49].

Da un'analisi dei comunicati stampa delle Questure italiane emerge che il sistema SARI Enterprise è ampiamente utilizzato nelle attività di polizia e, in alcuni casi, le persone identificate sono cittadini stranieri presenti in Italia: tra i soggetti coinvolti vi sono persone di etnia rom, persone di origine algerina, e persone nate in Italia da genitori stranieri. Non sempre però la polizia spiega il motivo per cui l'immagine è già presente nel database AFIS, in alcuni casi è indicato che i soggetti coinvolti sono stati fotosegnalati in precedenza per aver commesso altri reati. I soggetti presenti in AFIS rientrano però anche in altre categorie. Infatti, secondo quanto previsto dal Testo unico delle disposizioni concernenti la disciplina dell'immigrazione e norme sulla condizione dello straniero (D.Lgs 286/1998), chi richiede il permesso di soggiorno o chi ne domanda il rinnovo è sottoposto a fotosegnalamento. Comprendere la composizione del database AFIS utilizzato con il sistema di riconoscimento facciale SARI è fondamentale per capire quali rischi corrono le persone che vi sono incluse. Il database AFIS è, infatti, utilizzato durante le indagini per cercare l'identità di un sospetto tra volti già noti alle autorità per aver commesso dei reati, includere nello stesso database migranti e richiedenti asilo rischia, di conseguenza, di criminalizzare ulteriormente tali soggetti. Secondo una recente inchiesta[50], 8 persone su 10 presenti nel database AFIS sarebbero stranieri, ovvero circa 2 milioni di cittadini italiani e 7 milioni di persone con cittadinanza diversa da quella italiana. Al momento in Italia vi è, quindi, il rischio che un richiedente asilo possa essere fermato e interrogato dalla polizia solo perché l'algoritmo del sistema SARI ha indicato un match con la foto di un soggetto schedato con il quale condivide solamente il colore della pelle. Alla luce di tali considerazioni, la composizione del database, la mancanza di informazioni e analisi sull'accuratezza degli algoritmi utilizzati e l'assenza di risposte precise da parte delle forze dell'ordine sollevano necessarie preoccupazioni sui rischi che il sistema SARI potrebbe produrre quando utilizzato su migranti e persone straniere presenti in Italia, soprattutto in luce delle novità introdotte dall'entrata in vigore dell'AI Act. Infatti, la recente legislazione sulla regolamentazione dell'intelligenza artificiale ha vietato la categorizzazione biometrica con lo scopo specifico di dedurre i dati sensibili e l'identificazione biometrica da remoto negli spazi pubblici, con alcune eccezioni riservate alle forze dell'ordine relative alla ricerca di criminali e vittime e alle minacce terroristiche[51], che potrebbero determinare, alla luce delle considerazioni sopra esposte, il riesame delle potenzialità di impiego del sistema *SARI in Real Time*. Nel contesto migratorio invece, pur affermando che i sistemi basati sull'IA non devono in alcun modo infrangere il principio di non-respingimento, l'AI Act si è limitato a inserire nella lista ad alto rischio[52] (senza proibirli) i poligrafi, la profilazione del rischio individuale, i sistemi per esaminare le richieste d'asilo e quelli per rilevare, riconoscere o identificare le

[49] Garante per la protezione dei dati personali, Parere sul sistema Sari Real Time, 25.3.2021, n. 127, doc. web n. 9575877, in www.garanteprivacy.it.

[50] Disponibile in www.asgi.it (ultima consultazione 03/07/24).

[51] Cfr. art. 5 dell'AI Act.

[52] Cfr. art. 6 dell'AI Act.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

persone ai confini[53]. Tuttavia, i legislatori dell'UE si sono rifiutati, almeno per il momento, di vietare sistemi dannosi come i sistemi di valutazione del rischio discriminatorio nella migrazione e l'analisi predittiva se utilizzata per facilitare i respingimenti. Inoltre, il divieto di riconoscimento delle emozioni non si applica al contesto migratorio, escludendo, pertanto, i casi documentati di macchine della verità IA alle frontiere. L'elenco dei sistemi ad alto rischio non tiene conto dei numerosi sistemi di IA utilizzati nel contesto della migrazione e che, di conseguenza, non risultano soggetti agli obblighi del presente regolamento. L'elenco esclude modelli pericolosi come i sistemi di identificazione biometrica, gli scanner di impronte digitali o gli strumenti di previsione utilizzati per prevedere, bloccare e limitare la migrazione. Sullo stesso fronte di lacune e incertezze, l'IA utilizzata in ausilio di banche dati su larga scala dell'UE in materia di migrazione, come Eurodac, il Sistema d'informazione Schengen e ETIAS, non dovrà essere conforme al regolamento fino al 2030, aprendo pericolosamente spazi di incertezze. La legge sull'IA non ha, poi, affrontato il modo in cui i sistemi di IA sviluppati da aziende con sede nell'UE possano avere un impatto sulle persone al di fuori dell'UE, nonostante le prove esistenti di violazioni dei diritti umani facilitate dalle tecnologie di sorveglianza sviluppate nell'UE e impiegate in Paesi terzi. Pertanto, allo stato attuale non sarà proibito esportare un sistema vietato in Europa al di fuori dei confini europei.

L'aspetto forse più dannoso della legge europea sull'IA è la creazione di un quadro giuridico parallelo quando l'IA viene impiegata dalle autorità di polizia, di immigrazione e di sicurezza nazionale. Grazie alle pressioni esercitate dagli Stati membri, dalle forze dell'ordine e dalle lobby dell'industria della sicurezza, queste autorità sono esplicitamente esentate dalle norme e dalle salvaguardie più importanti della legge sull'IA.

5. Conclusioni

Il progresso tecnologico ha contribuito ad accentuare i mille volti della vulnerabilità del migrante, rendendolo preda prelibata e cavia inconsapevole delle nuove sperimentazioni.

Gli Stati sono in grado di giustificare i crescenti esperimenti tecnologici in materia di migrazione perché i migranti sono stati storicamente considerati come una popolazione da gestire e quantificare. La stessa retorica della gestione della migrazione implica che i rifugiati e i migranti debbano essere sorvegliati e controllati, in quanto considerati una minaccia alla sovranità nazionale, soprattutto in tempi in cui sempre più gli Stati si rivolgono verso l'interno e reificano il loro potere. Il concetto di "esclusione inclusiva" di Agamben[54], in cui lo Stato è in grado di dividere e separare le popolazioni sulla base della figura del fuorilegge, che si trova al di fuori dei confini della vita sociale e politica dello Stato, rafforza ulteriormente il modo in cui immaginiamo che la differenziazione dei diritti sia naturale quando viene utilizzata per giustificare interventi e sperimentazioni ai margini per il cosiddetto bene comune. Il

[53] D. Ozkul, *Automating Immigration and Asylum: The Uses of New Technologies in Migration and Asylum Governance in Europe*. Oxford: Refugee Studies Centre, University of Oxford, 2023; J. Laux, S. Wachter, B. Mittelstadt, *Trustworthy artificial intelligence and the European Union AI Act: On the conflation of trustworthiness and acceptability of risk*, in *Regulation & Governance,* 18.1/2024; A. Mantelero, *The Fundamental Rights Impact Assessment (FRIA) in the AI Act: Roots, legal obligations and key elements for a model template*, in *Computer Law & Security Review* 54, 2024.

[54] G. Agamben, *State of Exception*, University of Chicago Press, Chicago 2005.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

potere ultimo dello Stato di decidere chi può entrare e a quali condizioni è rafforzato dalla continua convinzione dell'imparzialità tecnologica, all'interno di una tensione intrinseca tra la prerogativa rivendicata dagli Stati nazionali sulla sovranità e la natura malleabile della tecnologia. Nella sua fluidità, la tecnologia è intrinsecamente contraria ai confini e, per estensione, alla sovranità, riverberando molto spesso i suoi effetti sulla definizione stessa di umanità nell'era digitale[55]. La distribuzione ineguale dei benefici che derivano dallo sviluppo tecnologico contribuisce a creare monopoli della conoscenza e a consolidare il potere e l'autorità conferiti allo Stato sovrano. Questi monopoli possono esistere perché non sussiste un regime normativo globale unificato che disciplini l'uso delle nuove tecnologie, creando laboratori per esperimenti ad alto rischio con un profondo impatto sulla vita delle persone più vulnerabili. Alla luce di tali considerazioni, le tecnologie di *lie-detecting* traducono l'immagine di un costrutto sociale, uno specchio in grado di riprodurre i problemi e i pregiudizi che sono già insiti nella società e che compromettono la pratica stessa della democrazia. L'interrogatorio biometrico prodotto dagli sviluppi dell'AI contribuisce a delineare un nuovo volto, una nuova figura, un corpo di cristallo del migrante, rotto in mille pezzi dal giudizio pronunciato da un avatar che si erge al ruolo di giudice naturalmente artificiale e che tenta di estrarre e tradurre i singoli cristalli in verdetti dell'intenzione, contribuendo ad aggiungere nuovi tasselli al mosaico della vulnerabilità: è la nuova frontiera per la nuova paura dell'artificiale? Un terrore al silicio?

[55] E. Zureik, K. Hindle, *Governance, Security and Technology: The Case of Biometrics*, in *Studies in Political Economy*, 2004.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Intelligenza artificiale e ingiustizia socio-linguistica: è necessaria una riflessione interdisciplinare

Francesca Morganti, Beatrice Zuaro*

ARTIFICIAL INTELLIGENCE AND SOCIO-LINGUISTIC INJUSTICE: THE NEED FOR AN INTERDISCIPLINARY REFLECTION

ABSTRACT: Artificial intelligence (AI) systems can – as stressed in a recent Council of Europe document – help safeguard regional and minority languages. This paper reflects on the potential of AI to address situations of linguistic marginalization and vulnerability, identifying the interdisciplinary collaboration as a necessary condition to push the discussion beyond the, certainly relevant, issues of translation and "access", and towards the bi-directional relationship between AI and more complex and severe forms of epistemic injustice.

KEYWORDS: Artificial intelligence; large language models; minority languages; linguistic injustice; interdisciplinarity.

ABSTRACT: I sistemi di intelligenza artificiale, come evidenziato in un recente documento del Consiglio d'Europa, possono aiutare – in un'ottica di protezione delle minoranze linguistiche e di eguaglianza sostanziale – a salvaguardare le lingue regionali o minoritarie. Nel contributo si tenta, appunto, di riflettere sul potenziale dei suddetti sistemi nel contrasto a situazioni di marginalità (o vulnerabilità) linguistica, ma, soprattutto, si individua l'approccio interdisciplinare come condizione necessaria per spostare i termini del discorso – oltre le questioni di "accesso" e traduzione, pure centrali – sulla relazione bi-direzionale che intercorre tra intelligenza artificiale e forme, più complesse e più gravi, di ingiustizia epistemica.

PAROLE CHIAVE: Intelligenza artificiale; modelli linguistici di grandi dimensioni; minoranze linguistiche; ingiustizia linguistica; interdisciplinarità.

* *Francesca Morganti: assegnista di ricerca in Diritto costituzionale e pubblico, Università degli Studi di Roma "Tor Vergata". Mail: francesca.morganti@outlook.com. Beatrice Zuaro: ricercatrice associata post-dottorale, Centre for Internationalisation and Parallel Language Use (CIP) dell'Università di Copenaghen e ricercatrice associata onoraria presso la Open University di Milton Keynes. Mail: bzu@hum.ku.dk.* *Il presente scritto è frutto del lavoro congiunto delle Autrici; tuttavia, i paragrafi 1 e 2 sono da attribuire a Francesca Morganti; il paragrafo 3, nelle sue articolazioni, a Beatrice Zuaro; il paragrafo 4 a entrambe. Contributo sottoposto a doppio referaggio anonimo.*

Downloaded from www.biodiritto.org.
ISSN 2284-4503

modelli linguistici di grandi dimensioni: potenzialità di utilizzo… – 3.1. … e criticità nell'applicazione. – 4. Intelligenza artificiale, minoranze (vecchie e nuove) e ingiustizia socio-linguistica: considerazioni conclusive.

1. Considerazioni sulla nozione costituzionale di "minoranza linguistica" (e sulla sua evoluzione)

E' stato sostenuto, in passato – e già in sede di Costituente – che una norma costituzionale posta a tutela delle minoranze linguistiche non fosse realmente necessaria: sembrava potesse bastare l'art. 3, co. 1, Cost., il quale dispone, tra l'altro, che tutti i cittadini «sono eguali davanti alla legge, senza distinzione (…) di lingua». Lo stesso Meuccio Ruini, Presidente della Commissione per la Costituzione, aveva fatto notare, nel corso del dibattito in Assemblea sull'emendamento proposto dall'on. Codignola[1], che «vi è già nell'articolo 2 delle dichiarazioni generali della Costituzione [*i.e.*, l'attuale art. 3 Cost.], il principio di eguaglianza di tutti i cittadini, indipendentemente dalla razza e dalla lingua. Altre garanzie in questo senso di una perfetta parità fra gli italiani vi sono in tutta la Costituzione. Una speciale disposizione per le minoranze etnico-linguistiche – *né ben si comprende il concetto di minoranza* – non sembra indispensabile, potendo rientrare nel concetto generale»[2]. A queste preoccupazioni circa la "ridondanza" di una norma specifica sulle minoranze linguistiche – passate in secondo piano nel dibattito costituente – la più autorevole dottrina ha sempre risposto, tra l'altro, che la proclamazione di cui all'art. 3, co. 1, Cost. «vale a stabilire il divieto di discriminazioni, ma non necessariamente anche quella tutela "positiva" cui le minoranze linguistiche, in quanto minoranze "volontarie", aspirano»[3].

[1] Nel testo approvato dalla Commissione per la Costituzione, infatti, non era presente alcuna disposizione sulle minoranze linguistiche; quest'ultima fu proposta in seguito, come emendamento, dall'on. Tristano CODIGNOLA, nella seguente forma: «La Repubblica garantisce il pieno e libero sviluppo, nell'ambito della Costituzione, delle minoranze etniche e linguistiche esistenti sul territorio dello Stato. / Gli enti autonomi regionali non possono, sotto nessuna forma, limitare o modificare i diritti fondamentali del cittadino sanciti dalla presente Costituzione, né emanare norme con essa in contrasto». La discussione a riguardo si svolse interamente nelle sedute del 1° e del 22 luglio 1947 dell'Assemblea.

[2] Così l'on. Meuccio RUINI, già Presidente della Commissione per la Costituzione, nel corso della seduta pomeridiana dell'Assemblea Costituente del 1° luglio 1947.

[3] Così A. PIZZORUSSO, *Art. 6*, in C. MORTATI *et al.*, *Principi fondamentali*, parte di G. BRANCA (a cura di), *Commentario della Costituzione*, Bologna-Roma, 1975, 304, dove con "minoranze volontarie" si intendevano «quei gruppi sociali i quali si trovano in contrasto con la maggioranza perché questa tende ad impedire loro di mettere in valore le caratteristiche che li differenziano dalla maggioranza stessa» (ivi, 304, nt. 7). Lo stesso A., ad ogni modo, colloca idealmente il principio di tutela delle minoranze, affidato all'art. 6 Cost., nel solco de «gli altri basilari principî espressi dagli art. 2 e 3», ovvero quello pluralistico e, appunto, quello di eguaglianza (ivi, 306 ss.); una sorta di ruolo di "raccordo" è svolto, anche in questo caso, dall'eguaglianza sostanziale: come evidenziato in C. cost, sent. 22 gennaio 1996, n. 15, il principio di tutela delle minoranze linguistiche «si situa al punto di incontro con altri principi, talora definiti "supremi", che qualificano indefettibilmente e necessariamente l'ordinamento vigente (sentenze nn. 62 del 1992, 768 del 1988, 289 del 1987 e 312 del 1983): il principio pluralistico riconosciuto dall'art. 2 – essendo la lingua un elemento di identità individuale e collettiva di importanza basilare – e il principio di eguaglianza riconosciuto dall'art. 3 della Costituzione, il quale, nel primo comma, stabilisce la pari dignità sociale e l'eguaglianza di fronte alla legge di tutti i cittadini, senza distinzione di lingua e, nel secondo comma, prescrive l'adozione di norme *che valgano anche positivamente* per rimuovere le situazioni di fatto da cui possano derivare conseguenze discriminatorie» (pt. 2 del *Considerato in diritto*, corsivo aggiunto); il passaggio appena citato è ripreso per intero, tra l'altro, nella fondamentale sent. 18 maggio 2009, n. 159 (a commento della

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Dalle parole di Ruini emergeva anche, tuttavia, come già all'epoca vi fossero dei dubbi sull'esatta ampiezza della nozione costituzionale di minoranza[4] – «*né ben si comprende il concetto di minoranza*» – e come si tendesse ad associarla – donde il richiamo al principio di eguaglianza formale, riferito ai soli cittadini, e alle plurime garanzie costituzionali di una «perfetta parità *tra gli italiani*» – alle minoranze linguistiche "storiche" o "autoctone": quella tedesca e ladina dell'Alto Adige/*Südtirol*, quella francese della Valle d'Aosta, quella slovena delle Province di Trieste e Gorizia[5]. Si pensava, in altri termini, soprattutto a gruppi minoritari sul piano linguistico, ma composti di cittadini italiani (che questi ultimi si considerassero o meno «parte di una comunità nazionale diversa da quella espressa dalla maggioranza dei loro concittadini»[6]).

Per lungo tempo, la tutela (costituzionalmente imposta e orientata) delle minoranze linguistiche, riferita a specifiche formazioni e ordinata in senso territoriale[7], si è concentrata – anche per ragioni storico-pratiche e connesse a vincoli internazionali[8] – quasi esclusivamente sulle suddette realtà; più o meno consapevolmente, il legislatore italiano si era idealmente collocato nel solco di una distinzione operata già in seno alla Commissione Forti[9]: quella fra "isole linguistiche", «disseminate tra la popolazione di lingua italiana e ambientate oramai da molte generazioni, tanto che solo la lingua parlata tradizionale e d'origine, che hanno mantenuto viva tra loro senza ostacoli, né rivendicazioni, né inconvenienti, le differenzia dalla circostante popolazione»[10], e minoranze etnico-linguistiche vere e proprie,

quale v., tra i molti, almeno R. TONIATTI, *Pluralismo sostenibile e interesse nazionale all'identità linguistica posti a fondamento di "un nuovo modello di riparto delle competenze" legislative fra Stato e Regioni*, in *Le Regioni*, 5, 2009, 1121 ss.), relativa (anche) alla portata di parametro "interposto" da riconoscersi alla l. n. 482 del 1999 (su questo specifico aspetto cfr., *ex multis*, V. PIERGIGLI, *La tutela delle minoranze linguistiche storiche nell'ordinamento italiano tra principi consolidati e nuove (restrittive) tendenze della giurisprudenza costituzionale*, 2010, disponibile su: www.associazionedeicostituzionalisti.it/old_sites/sito_AIC_2003-2010/dottrina/libertadiritti/Piergigli.pdf [ultima consultazione 07/07/2024], spec. 7).

[4] Riflette sulla «fluidità» e «adattabilità» fisiologiche del concetto di "minoranza", *ex multis*, V. PIERGIGLI, *Rileggendo l'opera di Alessandro Pizzorusso sulle minoranze linguistiche: le "nuove minoranze" tra identità e integrazione*, in *Nomos*, 1, 2019, 3.

[5] Non è un caso, tra l'altro, che l'on. CODIGNOLA, primo propositore di una norma costituzionale sulle minoranze linguistiche (v. *supra*, nt. 1), originariamente intendesse quest'ultima come alternativa alla previsione di Regioni a statuto speciale: «[r]itenevo e ritengo tuttora che il sistema di adottare degli statuti speciali per alcune Regioni italiane sia un sistema sotto molti aspetti criticabile e discutibile. […] Io quindi proponevo che lasciando immutata la situazione esistente, la Costituzione si limitasse ad una affermazione di garanzia delle minoranze etniche e linguistiche, minoranze quasi esclusivamente di confine, residenti cioè su territori mistilingue, sia italo-francesi, sia italo-slavi, sia italo-austriaci» (così nella seduta pomeridiana dell'Assemblea Costituente del 1° luglio 1947).

[6] A. PIZZORUSSO, *Minoranze e maggioranze*, Torino, 1993, 62.

[7] Cfr., *ex multis*, A. PIZZORUSSO, *Art. 6*, cit., *passim*, spec. 311.

[8] Ivi, 299 ss.

[9] Il riferimento è, ovviamente, alla Commissione per studi attinenti alla riorganizzazione dello Stato, presieduta da Ugo Forti (docente di Diritto amministrativo presso l'Università di Napoli), istituita nel novembre 1945 presso il Ministero per la Costituente al fine di «predisporre gli elementi per lo studio della nuova Costituzione» (cfr. artt. 2 e 5, d. lt. 31 luglio 1945, n. 435).

[10] Cfr. Commissione per studi attinenti alla riorganizzazione dello Stato, *Relazione all'Assemblea Costituente*, vol. I, *Problemi costituzionali, organizzazione dello Stato*, Roma, 1946, 179; con "isole linguistiche" si intendevano, in particolare, quelle «albanesi, catalane e greche dell'Italia meridionale e insulare».

Downloaded from www.biodiritto.org.
ISSN 2284-4503

alle quali, lungi dall'essere percepite come «mero fatto folcloristico», era ed è attribuito «un preciso rilievo sul piano giuridico e politico»[11].

La stessa l. n. 482 del 1999[12], prima legge generale di attuazione dell'art. 6 Cost., non faceva che rimarcare e confermare lo storico «*favor* verso le minoranze linguistiche *riconosciute*», delle quali pure ampliava il novero[13], e il corrispondente «atteggiamento agnostico [del medesimo legislatore], quando addirittura non palesemente ostile, verso le restanti comunità di lingua e cultura minoritaria»[14]. Come evidenziato già al tempo della sua approvazione, la legge in esame – nel cui titolo sono menzionate, non a caso, le «minoranze linguistiche *storiche*»[15] – evitava (deliberatamente?) di affrontare il problema delle "nuove minoranze", «costituite da lavoratori immigrati e rifugiati in primo luogo»[16].

2. Minoranze linguistiche storiche e "nuove minoranze": quali sfide per il diritto?

Occorre chiedersi, in via preliminare, se la scelta di tener fuori le "nuove minoranze" fosse in qualche modo necessitata, e se nella nozione costituzionale di "minoranza" sia ricompreso, come è stato sostenuto, un implicito riferimento allo *status* di cittadini dei componenti il gruppo minoritario. Potrebbe sembrare di sì, almeno guardando alle definizioni originariamente proposte dalla più autorevole dottrina: «per minoranza in senso giuridico», scriveva, *e.g.*, Pizzorusso, «si intende una frazione del popolo la quale costituisce un gruppo sociale, posto in condizioni di inferiorità nell'ambito della comunità statale, i cui membri, legati allo stato dal rapporto di cittadinanza (o eccezionalmente da quello di sudditanza, di stabile residenza, *etc.*), ricevono dall'ordinamento giuridico di esso un trattamento particolare diretto ad eliminare la situazione minoritaria ovvero ad istituzionalizzarla e disciplinarla nell'ambito dello stato stesso»[17]. Tre parrebbero essere, secondo questa ricostruzione, gli elementi distintivi di una minoranza in senso giuridico: *(i)* la condizione di «inferiorità» nella quale versano i suoi membri, senz'altro collocati in una situazione "di fatto" da cui possono derivare «conseguenze discriminatorie» (C. cost., sentt. nn. 15/1996 e 159/2009[18]); *(ii)* il «trattamento particolare» di cui dovranno essere

[11] Come nota V. PIERGIGLI, *La tutela delle minoranze linguistiche storiche nell'ordinamento italiano*, cit., 1.

[12] L. 15 dicembre 1999, n. 482 («Norme in materia di tutela delle minoranze linguistiche storiche»), a commento della quale v. almeno, *ex multis*, S. BARTOLE, *Le norme per la tutela delle minoranze linguistiche storiche*, in *Le Regioni*, 6, 1999, 1063 ss.; V. PIERGIGLI, *La legge 15 dicembre 1999, n. 482: un traguardo per le minoranze linguistiche (finora) debolmente protette*, in *Quaderni costituzionali*, 1, 2000, 126 ss.; E. PALICI DI SUNI PRAT, *La legge italiana sulla tutela delle minoranze linguistiche storiche nel quadro europeo*, in *Diritto pubblico comparato ed europeo*, 1, 2000, 101 ss.; E. MALFATTI, *La legge di tutela delle minoranze linguistiche: le prospettive ed i problemi ancora aperti*, in *Rivista di diritto costituzionale*, 2001, 109 ss.

[13] Cfr., *e.g.*, l'art. 2 della legge in esame, che affida alla Repubblica la tutela della lingua e della cultura «delle popolazioni albanesi, catalane, germaniche, greche, slovene e croate e di quelle parlanti il francese, il franco-provenzale, il friulano, il ladino, l'occitano e il sardo».

[14] Come osserva ancora V. PIERGIGLI, *La tutela delle minoranze linguistiche storiche nell'ordinamento italiano*, cit., 1. La stessa Corte costituzionale legava, già nella sent. 20 gennaio 1982, n. 28, la «operatività normativa» dell'art. 6 Cost. al carattere "riconosciuto" – legislativamente, s'intende – della minoranza in questione (pt. 2 del *Considerato in diritto*); nello stesso senso v. anche, sempre con riferimento alla minoranza slovena insediata nel Friuli-Venezia Giulia, C. cost., sent. 5 febbraio 1992, n. 62, *passim*.

[15] Corsivo aggiunto.

[16] S. BARTOLE, *op. cit.*, 1065.

[17] A. PIZZORUSSO, *Le minoranze nel diritto pubblico interno*, Milano, 1967, 193.

[18] Sulle quali più diffusamente *supra*, in nota 3.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

destinatari *al fine di eliminare o istituzionalizzare quella posizione "vulnerabile"*; *(iii)* il rapporto di cittadinanza che pure lega i componenti del gruppo minoritario allo Stato, solo «eccezionalmente» rimpiazzabile da vincoli di "sudditanza" o di stabile residenza. Di lì a qualche anno, Capotorti, in qualità di *special rapporteur* della *Sub-Commission on Prevention of Discrimination and Protection of Minorities* presso le Nazioni Unite, avrebbe qualificato la minoranza, all'interno di uno studio dedicato, come gruppo «*numerically inferior to the rest of the population of a State, in a non-dominant position, whose members* – being nationals of the State – *possess ethnic, religious or linguistic characteristics differing from those of the rest of the population and show, if only implicitly, a sense of solidarity, directed towards preserving their culture, traditions, religion or language*»[19]; tornavano, seppure in veste parzialmente diversa, gli elementi di cui sopra: posizione "non-dominante", solidarietà ed esigenze di tutela, cittadinanza.

Lo stesso Pizzorusso, tuttavia, come è stato sottolineato[20], ha parzialmente aggiornato, nel tempo, le proprie posizioni[21], e ha rilevato come «problemi analoghi a quelli propri delle minoranze possano porsi nei confronti di popolazioni immigrate, cui il diritto di cittadinanza – almeno entro certi limiti – può essere legittimamente negato, ma alle quali non può però essere comunque negato l'esercizio dei diritti fondamentali di libertà»; proprio il «diniego della cittadinanza» agli «appartenenti a un gruppo sociale *sociologicamente configurabile come una potenziale minoranza*», aggiungeva l'A., può rappresentare, talvolta, «uno strumento di discriminazione nei loro confronti»[22]. La marginalità linguistica, dunque – e le corrispondenti esigenze di tutela – sono spesso aggravate e acuite dalla non-cittadinanza, rendendo necessario, tanto più nell'attuale contesto socio-economico-politico, che lo Stato non ignori e, anzi, si faccia particolare carico delle condizioni di vulnerabilità multipla o multifattoriale, le quali espongono la persona a forme *intersezionali* di discriminazione (potenziale)[23].

---

[19] F. CAPOTORTI, *Study on the Rights of Persons Belonging to Ethnic, Religious and Linguistic Minorities*, New York, 1979, § 568, enfasi aggiunta.

[20] In particolare, da V. PIERGIGLI, *Rileggendo l'opera di Alessandro Pizzorusso sulle minoranze linguistiche*, cit., *passim*, spec. 4 ss.

[21] Propendono per una interpretazione "estensiva" dell'art. 6 Cost., tra gli altri, G. DE VERGOTTINI, *Verso una nuova definizione del concetto di minoranza*, in *Regione e governo locale*, 1-2, 1995, 9 ss., *passim*; M. COSULICH, *Lingue straniere e lingue minoritarie nell'ordinamento repubblicano*, in *Quaderni regionali*, 2, 2012, 133 ss., il quale, in parte rifacendosi a M. UDINA, *Sull'attuazione dell'art. 6 della Costituzione per la tutela delle minoranze linguistiche*, in *Giurisprudenza costituzionale*, 1974, 3599 ss., parla di «carattere aperto» dell'art. 6 Cost., «suscettibile di essere applicato a un novero di minoranze linguistiche ben più ampio di quello cui pensava il costituente» (143-4); C. GALBERSANINI, *La tutela delle nuove minoranze linguistiche: un'interpretazione evolutiva dell'art. 6 Cost.?*, in *Rivista AIC*, 3, 2014, spec. 7 ss.

[22] Così A. PIZZORUSSO, *Minoranze e maggioranze*, cit., 62, corsivo aggiunto; secondo G. DE VERGOTTINI, *op. cit.*, 12, nt. 5, la «carenza di tutela» nei confronti nelle nuove minoranze, e dunque delle comunità di stranieri immigrati, sarebbe spiegabile, oltre che con «la non insistenza su una data porzione di territorio» (v. *infra*, nt. 24), proprio con «la mancanza del requisito della cittadinanza italiana».

[23] La discriminazione *intersezionale*, com'è noto, non è il risultato della mera sommatoria di plurimi fattori di discriminazione compresenti (nei casi di specie, oltre alla lingua: "razza", origine etnica, condizioni sociali, eventualmente genere): quando più tratti identitari marginalizzanti si combinano, infatti, la loro incidenza tenderà a essere superiore alla somma delle discriminazioni e dello stigma risultanti da ciascun fattore considerato singolarmente; nelle parole di Kimberlé CRENSHAW, fondatrice ideale dell'intersezionalismo, che pensava soprattutto all'intrecciarsi del sesso e della "razza": «*the intersectional experience is greater than the sum of racism and sexism*» (K. CRENSHAW, *Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of*

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Va riconosciuto, nondimeno, come il carattere "diffuso" delle nuove minoranze[24] – la circostanza, in altri termini, che, a differenza delle minoranze storiche, esse presentino notevoli elementi di eterogeneità e non siano concentrate in specifiche aree – possa rendere particolarmente difficoltosi (e dispendiosi) interventi puntuali di tutela: è più agevole, a mero titolo di esempio, garantire che «l'insegnamento nelle scuole materne, elementari e secondarie [sia] impartito nella lingua materna italiana o tedesca degli alunni»[25] in un territorio circoscritto e caratterizzato da insediamenti "antichi" (l'Alto Adige/*Südtirol*, nel caso di specie), che prevedere misure assimilabili, o anche assai più contenute, per una molteplicità di lingue e sull'intero territorio nazionale; lo stesso può dirsi, *mutatis mutandis*, per la redazione di documenti ufficiali, per le garanzie "linguistiche" all'interno del processo *etc.*

Può essere utile, forse, estendere alle nuove minoranze considerazioni svolte – con riferimento, in quel caso, alle cc.dd. lingue non territoriali – nella Relazione esplicativa[26] allegata alla Carta europea delle lingue regionali o minoritarie, adottata oltre trent'anni fa nell'ambito del Consiglio d'Europa[27]. Con lingue «non territoriali» (o "sprovviste di territorio") si intendono, in base all'art. 1, lett. *c*, della Carta, «le lingue usate da alcuni cittadini dello Stato che differiscono dalla(e) lingua(e) usata(e) dal resto della popolazione di detto Stato ma che (…) non posso essere ricollegate a un'area geografica particolare di quest'ultimo»; come si evidenzia nella Relazione esplicativa, al § 78, se è vero che «alcune delle disposizioni contenute nella Carta possono essere applicate senza difficoltà anche alle lingue non-territoriali» (si pensi, *e.g.*, al «riconoscimento di quei linguaggi» o alle «misure volte a sviluppare uno spirito di rispetto, comprensione e tolleranza nei loro riguardi»), non sarà invece possibile – o sarà, comunque, assai complesso – estendere loro le disposizioni concernenti, tra l'altro, l'utilizzo delle lingue minoritarie nella vita pubblica (insegnamento, giustizia, servizi pubblici). Questo non vuol dire, si legge comunque al § 76, che la Carta, per quanto «riguardi principalmente le lingue storicamente identificate con una particolare area geografica dello Stato», intenda «ignorare le lingue tradizionalmente parlate al suo interno ma prive di una base territoriale precisa»: vi sarà bisogno, tuttavia, per tutelarle, di «*certain adjustments*» alle misure di sostegno tradizionali (§ 77).

Lo stesso può dirsi, con tutti i *caveat* del caso, relativamente al patrimonio linguistico e culturale delle nuove minoranze: le difficoltà pratiche – non ultima, la non-territorialità – e gli ostacoli *lato sensu*

Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics, in *University of Chicago Legal Forum*, 1, 1989, 140). Sulla nozione di intersezionalità v., nella dottrina italiana, almeno B.G. Bello, *Diritto e genere visti dal margine: spunti per un dibattito sull'approccio intersezionale al diritto antidiscriminatorio in Italia*, in *Diritto e questioni pubbliche*, 2, 2015, 141 ss.; Ead., *Intersezionalità. Teorie e pratiche tra diritto e società*, Milano, 2020; nonché, per un'intervista della stessa B.G. Bello e di L. Mancini a Kimberlé Crenshaw, *Talking about Intersectionality. Interview with Kimberlé Chrenshaw*, in *Sociologia del diritto*, 2, 2016, 11 ss.

[24] Nelle parole di G. de Vergottini, *op. cit.*, 11, «la minoranza formata da stranieri immigrati si configura in modo sensibilmente diverso rispetto alle abituali minoranze autoctone», in particolare perché «le comunità di immigrati non hanno un legame diretto con un preciso territorio regionale di insediamento; le stesse comunità non hanno un legame stabile e continuativo col territorio nazionale essendo (…) caratterizzate da sensibili avvicendamenti e non essendovi certezza sulla continuità futura della loro presenza».

[25] Così l'art. 19 dello Statuto speciale per il Trentino-Alto Adige/*Sonderstatut für Trentino-Südtirol*.

[26] Reperibile su: www.coe.int/it/web/european-charter-regional-or-minority-languages/testo-della-carta (ultima consultazione 07/07/2024). Il testo originale è in inglese, le porzioni citate nel testo sono tradotte in italiano dall'A.

[27] Trattato internazionale concluso a Strasburgo il 5 novembre 1992 nell'ambito del Consiglio d'Europa, sottoscritto dall'Italia nel 2000 ma mai ratificato.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

ideologici alla loro promozione e tutela richiedono, oltre che una scelta politica "a monte", l'adozione, "a valle", di strumenti altrettanto *nuovi*. Tra questi – come è stato suggerito, non a caso, in un recente rapporto sull'implementazione della Carta europea delle lingue regionali o minoritarie redatto dalla linguista Miriam Gerken per conto del Consiglio d'Europa[28] – non può non figurare l'intelligenza artificiale (IA), e in particolare quella sua branca che ha a che vedere con l'elaborazione del linguaggio naturale (NLP, da *natural language processing*); rimane inteso, come è usuale quando si fa ricorso alla tecnologia nell'erogazione di servizi o per agevolare l'accesso a questi ultimi, che i vantaggi pratici rischiano, se non si orientano opportunamente gli interventi e non si operano eventuali, necessarie correzioni, di essere "controbilanciati" da costi notevoli in termini di giustizia sociale (e di diritti).

3. Intelligenza artificiale e modelli linguistici di grandi dimensioni: potenzialità di utilizzo…

La Come evidenziato nel sopracitato rapporto di Gerken[29], «lo scopo dell'NLP è lo sviluppo di programmi in grado di leggere, processare, analizzare e infine comprendere i linguaggi naturali in tutta la loro complessità». Un obiettivo così ambizioso ha comportato, negli ultimi anni, un significativo investimento di risorse materiali e intellettuali, tra cui, notoriamente, ingenti quantità di dati rappresentativi delle modalità umane di comunicazione linguistica. Sebbene i sistemi di IA generativa (*generative artificial intelligence*), infatti, siano in grado di produrre contenuti originali, questi ricalcano informazioni esistenti e di matrice umana.

I programmi attualmente più efficaci nel comunicare tramite linguaggio umano sono noti come modelli linguistici di grandi dimensioni (*large language models*, o LLM), e funzionano convertendo un *input* testuale[30] in una combinazione numerica che viene processata da una rete neurale (*neural network*) e successivamente riconvertita in testo *output*. I milioni di numeri e parametri che, interconnessi, costituiscono la rete neurale di un LLM non sono generati tramite programmazione manuale, ma mediante un addestramento (*training*) che, entro certi termini, ricalca le modalità di apprendimento di un infante: al modello vengono fornite grandi quantità di informazioni, generalmente recuperate dalla rete, in modo che possa allenarsi a riconoscere quali parole fanno seguito, di solito, a un determinato *input*. In fase di addestramento, il modello conduce continui tentativi di produrre stringhe testuali sensate e pertinenti, migliorando gradualmente il suo livello di accuratezza (in un processo noto come *back-propagation*); il sistema è ulteriormente addestrato tramite *feedback* umano (*reinforcement learning with human feedback*) volto a valutarne la *performance*[31]. Ultimato l'addestramento, il modello viene

[28] M. Gerken (per conto del Consiglio d'Europa - Segretariato della Carta europea delle lingue regionali o minoritarie), *Facilitating the implementation of the European Charter for Regional or Minority Languages through artificial intelligence*, 2022, disponibile su: edoc.coe.int/en/minority-languages/11416-facilitating-the-implementation-of-the-european-charter-for-regional-or-minority-languages-through-artificial-intelligence.html (ultima consultazione 07/07/2024).

[29] M. Gerken, *op. cit.*, 4.

[30] Si noti, tuttavia, l'esistenza di modelli multimodali (MLM), in grado di operare, sia in *input* che in *output*, su altre modalità oltre quella testuale.

[31] Da questo "giudizio" umano, anziché da una propria comprensione del senso della morale, deriva, ad esempio, il rifiuto di modelli come GPT di fornire indicazioni volte alla commissione di reati.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

"congelato", non traendo di fatto ulteriori indicazioni addestranti dall'utilizzo successivo[32]. Se gli iniziali modelli linguistici di piccole dimensioni si limitavano alla predizione testuale, i modelli di grandi dimensioni hanno dimostrato abilità emergenti (*emergent abilities*), ossia abilità – come quella di produrre poesia – inaspettate e non-programmate, che il sistema trae autonomamente dall'osservazione di *pattern* nelle informazioni su cui è addestrato. In questo modo, tra l'altro, i LLM sono in grado di generare risposte ragionevolmente sensate anche per scenari sui quali non erano stati addestrati.

Lo sviluppo e la diffusione dei LLM[33] e dei prodotti che ne forniscono l'interfaccia utente (*e.g.*, ChatGPT) ha parzialmente democratizzato l'uso dell'IA generativa, rendendola fruibile anche da coloro che non abbiano conoscenze tecniche specifiche. Questo ha portato a varie applicazioni di rilievo, ad esempio in ambito educativo, dove l'IA e i LLM hanno dimostrato del potenziale nell'assistere forme di apprendimento *online*[34] o di pedagogia basata sullo stimolo della curiosità[35], nonché come *partner* di conversazioni in lingua straniera[36].

Sebbene la maggior parte dei LLM continui ad essere addestrata su *corpora* in lingua in inglese, vanno celebrati tentativi di rendere la tecnologia più accessibile anche a persone di altre lingue e culture, ad esempio in modelli come il francese FlauBERT, il coreano KLUE-BERT, l'arabo AraBERT e gli indonesiani IndoLEM e IndoBERT. A Scao e altri, in particolare, si deve la creazione di BLOOM, il primo modello addestrato in maniera trasparente su 46 linguaggi naturali e 13 di programmazione; uno strumento *open source* dalle enormi potenzialità[37]. Similmente, un importante passo in direzione dell'accessibilità linguistica è stato compiuto con CINO, il primo *pre-trained model* multilingue (MPLM) a integrare dati non solo in cinese mandarino, ma anche in cantonese e in altre sei lingue riconducibili a minoranze etniche[38].

Ai nostri fini devono inoltre evidenziarsi – previo adeguato investimento di risorse e competenze – le notevoli potenzialità di impiego delle tecnologie linguistiche (e dei MPLM in particolare) a supporto delle minoranze linguistiche. Nel suo *report*, *e.g.*, Gerken fa riferimento al possibile utilizzo di traduttori automatici per rendere facilmente accessibili documenti ufficiali e altri testi; di *chatbot* per facilitare la comunicazione fra autorità amministrative e utenti (simili applicazioni sono già in uso, tra l'altro, in varie città tedesche); di sistemi di sintesi vocale (*speech syntesis*) per gli annunci di trasporto pubblico – che potrebbero essere, quindi, agevolmente forniti in molteplici lingue – e persino nei *media*, per la generazione automatica di sottotitoli. Simili strumenti potrebbero essere impiegati, autonomamente o in commistione con traduzioni umane più specializzate (e accurate), per mitigare l'onere pratico ed

[32] Ragion per cui si parla di *pre-trained models*; è verosimile che sviluppi futuri consentiranno ai modelli di continuare l'addestramento oltre questa fase iniziale.

[33] Datata al 2018 da E. Kasneci *et al.*, *ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education*, in *Learning and Individual Differences*, 3, 2023, 102 ss.

[34] M. Gerken, *op. cit.*, 12.

[35] R. Abdelghani *et al.*, *GPT-3-Driven Pedagogical Agents to Train Children's Curious Question-Asking Skills*, in *International Journal of Artificial Intelligence in Education*, 2, 2024, 483 ss.

[36] R. El Shazly, *Effects of Artificial Intelligence on English Speaking Anxiety and Speaking Performance: A Case Study*, in *Expert Systems*, 3, 2021.

[37] Si v., per un resoconto completo, T. Le Scao *et al.*, *Bloom: A 176B-Parameter Open-Access Multilingual Language Model*, 2023, disponibile su: inria.hal.science/hal-03850124/ (ultima consultazione 10/07/2024).

[38] Z. Yang *et al.*, *CINO: A Chinese Minority Pre-Trained Language Model*, in *Proceedings of the 29th International Conference on Computational Linguistics*, 2022, 3937 ss.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

economico degli interventi di sostegno linguistico[39]. Questo potrebbe a sua volta agevolare l'accesso e, entro certi termini, l'utilizzo delle lingue minoritarie, potenzialmente con esiti benefici sulla loro "vitalità" e integrazione nel più ampio tessuto socio-linguistico nazionale.

3.1 … e criticità nell'applicazione

Tuttavia, nonostante l'IA sia in grado, come visto, di offrire un promettente contributo sociale nella tutela della "diversità" – quantomeno linguistica – è altresì importante riflettere sulle significative limitazioni dimostrate dalle tecnologie in esame in vari ambiti, e sulle loro implicazioni. Tristemente noti sono, ad esempio, i casi delle decisioni discriminatorie adottate dai sistemi di valutazione assistiti dall'IA presso Amazon, che sistematicamente penalizzava le candidate in base al genere, e presso Ofqual (l'O*ffice of Qualifications and Examinations Regulation* britannico), che altrettanto sistematicamente sfavoriva, invece, gli esaminati appartenenti alla classe operaia e a minoranze etniche[40].
Se, da un lato, si può desumere che l'IA riproduca esempi di pregiudizio e ingiustizia sociale "estratti" dal mondo reale, è la carenza stessa di dati sufficientemente rappresentativi, dall'altro, che può indurla a *performance* discriminatorie: si pensi, *e.g.*, ai sistemi di riconoscimento facciale già in uso in Nord America, la cui affidabilità fluttua significativamente sulla base dell'etnia o del genere della persona[41].
È opportuno, in questo senso, distinguere fra parzialità, intesa come iniquità, e *bias*: sebbene l'IA possa, teoricamente, essere considerata imparziale, essa è senza dubbio prone a *bias*, o «errori sistematici (non casuali) di misurazione risultanti in dissimili livelli di accuratezza tra un gruppo e un altro, a fronte della realtà corrispettiva»[42].
Guardando specificamente alle tecnologie di linguaggio e ai LLM, c'è ampia evidenza, nella letteratura, che la mera inclusione nel periodo di addestramento del modello di lingue e culture altre rispetto a quella anglosassone – di fatto dominante nel settore – non è di per sé sufficiente a "sradicare" i *bias*. Modelli addestrati con *data set* in varie lingue (*e.g.*, GermanBERT, German GPT-2, RobBERT)[43] semplicemente trattengono gli stereotipi sessisti, agisti e razzisti delle società di riferimento; su venticinque diversi modelli testati da Adewumi e altri, tutti hanno evidenziato problematiche relative a stereotipi, misoginia, razzismo, agismo, nonché discriminazioni di genere, religione, cultura[44].
Per quanto esistano strumenti di detossificazione (*detoxification*) dei LLM, che prevedono l'epurazione dei *data set* di addestramento da ogni esempio di linguaggio potenzialmente discriminatorio, recenti analisi hanno gettato luce sulla scarsa affidabilità di queste pratiche, soprattutto a causa di correlazioni spurie che portano a segnalare come problematiche nozioni e forme espressive devianti dallo *standard*

[39] M. Gerken, *op. cit.*, *passim*.
[40] Come riportato da M. Nihei, *Epistemic Injustice as a Philosophical Conception for Considering Fairness and Diversity in Human-Centered AI Principles*, in *Interdisciplinary Information Sciences*, 1, 2022, 35 ss.
[41] M. Nihei, *op cit.*, 35.
[42] Secondo T. Adewumi *et al.*, *Fairness and Bias in Multimodal AI: A Survey*, 2024, disponibile su: arxiv.org/abs/2406.19097 (ultima consultazione 10/07/2024), che si rifà a B.M Booth *et al.*, *Bias and Fairness in Multimodal Machine Learning: A Case Study of Automated Video Interviews*, in *Proceedings of the 2021 International Conference on Multimodal Interaction*, 2021, 268 ss.; tr. it. dell'A.
[43] Incluso l'italiano; si v. T. Adewumi *et al.*, *op. cit*., per un resoconto più completo.
[44] T. Adewumi *et al.*, *op. cit.*

Downloaded from www.biodiritto.org.
ISSN 2284-4503

percepito (generalmente, il linguaggio della classe dominante). È questo il caso, ad esempio, di DAPT, che, adottato nella detossificazione di LLM in lingua inglese, provoca un aumento sproporzionato della perplessità (*perplexity)* e conseguente riduzione dell'efficacia del modello sui testi in inglese afro-americano (AAE) o contenenti menzioni di identità minoritarie, rispetto a quelli nell'inglese in uso presso le demografiche bianche (WAE); così facendo, detossificatori come DAPT «addestrano i modelli non solo a dimenticare la tossicità, ma anche l'AAE e le menzioni di identità minoritarie»[45], di fatto ereditando loro stessi i *bias* che sarebbero sviluppati per controllare[46].

I *bias* sopra citati, e le difficoltà incontrate nei tentativi di sradicarli, rappresentano, con ogni evidenza, un serio ostacolo alla promozione dei LLM quali strumenti di tutela linguistica e sociale. Le tecnologie di linguaggio sono state definite come «intrinsecamente politiche, poiché fautrici di processi di profondo cambiamento sociale»[47]; in quanto tali, esse godono di una posizione privilegiata relativamente alla loro abilità di influenzare (ed essere influenzate) da specifiche comunità. Questa influenza può dirsi "direzionata", poiché è tipicamente esercitata da comunità con lingue "ad alte risorse" a discapito delle lingue "a basse risorse". Il capitale linguistico e sociale di cui beneficiano le lingue "ad alte risorse" non è semplicemente riconducibile al numero di parlanti, ma è, piuttosto, l'esito di fenomeni imperialisti e neo-coloniali: basti pensare che il kiswahili, che con circa 80 milioni di locutori costituisce una delle principali lingue africane, è rappresentato su Wikipedia da tante pagine quante il bretone, una lingua protetta del gruppo celtico, attualmente parlata da circa 200 mila persone[48]. In questo senso, l'avanzamento tecnologico in ambito comunicativo (e non solo) altera i convenzionali parametri di marginalità, ampliando e complicando l'orizzonte dell'ingiustizia linguistica, e costringendo necessariamente ad una riflessione più intersezionale.

Sebbene si prospettino, infatti, significativi margini di miglioramento circa il repertorio di lingue accessibili tramite i LLM e il controllo dei potenziali *bias* loro associati, l'affidabilità di queste tecnologie rimarrà contenuta fintantoché esse continueranno a essere modellate su lingue e culture (occidentali) dominanti, venendo poi semplicemente "traslate" in contesti minoritari (quali che siano le circostanze socio-culturali che li rendono tali) [49]. Questo tipo di operazione rinforza forme di "ingiustizia epistemica", con la quale si identificano pratiche discriminatorie che screditano il contributo e la capacità conoscitiva di individui e comunità, distorcendone la percezione[50].

---

[45] A. Xu *et al.*, *Detoxifying Language Models Risks Marginalizing Minority Voices*, in *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2021, 2393, tr. it. dell'A.

[46] Si noti, infatti, che, secondo quanto riportato da A. Xu *et al.*, *op. cit*., questo è il caso anche quando il modello è sottoposto a valutazione umana (tramite i cosiddetti *crowdworkers*), e che aumentare l'intensità di detossificazione finisce per esacerbare ulteriormente il problema.

[47] Così P. Helm *et al.*, *Diversity and Language Technology: How Language Modeling Bias Causes Epistemic Injustice*, in *Ethics and Information Technology*, 2024, 7, rifacendosi a L. Winner, *The Whale and the Reactor: A Search for Limits in an Age of High Technology*, Chicago-Londra, 1988.

[48] Riprendendo l'esempio citato da P. Helm *et al.*, *op. cit*., 2.

[49] Per una discussione più ampia sull'ecologia linguistica globale e il ruolo della tecnologia, v. S. Bird, *Local Languages, Third Spaces, and Other High-Resource Scenarios*, in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, 2022, 7817 ss.

[50] Concetto che si deve alla filosofa Miranda Fricker; v., in particolare, M. Fricker, *Epistemic Injustice: Power and the Ethics of Knowing*, Oxford, 2009. Prima di lei, Gayatri Chakravorty Spivak aveva parlato di «colonizzazione

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Non solo, dunque, l'attuale diffusione delle tecnologie di linguaggio favorisce le lingue "ad alte risorse" (particolarmente quelle occidentali, e più di tutte l'inglese) nei vari modi sopra discussi, ma "offusca" anche la necessità di coinvolgere nella ricerca e nello sviluppo le lingue "a basse risorse" e le loro comunità di riferimento. Come già accennato, ad esempio, nonostante le tecnologie per la traduzione automatica (*e.g.*, Google Translate) possano essere percepite come altamente performanti in un'ampia casistica di operazioni, in realtà esse incorrono in varie difficoltà legate alla gestione degli elementi intraducibili, a carenze lessicali e alla necessità di ricorrere all'inglese come lingua *pivot* (in funzione intermediaria nella traduzione fra altre due lingue); l'utente è indotto a credere che la macchina abbia eseguito la traduzione in maniera efficace anche quando questo non è il caso, di fatto non ottenendo contezza dell'errore – e quindi della necessità di cercare una soluzione alternativa – e potenzialmente traendo impressioni fallaci circa la lingua (e cultura) *target*[51].

Le problematiche tecniche, epistemiche ed etiche discusse finora si possono ricondurre, almeno in parte, alla tendenza del settore dell'IA a procedere in relativo isolamento metodologico e teoretico, anche a causa della paucità di ricerca in questo ambito proveniente da altre discipline. Se è vero che un riscontro viene fornito dalle compagnie *high-tech* e dai linguisti computazionali che investono e lavorano per migliorare la *performance* di queste tecnologie, e che, soprattutto in ambito europeo, si stanno compiendo i primi tentativi di regolazione, complessivamente le iniziative volte ad accrescere rappresentatività e tutele rimangono al momento inadeguate. L'informatica, come d'altronde è il caso anche di altri settori, continua ad essere dominata dall'influenza anglosassone, che accentra su di sé ricerca e sviluppo, relegando a un ruolo marginale il contributo derivato dallo studio di altre comunità linguistiche[52]. Emerge chiaramente, dunque, la necessità di ricorrere a una maggiore integrazione delle varie competenze disciplinari coinvolte, così da ampliare la riflessione intorno al ruolo dei LLM quali strumenti di accesso e tutela linguistica (e culturale), e informarne le future implementazioni.

4. Intelligenza artificiale, minoranze (vecchie e nuove) e ingiustizia socio-linguistica: considerazioni conclusive

L'intelligenza artificiale, lo si è detto, può essere utilizzata a sostegno delle minoranze linguistiche, agevolando la fruizione di servizi (e, in parallelo, il godimento di diritti) da parte di queste ultime e di coloro che le compongono: le piattaforme di *online learning*, a mero titolo di esempio, facilitano l'apprendimento delle lingue regionali o minoritarie, contribuendo a preservarle e a promuoverle[53]; la traduzione automatica (*machine translation*) di atti e documenti ufficiali, anche giudiziari, appare particolarmente promettente, dato il linguaggio formulaico che spesso li contraddistingue[54]; diverse applicazioni di NLP – si pensi ai *chatbot* – sembrano in grado, a certe condizioni, di "avvicinare" le autorità

epistemica»; cfr. G.C. SPIVAK, *Can the Subaltern Speak*, in L. GROSSBERG, C. NELSON (a cura di), *Marxism and the Interpretation of Culture*, Chicago, 1988.

[51] Offrono una dettagliata descrizione di questo tipo di problemi, fornendo esempi in varie lingue, P. HELM *et al.*, *op. cit.*, 10 ss.

[52] P. HELM *et al.*, *op. cit.*, 12, parlano di un riposizionamento dell'inglese da *lingua franca* della comunicazione scientifica a oggetto *standard* della ricerca scientifica.

[53] M. GERKEN, *op. cit.*, 12 ss.

[54] Ivi, spec. 8 ss. e 14.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

amministrative agli utenti, consentendo a questi ultimi di comunicare, almeno in uno stadio preliminare o in fasi circoscritte dell'interazione, nella loro lingua d'elezione[55]. Gli esempi potrebbero continuare.

In ottica giuridico-costituzionale – non tanto e non solo in chiave di eguaglianza formale e non-discriminazione, ma anche e soprattutto di eguaglianza sostanziale e "rimozione degli ostacoli"[56] – l'implementazione di simili misure "positive" sarebbe senz'altro auspicabile, tantopiù a supporto delle minoranze "sprovviste di territorio" e delle nuove minoranze, la cui eterogeneità e non-territorialità, come visto[57], indebolisce le potenzialità d'inclusione degli strumenti tradizionalmente impiegati a questi fini (quando non ne rende troppo complessa o dispendiosa l'attuazione). L'eguaglianza, poi, torna rilevante come "filtro", come condizione o circostanza pratica di godimento delle situazioni giuridiche di vantaggio, e *in primis* dei diritti sociali: la parziale digitalizzazione dei servizi, in questo caso, contribuisce a renderli fruibili «senza distinzione (…) di lingua». C'è, ovviamente, un però.

I sistemi di IA, come ampiamente anticipato, spesso tendono non solo a riprodurre, ma ad «accelerare, esacerbare e amplificare l'impatto di "forze di oppressione" preesistenti»[58], in primo luogo (ma non esclusivamente) perché ricevono o sono addestrati su dati non sufficientemente rappresentativi delle differenze e disparità presenti nella realtà fisica e sociale, o su dati *biased* esibiti e recepiti come "neutri". Si è precedentemente fatto riferimento agli strumenti di traduzione automatica: il limitato orizzonte teorico che fa da sfondo allo sviluppo di tali tecnologie può portare a ridurre il linguaggio ai suoi aspetti più meccanici, trascurandone l'intrinseco carico culturale e identitario, con esiti in alcuni casi grossolani, non solo dal punto di vista linguistico, ma sociale. Di recente, a titolo di esempio, un gruppo di ricercatrici e ricercatori ha formato e annotato un *corpus* di testi giuridici redatti nella Provincia multilingue di Bolzano e tradotti (dall'italiano al tedesco o viceversa) con l'ausilio di sistemi di traduzione automatica[59]; gli errori più comuni in cui incorrevano le macchine – perlopiù perché addestrate su testi giuridici (anche) in lingua tedesca, ma non riferiti o riferibili alla "piccola" e peculiare realtà del *Südtirol* – erano soprattutto di natura «*terminological, phraseological and semantic*», ma comprendevano pure, tra l'altro, «*omissions, additions and problems related to gender-sensitive language*», oltre a diversi «*context-related issues*»[60].

[55] Ivi, 13 ss.

[56] Sul rapporto tra gli artt. 3, co. 2, e 6 Cost. v. *supra*, § 1, spec. nt. 3.

[57] *Supra*, spec. § 2.

[58] W. So, C. D'Ignazio, *Race-Neutral vs Race-Conscious: Using Algorithmic Methods to Evaluate the Reparative Potential of Housing Programs*, in *Big Data & Society*, 2, 2023, 5, tr. it. dell'A. Sul punto v. anche, *ex multis*, C. D'Ignazio, L.F. Klein, *Data Feminism*, Cambridge (MA), 2023, *passim*; R. Benjamin, *Race After Technology*, Cambridge, 2019, *passim*; V. Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police and Punish the Poor*, New York, 2018, *passim*.

[59] V. F. De Camillis *et al.*, *The MT@BZ Corpus: Machine Translation & Legal Language*, in *Proceedings of the 24th Annual Conference of the European Association for Machine Translation*, 2023, 171 ss., nonché, anche per un quadro più completo della storia e della pratica della traduzione istituzionale in Alto Adige, F. De Camillis, *200 Years of Institutional Translation in South Tyrol: From Civil Servants to Machines?*, in R. Martínez-Carrasco, A. Borja, Ł. Biel (a cura di), *Repensar la (des)globalización y su impacto en la traducción: desafíos y oportunidades en la práctica de la traducción jurídica*, n. mon. di *Monografías de Traducción e Interpretación*, 2024, 108 ss.

[60] F. De Camillis, *op. cit.*, 128.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Anche per questo sarà sempre necessario, fuor d'esempio, in prima battuta verificare la qualità e la rappresentatività dei dati di addestramento del sistema. Non basta, a tal fine, "depurare" i dati stessi, rimuoverne eventuali tratti stigmatizzanti e assicurarsi che siano rappresentativi; occorre, in aggiunta a ciò, arricchire i dati di contesto – storico, culturale e socio-economico – perché il sistema possa, non solo divenire meno dannoso, ma rivelarsi uno strumento funzionale e, idealmente, un veicolo positivo di eguaglianza[61]. Questo compito, al momento, è parzialmente assolto da operazioni di umana annotazione dei *data set* e valutazione della *performance* del modello (*i.e.*, il sopracitato *reinforcement learning with human feedback)*; le modalità e i parametri decisionali adottati nell'implementazione di tali interventi, tuttavia, sono spesso caratterizzati da scarsa chiarezza e, come si è già visto, di fatto presentano un'efficacia limitata.

Alla luce delle considerazioni finora svolte, emerge con chiarezza la necessità di una più significativa integrazione, nello sviluppo di tecnologie basate sull'NLP, di competenze che sappiano tener contro del quadro linguistico e giuridico caratterizzante le realtà fisico-sociali all'interno delle quali si propone l'adozione dell'IA. Una più stabile e sistematica *collaborazione interdisciplinare tra sviluppatori, linguisti e giuristi*, in particolare, potrebbe incoraggiare la (auspicabile) riconcettualizzazione del contributo che l'IA è in grado di apportare alla facilitazione della comunicazione umana, chiarendo, tra l'altro, quali comunità possano effettivamente o maggiormente beneficiarne (e quali ne risultino escluse).

Sarà necessario, da questo punto di vista, rivalutare le attuali definizioni di "minoranza" e "marginalizzazione", nonché contestualizzare la discussione sull'eguaglianza linguistica – oltre i semplici termini dell'"accesso" – nel più ampio e complesso quadro dei fenomeni neocoloniali e dell'ingiustizia epistemica. Solo realizzando questi interventi, tanto materiali quanto teorici, la collaborazione tra competenze multi-disciplinari potrà, forse, riuscire nel gravoso compito di incoraggiare il progresso tecnologico senza, però, perdere di vista le complessità etiche del suo rapido incedere nel contesto sociale.

[61] «*This is in line with work that advocates for moving away from the narrow idea of "bias" toward a more robust conceptual, computational, and historical modeling of "power" in algorithms and machine learning. […] The use of these methods is grounded in a theoretical shift – instead of conceiving of fairness as the absence of, or expunging of, racial classification from computational systems (race-neutral) we move toward an* antisubordination approach*, which contends that equal citizenship is not possible under the current social structure and requires the dismantling of racial stratification precisely by examining and attending to its racial effects (race-conscious)*» (così W. So, C. D'Ignazio, *op. cit.*, 5, enfasi aggiunta, con particolare riferimento alle discriminazioni per motivi di "razza", ma introducendo argomenti senz'altro estensibili, con alcuni aggiustamenti, alle altre forme e agli altri fattori di discriminazione).

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Discriminazioni algoritmiche e tutela dei consumatori vulnerabili nell'accesso al credito

Giulia Curcuruto, Paolo Inturri*

ALGORITHMIC DISCRIMINATION AND PROTECTION OF VULNERABLE CONSUMERS IN ACCESS TO CREDIT
ABSTRACT: Consumer credit carries risks of discrimination against certain categories of customers. This was originally due to people's involvement in evaluating credit applications. As a consequence, since the 80s intermediaries have used automated data processing systems, which are today complemented by AI models. Despite the several advantages, AI entails the risk of discriminatory consumer classification. This paper proposes an analysis of credit-scoring AI systems under the spectrum of Italian and EU legislation to understand whether they provide an adequate level of protection against the risk of algorithmic discrimination, also from a remedial point of view.

KEYWORDS: AI; consumer credit; discrimination; credit scoring; algorithms.

ABSTRACT: Il credito al consumo si caratterizza per il rischio di discriminazioni nei confronti di particolari categorie di clienti. Ciò in origine era dovuto all'impiego di persone fisiche nell'attività di verifica prodromica alla concessione del credito. Anche per tale ragione già a partire dagli anni 80' gli intermediari si sono avvalsi di sistemi automatizzati di elaborazione dati, oggi affiancati da modelli di IA. A fronte dei vantaggi, l'IA comporta rischi di classificazioni discriminatorie dei consumatori. Lo scritto propone un'analisi dell'impiego di sistemi di IA nella verifica del merito creditizio dei consumatori, alla luce della legislazione nazionale e sovranazionale, onde comprendere se fornisca un adeguato livello di tutela contro il rischio di discriminazione algoritmica, anche sotto il profilo rimediale.

PAROLE CHIAVE: IA; credito al consumo; discriminazione; credit scoring; algoritmi.

* *Giulia Curcuruto: dottoranda di ricerca in diritto commerciale, Università di Catania. Mail: giulia.curcuruto@phd.unict.it; Paolo Inturri: dottorando di ricerca in diritto costituzionale, Università di Catania. Mail: paolo.inturri@phd.unict.it.* *Sebbene il lavoro sia il risultato di una riflessione comune, i paragrafi 2, 3, 3.1, 3.2 e 3.3 sono da attribuire a Paolo Inturri e i paragrafi 4 e 5 a Giulia Curcuruto. I paragrafi 1 e 6 sono riferibili ad entrambi. Contributo sottoposto a doppio referaggio anonimo.*

Downloaded from www.biodiritto.org.
ISSN 2284-4503

1. Introduzione

Nel settore del credito la necessità di accertare il grado di solvibilità della parte debitoria risulta impellente per almeno due motivi: la massimizzazione del profitto dell'impresa finanziatrice e il mantenimento della stabilità delle condizioni macroeconomiche di mercato[1].

Per tale ragione, prima di stipulare un contratto di credito[2] i finanziatori espletano un procedimento di valutazione del merito creditizio (*creditworthiness*), teso a comprendere la probabilità di adempimento del potenziale beneficiario di un'erogazione.

In Italia il merito creditizio viene stimato impiegando per lo più modelli statistici econometrici[3]. Tuttavia, nel biennio 2023-2024 le imprese che operano nel *Fintech credit*[4] hanno investito 901 milioni di euro in progetti di sviluppo di tecnologie innovative, il 16,5% dei quali è destinato all'intelligenza artificiale (di seguito IA)[5].

La prassi dei Paesi in cui l'IA è già in uso nel settore dell'accesso al credito ha evidenziato il rischio di discriminazione che questa comporta per i consumatori appartenenti a categorie vulnerabili.

Per tale ragione, il presente scritto indaga le cause della discriminazione algoritmica dei consumatori nel procedimento di verifica del merito creditizio, quale premessa per comprendere se l'attuale quadro normativo nazionale e sovranazionale fornisca un adeguato livello di tutela, anche sotto il profilo rimediale.

2. Gli algoritmi di credit scoring

Per larga parte del XX secolo, il procedimento di verifica del merito creditizio veniva espletato ricorrendo esclusivamente a tecniche di valutazione discrezionali (*judgemental system*)[6], così denominate in quanto l'analisi di ciascuna istanza di credito è rimessa alla valutazione individuale di una persona fisica.

[1] Sull'incidenza delle prassi di scorretta valutazione del merito di credito nella crisi dei mutui *sub prime* v., *ex multis,* F. CAPRIGLIONE, *I "prodotti" di un sistema finanziario evoluto. Quali regole per le banche? Riflessioni a margine della crisi causata dai mutui sub-prime*, in *Banca borsa titoli di credito*, 1, 2008, 20 ss.

[2] L'attività valutativa, sia pure con dinamiche parzialmente diverse, investe anche le fasi del rapporto di credito successive alla sua instaurazione: svolgimento (monitoraggio e controllo delle sopravvenienze) e chiusura (recupero, anche in via forzata, dell'erogato). Sul punto si rinvia a A.A. DOLMETTA, *La valutazione del merito del credito nell'accesso al servizio. La prospettiva del contratto di impresa*, in *Banca borsa titoli di credito*, 3, 2023, 307.

[3] E. BONACCORSI DI PATTI ET AL., *Intelligenza artificiale nel credit scoring. Analisi di alcune esperienze nel sistema finanziario italiano*, in *Questioni di economia e finanza (occasional papers*), 721, 2022, 30.

[4] Con l'espressione ci si riferisce all'impiego di vari strumenti frutto dell'innovazione tecnologica nel settore del credito. Sul tema v. M. CIAN, C. SANDEI, *Diritto del Fintech*, Padova, 2020.

[5] A. SCOGNAMIGLIO, M. BERRUTI, (a cura di), *Indagine Fintech nel sistema finanziario italiano,* Roma, 2024, 7-8.

[6] Per un esempio di *judgemental system* v. C.A. BANA ET AL., *Qualitative Modelling of Credit Scoring: A Case Study in Banking*, in *European Research Studies*, 5, 1–2, 2002, 37–51.

Downloaded from www.biodiritto.org.

ISSN 2284-4503

Al netto dei benefici di tali sistemi[7], l'elevato grado di soggettività coinvolto implica, tra i principali rischi, quello di discriminare i consumatori nell'accesso al credito[8].

Successivamente, l'evoluzione dell'impresa bancaria verso un assetto più impersonale e l'attenzione nei confronti di modelli matematici capaci di oggettivizzare e automatizzare il processo di valutazione del merito creditizio hanno condotto al progressivo abbandono dei *judgemental system* in favore dei modelli di *credit scoring*[9].

Con quest'ultima espressione ci si riferisce all'impiego di metodi statistici per l'elaborazione dei dati rilevanti in *output* numerici che indicano il profilo di rischio, affidabilità e puntualità nei pagamenti di ciascun potenziale cliente. Tipicamente, maggiore è il punteggio (*credit score*) ottenuto, minore è il rischio di inadempimento[10].

I sistemi di *credit scoring* si basano su modelli statistici econometrici in grado di restituire il punteggio di credito di un consumatore grazie all'esclusiva analisi delle variabili statisticamente correlate con le performance di pagamento ovverosia tramite l'elaborazione dei c.d. dati finanziari, tradizionalmente detenuti dai finanziatori e dalle banche dati creditizie[11].

Ciononostante, la maggiore oggettività che contraddistingue i modelli in esame non li impermeabilizza dal rischio di generare risultati discriminatori. Difatti, sono esposti a errori di classificazione, tali per cui la presenza casuale di alcune caratteristiche porterà taluni richiedenti a sembrare non meritevoli, quando è ragionevole presumere il contrario. Tuttavia, si tratta di un problema attenuabile sfruttando uno storico degli errori ai fini di un aggiornamento costante. Inoltre, tali modelli permettono di ricostruire l'influenza delle variabili considerate rispetto al risultato prodotto, il che non è del tutto possibile con riferimento alle decisioni umane[12].

Nonostante le performance garantite dai sistemi di *credit scoring* tradizionali, la naturale ricerca della massimizzazione del profitto ha condotto le imprese che operano nel settore del credito a cercare nell'IA uno strumento per ottenere risultati maggiormente accurati, anche in ragione della varietà e della tipologia di dati che riesce a processare[13].

---

[7] Per un analisi dei *judgemental system* v. G.G. CHANDLER, J.Y. COFFMAN, *A comparative analysis of empirical vs. judgemental credit evaluation,* in *The Journal of Retail Banking*, 1, 2, 15-26.

[8] Con particolare riferimento alle discriminazioni razziali negli U.S.A. v., *ex plurimis*, L. RICE, D. SWESNIK, *Discriminatory Effects of Credit Scoring on Communities of Color*, in *Suffolk University Law Review*, 46, 3, 2013, 935-942.

[9] Per una breve ricostruzione della storia del *credit scoring* v. L.C. THOMAS, D.B. EDELMAN, J.N. CROOK, *Credit scoring and its applications*, Philadelphia, 2017, 2-5.

[10] THE WORLD BANK GROUP, *Credit scoring approaches guidelines*, 2019, 3.

[11] Ad esempio, rientrano in tale categoria i dati bancari transazionali (registrazioni di ritardi di pagamento di crediti attuali e passati, importi e scopo dei prestiti, la storia creditizia), il numero di istanze di credito registrate dalle banche dati creditizie, i dati commerciali (scritture contabili).

[12] H.A. ABDOU, J. POINTON, *Credit scoring, statistical techniques and evaluation criteria: a review of the literature*, in *Intelligent Systems in Accounting, Finance & Management*, 18, 2-3, 4-6.

[13] Sul punto cfr. *ex plurimis,* A. FUSTER ET AL., *Predictably unequal? The effects of machine learning on credit markets*, in *The Journal of Finance*, 77, 1, 2022, 5-47. Critici sulle *performance* dell'IA nel *credit scoring* A. WANG ET AL., *Against Predictive Optimisation: On the Legitimacy of Decision-Making Algorithms that Optimize Predictive Accuracy*, 16 febbraio 2023.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

In particolare, la capacità dell'IA di sfruttare i *big data* permette di includere nel processo di *credit scoring* anche i dati alternativi che ogni utente genera attraverso l'uso di internet[14].
In tal modo, le tipologie di dati che l'IA consente di sfruttare nel *credit scoring* sono riconducibili alle seguenti quattro categorie: finanziari strutturati[15]; non finanziari strutturati[16]; finanziari non strutturati[17] e non finanziari non strutturati[18].
Proprio la capacità di sfruttare dati alternativi permette all'IA di determinare il punteggio di credito estraendo delle variabili che non hanno una chiara relazione economica e che, pertanto, non potrebbero essere considerate da un modello basato unicamente su dati finanziari[19]. Ad esempio, l'impiego di dati ricavati dagli *smartphone* (geolocalizzazione, transazioni) e dai *social media* (numero e frequenza dei post, interazioni) consente all'IA di tracciare lo stile di vita dell'utente anche in termini di spese e propensione a ripagare i debiti[20].
Inoltre, ciò si traduce in un accrescimento dell'inclusione finanziaria, permettendo di valutare consumatori altrimenti esclusi o penalizzati[21]. Difatti, il solo impiego di dati finanziari comporta l'inevitabile estromissione dal credito di quei soggetti che siano sprovvisti, per le ragioni più varie, di una storia finanziaria (giovani adulti, immigrati, minoranze etniche)[22].

3. La discriminazione algoritmica nell'accesso al credito

Numerosi studi economici hanno da tempo evidenziato come il rischio di discriminazione, basato su fattori quali razza, etnia e genere, sia endemico al procedimento di valutazione del merito creditizio, indipendentemente dal modello adottato[23].
Dal punto di vista tipologico è possibile distinguere tra discriminazione formale (*disparate treatment*) e sostanziale (*disparate impact*)[24]: nel primo caso, due soggetti nella medesima situazione vengono trattati diversamente, proprio in ragione dell'appartenenza ad una specifica classe; nel secondo, la

[14] Cfr. Per approfondire v. AA.VV., *On the Rise of Fintechs: Credit Scoring Using Digital Footprints*, in *The Review of Financial Studies*, 2020, 33, 2020, 2845-2897.
[15] Ad esempio, indicatori patrimoniali ed economico-finanziari sull'andamento dei conti, dei pagamenti, del mercato.
[16] Come i dati di tipo sociodemografico.
[17] Si tratta dei dati ricavati dall'analisi delle transazioni e dall'attività di *open banking*.
[18] È il caso del *digital footprint*, dei dati di navigazione e dei social network.
[19] E. BONACCORSI DI PATTI ET AL., *op. cit.*, 14-15.
[20] In generale sulla possibilità di utilizzare i *big data* per ricavare il comportamento degli agenti economici v. B. DESAMPARADOS, J. DOMENECH, *Big Data Sources and Methods for Social and Economic Analyses*, in *Technological Forecasting and Social Change*, 130, 2018, 99–113.
[21] Sui vantaggi in termini di inclusione finanziaria dell'impiego di dati alternativi v. M.M. SMITH, C. HENDERSON, *Beyond Thin Credit Files*, in *Social Science Quaterly,* 99, 2018, 24-42.
[22] Così, E. TEDESCHI, *Credit scoring algoritmico: benefici e rischi della creditworthiness tramite artificial intelligence alternative data e digital footprints*, in C. CAMARDI (a cura di), *La via europea per l'intelligenza artificiale. Atti del Convegno del Progetto Dottorale di Alta Formazione in Scienze Giuridiche*, Venezia, 2021, 328.
[23] Per una rassegna di letteratura in materia, con particolare riferimento alla discriminazione algoritmica, v. A.C.B. GARCIA, M.G.P. GARCIA, R. RIGOBON, *op. cit*.
[24] S. BAROCAS, A.D. SELBST, *Big Data's Disparate Impact*, in *California Law Review*, 104, 3, 2016, 694-711.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

diseguaglianza costituisce il risultato di una pratica commerciale formalmente neutra, ma fattivamente svantaggiosa per determinati gruppi sociali.
Sotto il profilo degli effetti, le discriminazioni nell'accesso al credito si traducono nel diniego dell'istanza di credito oppure in una discriminazione di prezzo[25].
Ciò che contraddistingue i procedimenti di *credit scoring* basati sull'IA sono le cause della discriminazione. Infatti, determinanti in tal senso non potranno che essere le scelte adottate nelle diverse fasi di sviluppo dell'algoritmo, le quali volontariamente o involontariamente possono determinare esiti discriminatori.

3.1. La fase di raccolta dei dati

Innanzitutto, nella fase di raccolta dei dati risulteranno fondamentali le decisioni che riguardano la composizione del *dataset* di addestramento.
Poiché gli algoritmi di ML apprendono per esempi, non possono che generare *output* influenzati dai dati su cui sono stati allenati. Dunque, se questi ultimi presentano dei bias, l'algoritmo potrebbe riprodurli nei risultati[26].
Nel caso in cui il *dataset* ricomprenda dati tradizionali i pregiudizi possono derivare da una inadeguata rappresentatività del campione di consumatori di riferimento, tale per cui specifiche categorie di individui risulteranno sovrarappresentate, sottorappresentate o escluse. Nonostante esistano approcci statistici consolidati per l'attività di campionamento, la generazione e la raccolta dei dati implica un ineliminabile grado di soggettività, che può viziarli con forme di bias storici o istituzionali.
Inoltre, taluni pregiudizi possono derivare dall'etichettatura dei dati di addestramento degli algoritmi di ML supervisionato, in cui l'algoritmo è addestrato alla modellazione della variabile dipendente attraverso dei dati di addestramento contrassegnati con delle *class label*.
Più nello specifico, durante il procedimento di etichettatura possono presentarsi dei casi di cui non è pacifico l'inserimento in una delle *class label* predeterminate. Così, ad esempio, se si cerca di determinare il merito creditizio dei consumatori utilizzando i dati relativi ai pagamenti delle rate delle carte di credito, risulta del tutto discrezionale il numero di rate inadempiute superato il quale un consumatore deve essere etichettato come immeritevole[27].
Invece, nell'ipotesi di un *dataset* composto esclusivamente da dati alternativi non è possibile realizzare un campionamento rappresentativo, poiché può essere costituito soltanto da quei consumatori che hanno lasciato una impronta digitale, ad eccezione, quindi, di tutti quegli individui esclusi o scarsamente presenti nell'infosfera[28] per ragioni, ad esempio, anagrafiche, culturali, economiche[29]. Tuttavia,

[25] Il fenomeno della discriminazione di prezzo si realizza allorquando il medesimo bene è venduto a prezzi differenti a diverse categorie di acquirenti, in presenza del medesimo costo di produzione. Così, A KOUTSOYIANNIS, *Modern Microeconomics*. Londra, 1979, 192. Sulle discriminazioni di prezzo nell'accesso al credito v. A.C.B. GARCIA, M.G.P. GARCIA, R. RIGOBON, *op. cit*., 3-4.
[26] Approfonditamente sul punto S. BAROCAS, A.D. SELBST, *op. cit*., 680-687.
[27] Sul punto v. J.H. DAVID, *Classifier Technology and the Illusion of Progress*, in *Statistical Science*, 21, 1, 2006, 10.
[28] L. FLORIDI, *La quarta rivoluzione. Come l'infosfera sta cambiando il mondo*, Milano, 2017.
[29] Sull'esclusione dai *big data* v. J. LERMAN, *Big Data and Its Exclusions*, in *Stanford Law Review. Online*, 66, 2013, 55-63.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

ove impiegati congiuntamente ai dati strutturati, incrementano la rappresentatività del campione grazie all'alto livello di ricchezza e granularità.
Infine, indipendentemente dalla tipologia di dati considerati, i risultati discriminatori possono risultare dalla selezione delle caratteristiche (*feature*) di *creditworthiness*[30].

3.2. La fase di sviluppo e addestramento

Ulteriori risultati discriminatori possono derivare dalla fase di sviluppo e addestramento del modello di IA[31].
A volte l'algoritmo aderisce con eccessiva fedeltà ai dati di allenamento, dando luogo a un modello incapace di prevedere accuratamente allorquando alimentato con *input* diversi dai dati di addestramento (c.d. *overfitting*).
Al contempo, è possibile che il modello sviluppato risulti troppo semplice per cogliere puntualmente la relazione tra le variabili di *input* e di *output*, producendo, conseguentemente, dei risultati connotati da un elevato tasso di errore (c.d. *underfitting*).
In entrambi i casi, il modello non potrà raggiungere un grado di generalizzazione tale per assegnare dei punteggi in grado di rispecchiare l'effettiva *creditworthiness* dei consumatori.
Ancora, può accadere che non venga adeguatamente considerata la tipologia di errore nei risultati generati dal modello.
Nello specifico, i sistemi di *credit scoring* possono dar luogo a diversi tipi di errore.
Oltre alle ipotesi di sovrastima e sottostima — quando ad un consumatore meritevole viene attribuito rispettivamente uno *score* maggiore o inferiore rispetto a quello corretto, con ciò che ne consegue sotto il profilo delle condizioni contrattuali — il sistema può dar luogo a falsi positivi, quando il credito viene concesso ad un consumatore immeritevole, e falsi negativi, quando il credito viene negato ad un consumatore meritevole.
Ebbene, durante l'addestramento è imprescindibile scegliere quali errori minimizzare ovverosia spiegare al modello se, ad esempio, sia preferibile che per errore ad un consumatore meritevole non venga concesso credito o, viceversa, che ad uno non meritevole venga concesso.
Indipendentemente dall'imperizia degli sviluppatori rispetto al problema *de qua*, la scelta non appare di poco conto. I finanziatori preferiscono che venga data prevalenza ai falsi negativi, in modo da minimizzare il rischio di impresa dovuto alle insolvenze. Tuttavia, ove tale errore sia dovuto a forme di bias contenuti nei dati, la mancata considerazione della tipologia di errore si traduce in una sistematica esclusione dal credito dei consumatori danneggiati dal pregiudizio.
Analoghi problemi possono derivare dalla mancata considerazione di variabili influenti sul risultato o nella scorretta aggregazione dei dati.
In breve, tali omissioni possono condurre al paradosso di *Simpson*[32] ovverosia l'inversione o la distorsione nella relazione tra due variabili quando si analizzano dati aggregati, rispetto a quando si

[30] Sul tema v. S. BAROCAS, A.D. SELBST, *op. cit.*, 688-690.
[31] Per una sintetica esposizione del problema v. E. BONACCORSI DI PATTI ET AL., *op. cit.*, 21- 22.
[32] E.H. SIMPSON, *The Interpretation of Interaction in Contingency Tables*, in *Journal of the Royal Statistical Society. Series B (Methodological)*, 13, 2, 1951, 238–41.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

considerano disaggregati. Ad esempio, se i dati di addestramento contengono solo esempi di consumatori meritevoli di età anziana, la mancata considerazione della variabile anagrafica potrebbe tradursi in risultati detrimenti per i giovani adulti.
Infine, è anche possibile che gli effetti discriminatori si realizzino allorché le variabili effettivamente necessarie per valutare il merito creditizio siano al contempo *proxy* circa l'appartenenza ad una classe di consumatori svantaggiata[33]. In altri termini, è possibile che i dati impiegati per la misura della *creditworthiness* costituiscano, al contempo, una variabile alternativa per profilare indirettamente i consumatori in termini di appartenenza ad una specifica classe, sicché una volta categorizzati in tal modo l'algoritmo potrebbe discriminarli in ragione dei bias esistenti a detrimento di tali gruppi.

3.3. La fase di analisi dell'output

L'ultima fase in cui può darsi luogo a fenomeni di discriminazione è quella dell'analisi dei risultati generati dall'algoritmo, nei casi in cui, la decisione di erogazione credito non sia totalmente automatizzata[34].
Numerose sono le ragioni che possono condurre ad una scorretta interpretazione dell'*output* generato dall'algoritmo, per cui, lungi dall'enumerarle[35], ci si limita a rilevare come esse siano la conseguenza dell'ineliminabile soggettività e fallibilità di ogni attività umana.
Tuttavia, ciò che contraddistingue in tale fase l'impiego dell'IA dai modelli econometrici tradizionali è la difficoltà di interpretarne i risultati e di ricostruirne i processi logici, in ragione della complessità del modello (c.d. *black box effect*).
Di conseguenza, in presenza di un risultato discriminatorio l'analista non potrebbe agevolmente comprenderne le cause, in assenza di sistemi di *explainable AI*[36], e, soprattutto, potrebbe manifestare ritrosie in tal senso derivanti dall'inconscio affidamento nelle superiori capacità computazionali dell'IA (c.d. *anchoring effect*)[37].
Alla luce di tutto quanto sopra affermato, emerge che i rischi di *output* discriminatori non possono essere cancellati del tutto secondo un modello di tutela *by design*, stante l'ineliminabile tasso di soggettività coinvolto.
In caso contrario, il rischio è quello di generare un ciclo di feedback in cui nel tempo la distorsione è confermata e rinforzata. In altri termini, la negazione sistematica del credito a danno di specifici gruppi sociali causata da un modello non a regola d'arte, contribuisce a determinare un bias storico nei dati, che si rifletterà nei campioni estratti dalla popolazione, sulla base dei quali verrà aggiornato lo stesso modello distorto[38].

[33] S. BAROCAS, A.D. SELBST, *op. cit*., 21-22.
[34] A.C.B. GARCIA, M.G.P. GARCIA, R. RIGOBON, *op. cit*., 5.
[35] Per una rassegna sintetica v. E. BONACCORSI DI PATTI ET AL., *op. cit.*, 21- 22.
[36] Sulla cd. *explainable AI* v. R. CALEGARI, A. OMICINI, G. SARTOR, *Explainable and Ethical AI: A Perspective on Argumentation and Logic Programming*, in M. BALDONI, S. BANDINI (a cura di), *AIxIA 2020 – Advances in Artificial Intelligence, AIxIA 2020*. *Lecture Notes in Computer Science*, Berlino, 2020, 12414.
[37] A. SIMONCINI, S. SUWEIS, *Il cambio di paradigma nell'intelligenza artificiale e il suo impatto sul diritto costituzionale*, in *Rivista di filosofia del diritto*, 1, 2019, 100.
[38] E. BONACCORSI DI PATTI ET AL., *op. cit.*, 20.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Tuttavia, orientare la programmazione e l'impiego delle tecnologie di IA in chiave antidiscriminatoria implica dei costi aggiuntivi per le imprese, i quali potrebbero non essere ritenuti adeguatamente bilanciati dai benefici che ne derivano (ad esempio in termini di *corporate reputation*).
Per tale ragione il paragrafo successivo è dedicato alla ricostruzione del quadro normativo in materia, onde comprendere se fornisca un adeguato livello di tutela avverso l'adozione di decisioni sul merito creditizio basate su *output* discriminatori generati dall'IA.

4. Il quadro normativo: l'ordinamento settoriale del credito

L'accertamento del grado di solvibilità degli aspiranti prenditori rileva, anzitutto, sul piano della disciplina prudenziale, per tale intendendosi l'insieme di regole volte garantire la stabilità degli intermediari e del sistema finanziario nel suo complesso. Invero, la corretta stima della *probability of default* costituisce il primo presidio per il contenimento del rischio di credito e consente l'esatta determinazione della dotazione patrimoniale regolamentare della banca[39].
A questo scopo, le disposizioni di vigilanza della Banca d'Italia richiedono al finanziatore di acquisire, in fase precontrattuale, *«tutta la documentazione necessaria per effettuare un'adeguata valutazione del merito di credito del prenditore, sotto il profilo patrimoniale e reddituale»* e di adottare procedure di sfruttamento delle informazioni che forniscano *«indicazioni circostanziate sul livello di affidabilità del cliente (ad es., attraverso sistemi di* credit scoring *e/o di* rating*)»*[40].
Le prescrizioni, pur non essendo riferite direttamente all'utilizzo dell'IA, sono da ritenere applicabili indipendentemente dalle tecniche di stima adottate dagli intermediari; se ne deduce, quale presupposto minimo ed indefettibile, che la valutazione algoritmica deve almeno essere in grado di restituire una rappresentazione veritiera della solvibilità del cliente.
Diversamente, gli Orientamenti ABE in materia di concessione e monitoraggio dei prestiti (EBA/GL/2020/06) prendono in considerazione l'ipotesi in cui la concessione avvenga a seguito del trattamento automatizzato dei dati, o persino avvalendosi di *tecnologie innovative.* In questi casi, gli intermediari devono dotarsi di politiche o di procedure dalle quali risultino le condizioni per l'applicazione di decisioni automatizzate[41]; devono essere in grado di comprendere il funzionamento dei modelli, i dati inseriti, le ipotesi, i limiti e i risultati; devono prevenire possibili distorsioni; devono accompagnare al modello automatizzato meccanismi di controllo del risultato e di *«override»* che incorpori il giudizio di esperti[42]. Nell'ipotesi (ulteriore e diversa) in cui vengano utilizzate tecnologie definite

[39] V., *ex multis,* C. Brescia Morra, *Il diritto delle banche. Le regole dell'attività,* Bologna, 2020, 202 ss.; P. Bontempi, *Diritto bancario e finanziario*, Milano, 2023, 93 ss.
[40] V. Circolare della Banca d'Italia n. 285 del 17 dicembre 2013 Parte I, Tit. IV, Cap. 3, Allegato A. Disposizioni identiche, con riferimento agli intermediari di cui al Titolo V del t.u.b., sono contenute nella Circolare della Banca d'Italia n. 288 del 3 aprile 2015, Parte I, Tit. III, Cap. I, Sez. VII, par. 2.
[41] In particolare, al paragrafo 4.3, punto 38, lettera g) degli Orientamenti viene specificato che «*Le politiche e procedure relative al rischio di credito dovrebbero specificare […] le condizioni per l'applicazione di decisioni automatizzate nel processo di concessione del credito, compresa l'identificazione dei prodotti, segmenti e limiti per i quali sono consentite le decisioni automatizzate*».
[42] Cfr. i punti 54 e 55 degli Orientamenti.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

innovative, come l'IA, gli enti devono, altresì, essere in grado di comprendere e prevenire i rischi specifici legati al funzionamento della tecnologia in uso[43].

Infine, all'interno della disciplina prudenziale, è utile volgere lo sguardo al Regolamento (UE) n. 575/2013 (c.d. *Capital Requirements Regulation*, di seguito CRR) applicabile agli intermediari che utilizzano, previa autorizzazione dell'autorità di vigilanza, modelli interni per la misurazione del rischio di credito (*internal ratings-based* o IRB)[44]. Difatti, fermo restando che tali norme sono applicabili soltanto ai modelli destinati al calcolo dei requisiti patrimoniali, esse forniscono indicazioni utili per lo sviluppo di buone prassi nell'impiego di sistemi di IA[45].

In particolare, l'art. 174 CRR descrive le caratteristiche che devono possedere i modelli statistici e gli altri metodi automatici per l'assegnazione delle esposizioni a classi o a pool relativi a debitori o ad operazioni. Le prescrizioni più interessanti ai nostri fini sono quelle relative alla selezione del campione di dati e alla necessarietà dell'intervento umano. Più in dettaglio, la norma richiede che i dati siano accurati, completi e pertinenti, nonché rappresentativi dell'effettiva popolazione di debitori o di esposizioni dell'ente e che il modello statistico sia combinato con la valutazione e la revisione umana[46] *«in modo da verificare le assegnazioni effettuate in base al modello e da assicurare che i modelli siano utilizzati in modo appropriato»* (lett. e).

Dall'insieme delle disposizioni che compongono la disciplina prudenziale può evincersi che l'intermediario, qualora intenda avvalersi dell'IA per l'adozione di decisioni sulla concessione del credito, dovrebbe essere in grado di comprendere il funzionamento dei modelli, di monitorare gli *output* e, eventualmente, di superare le determinazioni dell'algoritmo. La finalità esclusiva di tali disposizioni resta, tuttavia, quella di garantire la stabilità dell'intermediario e, in ultima istanza del sistema, essendo estranea agli scopi della normativa prudenziale la tutela dei soggetti vulnerabili, demandata alla normativa sulla trasparenza delle operazioni e dei servizi bancari e finanziari.

Quest'ultima, in via diretta, persegue l'obiettivo di assicurare la correttezza del comportamento contrattuale degli intermediari nei confronti dei clienti e, soltanto in via indiretta, concorre alla loro sana e prudente gestione[47]. Le norme in tema di trasparenza sono contenute nel Titolo VI del Testo unico

[43] Cfr. il punto 53 degli Orientamenti.

[44] In questa sede basti ricordare che l'accordo di Basilea del 2004 ha introdotto la possibilità di adottare due diversi metodi per misurare la rischiosità delle esposizioni: il metodo standard e il metodo dei *rating* interni. Il metodo standard prevede la suddivisione delle esposizioni in classi di rischio, a tali esposizioni viene attribuita una ponderazione sulla base della valutazione del merito creditizio fatta da agenzie esterne di valutazione del merito di credito (ECAI), riconosciute sulla base di una procedura stabilita da un regolamento europeo. Il metodo dei *rating* interni prevede, invece, due varianti: "di base" e "avanzato". Nel primo, la banca stima direttamente la probabilità di insolvenza; nel secondo, destinato unicamente agli intermediari che soddisfano determinati requisiti, è rimessa alla banca anche la stima delle variabili di rischio. V., per un quadro generale, C. BRESCIA MORRA, *op. cit.*, p. 205.

[45] E. BONACCORSI DI PATTI ET AL, *op. cit.*, 21 ss.

[46] Critico sul requisito normativo del controllo umano B. GREEN, *The flaws of policies requiring human oversight of government algorithms*, in *Computer Law & Security Review*, 45, 2022, 1-22.

[47] Sul tema v., *ex multis*, A. MIRONE, *Le regole dell'attività: la tutela del cliente*, in M. CIAN (a cura di), *Manuale di diritto commerciale*, Torino, 2021, 734 ss.; A. MIRONE, *La trasparenza bancaria*, Padova, 2012.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

bancario e, nello specifico, gli articoli 124 *bis*[48] e 120 *undecies*[49] sono dedicati alla verifica del merito creditizio.

Entrambe le disposizioni, pur fornendo indicazioni sulla quantità e la qualità delle informazioni da utilizzare ai fini di una corretta valutazione, non prendono in considerazione l'ipotesi in cui l'elaborazione dei dati avvenga in maniera automatizzata oppure ricorrendo all'IA. L'unico laconico avvertimento è contenuto all'art. 120 *undecies*, comma 5, ai sensi del quale «*quando la domanda di credito è respinta, il finanziatore informa il consumatore senza indugio del rifiuto e, se del caso, del fatto che la decisione è basata sul trattamento automatico di dati*».

Il vigente quadro normativo è stato, perciò, reputato insufficiente a fornire un adeguato livello di tutela del consumatore alla luce della digitalizzazione che ha investito il mercato del credito, ragion per cui il legislatore dell'U.E. è tornato sul tema con la Direttiva 2023/2225/UE (di seguito CCD II)[50].

Innanzitutto, il legislatore euro-unitario ha fissato rigidi limiti alla tipologia di dati che possono essere utilizzati, facendo divieto di impiegare quelli «*che rivelino l'origine razziale o etnica, le opinioni politiche, le convinzioni religiose o filosofiche, o l'appartenenza sindacale, nonché trattare dati genetici, dati biometrici intesi a identificare in modo univoco una persona fisica, dati relativi alla salute o alla vita sessuale o all'orientamento sessuale della persona*» e i dati tratti dai *social network*[51].

---

[48] L'art. 124 *bis* è stato introdotto in recepimento dell'art. 8 della Direttiva 2008/48/CE, oggi abrogata e sostituita dalla Direttiva 2023/2225/UE, relativa ai contratti di credito ai consumatori. L'articolo in parola prevede che, prima della conclusione del contratto di credito, il finanziatore debba valutare il merito creditizio del cliente sulla base di informazioni adeguate o fornite dallo stesso o, se necessario, reperite attraverso la consultazione di banche dati. Il secondo comma impone al finanziatore di aggiornare le informazioni finanziarie di cui dispone se le parti convengono di modificare l'importo totale del credito dopo la conclusione del contratto e di ripetere la valutazione qualora si proceda ad un aumento significativo dell'importo totale del credito. V. R. De Chiara, *Commento all'art. 124 bis*, in F. Capriglione (a cura di), *Commentario al Testo Unico delle leggi in materia bancaria e creditizia*, III, Padova, 2016, 2161.

[49] L'art. 120 *undecies* è stato introdotto in recepimento dell'art. 18 della Direttiva 2014/17/UE in merito ai contratti di credito ai consumatori relativi a beni immobili residenziali. La norma ha imposto al finanziatore lo svolgimento di *«una valutazione approfondita del merito creditizio del consumatore, tenendo conto dei fattori pertinenti per verificare le prospettive di adempimento da parte del consumatore degli obblighi stabiliti dal contratto»*. Le informazioni da utilizzare per la valutazione devono riguardare la situazione economica e finanziaria e devono essere necessarie, sufficienti, proporzionate e verificate. Tali informazioni possono essere fornite dallo stesso consumatore al quale il finanziatore può richiedere chiarimenti (comma 2). L'articolo richiede che la capacità di rimborso debba essere valutata prescindendo dall'eventuale presenza di garanzie. Inoltre, detta valutazione deve essere ripetuta, sulla base di informazioni aggiornate, prima di procedere ad un aumento significativo dell'importo concesso a credito (comma 4). V.R. Grasso, *Articolo 120-undecies. Verifica del merito creditizio*, in F. Capriglione (a cura di), *Commentario al Testo Unico delle leggi in materia bancaria e creditizia*, III, Padova, 2016.

[50] La Direttiva 2023/2225/UE è stata pubblicata il 30 ottobre 2023 sulla G. U. dell'U. E. e dovrà essere adottata dagli stati membri entro il 20 novembre 2025, sarà applicabile ai contratti di credito in corso al 20 novembre 2026. Esplicativo è il considerando n. 4 *«La digitalizzazione ha contribuito a sviluppi di mercato che non erano previsti quando la direttiva 2008/48/CE è stata adottata. I rapidi sviluppi tecnologici registrati dall'adozione di tale direttiva hanno difatti apportato cambiamenti significativi al mercato del credito al consumo, sia sul versante dell'offerta che su quello della domanda, come la comparsa di nuovi prodotti e l'evoluzione del comportamento e delle preferenze del consumatore».*

[51] Il principio è codificato, oltre che al considerando n. 57, negli articoli 18, comma 3 *(«Obbligo di verifica del merito creditizio del consumatore»)* secondo il quale *«i social network non sono considerati una fonte esterna ai fini della presente direttiva»* e 19 *(«Banche dati»)* secondo il quale *«I creditori e gli intermediari del credito non trattano le categorie particolari di dati di cui all'articolo 9, paragrafo 1, del regolamento (UE) 2016/679 e i dati*

Downloaded from www.biodiritto.org.
ISSN 2284-4503

In secondo luogo, la tutela del consumatore è stata realizzata attraverso la previsione di una serie di obblighi informativi a carico del finanziatore[52]. In particolare, gli aspiranti prenditori dovranno essere informati ogni volta che la concessione, la determinazione del *pricing* o il diniego del prestito sia l'esito di una procedura di trattamento automatizzato dei dati *«in modo che gli stessi possano prendere in considerazione i potenziali rischi nella loro decisione di acquisto»*[53].
Inoltre, al consumatore è attribuito un diritto di reazione[54], consistente nella possibilità di chiedere una spiegazione "chiara e comprensibile" del funzionamento del trattamento automatizzato comprese le sue principali variabili, la sua logica, e i suoi rischi, di ottenere una revisione della procedura, verosimilmente attraverso un operatore umano (pur non essendo specificato nella norma).
Infine, la CCD II, per la prima volta, include una disposizione esplicitamente finalizzata a prevenire condotte discriminatorie nell'accesso al credito in ragione *«della cittadinanza o del luogo di residenza o per qualsiasi motivo di cui all'articolo 21 della Carta dei diritti fondamentali dell'Unione europea»*[55].

5. La normativa non settoriale. Il GDPR e l'Artificial Intelligence Act

Al di fuori dell'ordinamento settoriale del credito, rivestono particolare importanza la normativa euro-unitaria in materia di privacy (Regolamento (UE) 2016/679, di seguito GDPR) e il recente Regolamento sull'Intelligenza Artificiale (di seguito AI Act).
Nello specifico, il GDPR, all'art. 15, par. 1, lett. h), riconosce all'interessato il diritto di essere informato dell'esistenza di un processo decisionale automatizzato che lo riguarda[56].
Ancora, all'art. 22, prevede il diritto a non essere sottoposto ad una decisione basata unicamente sul trattamento automatizzato dei dati[57]; nelle ipotesi in cui ciò sia consentito[58] al titolare deve essere

personali trattati dai social network che potrebbero essere contenuti nelle banche dati di cui al paragrafo 1 del presente articolo».

[52] Tra le disposizioni, assume particolare rilevanza l'art. 10 («*Informazioni precontrattuali*») il quale prevede che il consumatore sia informato (in maniera chiara e comprensibile) se «*il prezzo è stato personalizzato sulla base di un trattamento automatizzato, inclusa la profilazione*» (par. 5 lett. m). Una disposizione identica è contenuta all'art. 11 par. 4 lett. h con riferimento ai contratti di credito di cui all'articolo 2, paragrafo 6 o 7.
E ancora all'articolo 18, comma 9 si aggiunge che «*Se del caso, il creditore è tenuto a informare il consumatore del fatto che la valutazione del merito creditizio è basata sul trattamento automatizzato di dati come anche del diritto del consumatore a una valutazione umana e della procedura per contestare la decisione».*

[53] Cfr. il considerando n. 46 della CCD II.

[54] È quanto previsto dall'art.18 comma 8, secondo il quale il consumatore ha il diritto*:* «*a) di chiedere ed ottenere dal creditore una spiegazione chiara e comprensibile della valutazione del merito creditizio, compresi la logica e i rischi derivanti dal trattamento automatizzato dei dati personali nonché la rilevanza e gli effetti sulla decisione; b) di esprimere la propria opinione al creditore; e c) di chiedere un riesame della valutazione del merito creditizio e della decisione relativa alla concessione del credito da parte del creditore».*

[55] Cfr. art. 6 CCD II.

[56] A ben vedere, la disposizione è accostabile all'art. 120 *undecies*, comma 5, ma ha un contenuto più ampio, poiché il diritto di accesso è riconosciuto indipendentemente dalla circostanza che la richiesta di credito sia stata rifiutata.

[57] Tra queste ipotesi il considerando n. 71 del GDPR include proprio «*il rifiuto automatico di una domanda di credito online».*

[58] Perché, ai sensi del paragrafo 2 dello stesso articolo, «*a) sia necessaria per la conclusione o l'esecuzione di un contratto tra l'interessato e un titolare del trattamento […] c) si basi sul consenso esplicito dell'interessato».*

Downloaded from www.biodiritto.org.
ISSN 2284-4503

riconosciuto almeno il diritto «*di ottenere l'intervento umano […] di esprimere la propria opinione e di contestare la decisione*». Un'interpretazione estensiva della norma, preferibile anche alla luce dell'art.18 comma 8, della CCD II, permetterebbe all'interessato di esigere dal titolare del trattamento la "spiegazione" della logica sottesa alla decisione algoritmica[59], che, tuttavia, non pare potersi estendere ad una piena *disclosure* del codice coperto da segreto industriale.

L'AI Act, invece, classifica «*i sistemi di IA utilizzati per valutare il merito di credito o l'affidabilità creditizia delle persone fisiche come sistemi di IA ad alto rischio, in quanto determinano l'accesso di tali persone alle risorse finanziarie o a servizi essenziali quali l'alloggio, l'elettricità e i servizi di telecomunicazione*».

I modelli ad alto rischio possono «*portare alla discriminazione di persone o gruppi e perpetuare modelli storici di discriminazione, ad esempio in base all'origine razziale o etnica, alle disabilità, all'età o all'orientamento sessuale, o dar vita a nuove forme di effetti discriminatori* [60]».

Per queste ragioni, i fornitori devono rispettare i requisiti previsti dal Titolo III, Capo II del Regolamento, volti a garantire un'elevata qualità dei dati, metodologie e pratiche idonee a prevenire distorsioni, la tracciabilità dei risultati nonché il controllo di esperti.

6. Conclusioni: quale rimedio contro la discriminazione algoritmica?

Nel prossimo futuro, è altamente probabile che un numero crescente di intermediari del credito italiani impiegherà sistemi di IA nei procedimenti di *credit scoring* e ciò determinerà un inevitabile aumento dei casi di discriminazione algoritmica.

L'ordinamento non prevede disposizioni che regolano appositamente il fenomeno. Al più, l'analisi sopra condotta evidenzia un quadro frammentato di prescrizioni nazionali e sovranazionali più propriamente volte a disciplinare l'IA e il settore del credito. Soltanto il già menzionato art. 6 della CCD II introduce un divieto di discriminazione nell'accesso al credito, rimettendone, tuttavia, agli Stati membri le concrete modalità di attuazione e tutela.

In tale ottica, al consumatore sono riconosciuti una serie di strumenti che gli consentono di venire a conoscenza delle logiche sottese alla decisione automatizzata. L'Arbitro Bancario e Finanziario, pronunciandosi su un caso di diniego del credito sulla base di sistemi di *credit scoring* automatici[61], ha ribadito il diritto del cliente a conoscerne le ragioni in osservanza dei doveri di correttezza, buona fede e "collaborazione attiva" che ormai segnano la disciplina della trasparenza bancaria, configurandosi in assenza di siffatta motivazione un diritto al risarcimento[62].

Cionondimeno, siffatti diritti conoscitivi presentano due ordini di problemi.

[59] L. AMMANNATI, G.L. GRECO, *Il credit scoring "intelligente": esperienze, rischi e nuove regole,* in *Rivista di Diritto Bancario*, 3, 2023, 479.

[60] V. Considerando n. 58.

[61] ABF, Collegio di coordinamento, 29 novembre 2013, n. 6182, in www.arbitrobancario-finanziario.it (ultima consultazione 30/06/2024); ABF, Collegio di Napoli, 7 aprile 2016, n. 3196, *ivi*; ABF, Collegio di Napoli, 21 settembre 2016, n. 8100, *ivi*; ABF, Collegio di Roma, 16 novembre 2021, n. 23570, *ivi*;

[62] ABF, Collegio di Bari, 08 ottobre 2021, n. 21103, in www.arbitrobancario-finanziario.it (ultima consultazione 30/06/2024).

Downloaded from www.biodiritto.org.
ISSN 2284-4503

In primo luogo, innanzi all'esercizio degli stessi i finanziatori possono opporre tanto l'inspiegabilità del sistema quanto il segreto industriale.

In secondo luogo, ci si chiede se le informazioni così ottenute possano essere prodromiche all'esercizio del diritto di difesa in sede giurisdizionale ovverosia si pone la questione circa i rimedi esperibili dal consumatore nel caso di diniego di credito a cagione di una discriminazione algoritmica.

A ben vedere, nel caso di specie i nuovi problemi delle applicazioni dell'IA celano la necessità di riaffrontare temi giuridici tradizionali come quello della sindacabilità della decisione della banca di concedere o denegare credito. Infatti, anche il *credit scoring* algoritmico, in quanto funzionale alla valutazione del merito di credito, rimane secondo l'opinione tradizionale nella sfera di assoluta discrezionalità della banca, non potendosi configurare un diritto al credito[63].

Tuttavia, tali considerazioni non possono condurre ad escludere il settore del credito dall'ambito di applicazione del diritto antidiscriminatorio e, dunque, dalla possibilità di esperire le azioni civili contro la discriminazione di cui agli art. 44 d.lgs. n. 286/1998, art. 4 d.lgs. n. 215/2003, art. 3 l. n. 67/2006, art. 55-*quinquies* d.lgs. n. 198/2006, regolate dal rito semplificato di cognizione ai sensi dell'art. 28 d.lgs. n. 150/2011[64].

In particolare, il co. 4 dell'articolo da ultimo citato consente di superare il problema dell'effetto *black box* e del segreto industriale: all'allegazione del ricorrente di elementi di fatto, desunti da dati di carattere anche statistico, dai quali si può presumere l'esistenza di *pattern* discriminatori il legislatore riconnette un'inversione dell'onere della prova tale per cui spetterà al soggetto finanziatore dimostrare che l'IA non ha generato un risultato secondo una logica discriminatoria.

Sul piano dei contenuti della sentenza, il co. 5 attribuisce il diritto di rivolgersi al giudice per ottenere: il risarcimento del danno anche non patrimoniale, la cessazione dell'atto discriminatorio pregiudizievole, ogni altro provvedimento idoneo a rimuoverne gli effetti e l'ordine di adozione di un piano di rimozione delle discriminazioni.

Più nel dettaglio, bisogna interrogarsi se la cessazione dell'atto discriminatorio e l'adozione di ogni altro provvedimento idoneo possano tradursi in un ordine di rivalutazione del richiedente previa eliminazione dei bias riscontrati nell'IA, nonché se l'ordine di adozione di un piano di rimozione delle discriminazioni possa consistere in una ridefinizione delle politiche e delle procedure interne volte a prevenire i rischi legati all'utilizzo dei modelli automatizzati[65].

---

[63] P. ABBADESSA, *Obbligo di far credito*, in *Enc. dir.*, XXIX, 1979, 529 ss.; N. SALANITRO, *Le banche e i contratti bancari*, in F. VASSALLI (diretto da) *Trattato di diritto civile*, VIII, 3 1983, 42; in giurisprudenza si rinvia a Trib. Brindisi, 07 agosto 2021, in www.ilcaso.it (ultima consultazione 30/06/2024), secondo il quale resta fuori dalla cognizione del Giudice la verifica della sussistenza dei presupposti, sotto il profilo contabile, per la concessione del finanziamento da parte della banca, al pari di ogni altra transazione economica fra privati.

[64] In generale, sull'applicabilità del diritto antidiscriminatorio alla valutazione del merito creditizio K. LANGENBUCHER, *Consumer Credit in The Age of AI–Beyond Anti-Discrimination Law*, in *ECGI Working Paper Series in Law*, 663, 2022.

[65] Dello stesso parere G. MATTARELLA, *Big data e accesso al credito degli immigrati: discriminazioni algoritmiche e tutela del consumatore*, in *Giurisprudenza commerciale,* 4, 2020, 711, il quale addirittura prospetta la possibilità di ottenere giudizialmente la stipula del contratto attraverso un risarcimento del danno in forma specifica oppure nel caso di condizioni contrattuali peggiori rispetto a quelle normalmente praticate di ottenere una rettifica del contratto.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Una risposta in senso affermativo non solo sembra consentita da un'interpretazione letterale delle disposizioni, ma soprattutto da un'interpretazione sistematica che tenga conto tanto dell'art. 6 della CCD II quanto dei principi costituzionali.

Infatti, l'art. 47 della Costituzione nel prevedere che la Repubblica «*disciplina, coordina e controlla l'esercizio del credito*» implica necessariamente che a tale disciplina compartecipino gli stessi principi costituzionali, sovranazionali e internazionali e con essi il principio di uguaglianza (artt. 3 Cost. e 21 Carta di Nizza e 14 CEDU)[66].

[66] Per una lettura del settore del credito alla luce dei principi costituzionali v. A. LANZAFAME, *Credito e costituzione: dal risparmio come «bene comune» al principio di accessibilità. Temi e problemi di democrazia economica*, in www.costituzionalismo.it (ultima consultazione 20/06/2024), 1, 2019.

Downloaded from www.biodiritto.org.

ISSN 2284-4503

Non discriminazione e diritto alla diversità: cosa c'è di nuovo per i disabili nell'era dell'AI?

Valentina Pagnanelli*

NON-DISCRIMINATION AND RIGHT TO DIVERSITY: WHAT'S NEW FOR DISABLED PEOPLE IN THE AI ERA?
ABSTRACT: In the last decade, the European Union has faced profound changes due to the Covid-19 pandemic, recent military conflicts, and technological advancements such as Big Data, cloud computing, IoT, and AI systems. These developments have led to a new legal framework for citizens and businesses, encompassing old and new rights. This paper examines the condition of disability. This particular condition of vulnerability indeed serves as a litmus test for the effectiveness of the right to non-discrimination and the mechanisms protecting it. It outlines the legal definition of disability, the key regulations, and the European legal context. A discussion on equality introduces the right to diversity. The paper concludes with reflections on AI's potential to enhance disability rights, emphasizing the importance of transparent, GDPR-compliant practices and ongoing collaboration among institutions and technology providers.

KEYWORDS: Artificial Intelligence; Disability; Nondiscrimination; Diversity; Data protection.

ABSTRACT: Negli ultimi dieci anni, l'Unione Europea ha attraversato trasformazioni profonde, segnate dalla pandemia di Covid-19, da conflitti internazionali e dai rapidi progressi tecnologici in ambiti quali Big Data, cloud computing, Internet of Things (IoT) e intelligenza artificiale (IA). Questi cambiamenti hanno determinato l'elaborazione di un nuovo quadro normativo che ridefinisce diritti "tradizionali" e introduce nuovi diritti per cittadini e imprese. Il presente studio si propone di analizzare la condizione delle persone con disabilità, considerata un importante banco di prova per valutare l'efficacia del diritto alla non discriminazione e dei relativi strumenti di tutela. A tal fine, vengono esaminati la definizione giuridica della disabilità, le principali normative di riferimento e il contesto giuridico europeo. La riflessione sull'uguaglianza conduce inevitabilmente a una valorizzazione del diritto alla diversità. Il contributo si conclude con una disamina del potenziale dell'intelligenza artificiale nel promuovere i diritti delle persone con disabilità, evidenziando l'importanza di adottare pratiche trasparenti e conformi al GDPR, nonché di favorire una collaborazione costante tra istituzioni e fornitori tecnologici.

* *Assegnista di ricerca in diritto costituzionale, Università di Firenze. Mail valentina.pagnanelli@unifi.it.* *Contributo sottoposto a doppio referaggio anonimo.*

Downloaded from www.biodiritto.org.
ISSN 2284-4503

PAROLE CHIAVE: Intelligenza artificiale; Disabilità; Non discriminazione; Diversità; Protezione dei dati personali.

SOMMARIO: 1. Introduzione – 2. Disabilità e non discriminazione nella Convenzione ONU e nelle politiche europee – 3. Disabilità e non discriminazione nel GDPR e nell'AI Act. 4. Principio di uguaglianza, minoranze e diritto alla diversità. 5. Dalla non discriminazione al diritto alla diversità. Una nuova missione per l'AI?

1. Introduzione

Nell' ultimo decennio il contesto giuridico, economico e sociale dell'Unione europea ha subito profondi cambiamenti[1]. Molti di essi sono conseguenze della pandemia di Covid-19[2] e dei recenti eventi bellici. Molti altri, sostanziali, dipendono dall'irrompere sulla scena globale di straordinarie innovazioni, legate a nuove tecnologie quali i *Big Data*[3], il *cloud e edge computing*, *Internet of Things*, i sistemi di Intelligenza Artificiale. Tutti gli accadimenti e tutte le innovazioni appena richiamati hanno visto una produzione normativa che ha in alcuni casi anticipato ed in altri seguito gli eventi, ma che di fatto delinea ora un preciso panorama giuridico entro cui cittadini e imprese europee agiscono, esprimono la propria personalità, sono chiamati a rispettare obblighi e godono di "antichi" e nuovi diritti[4]. Questo contributo intende esaminare una specifica condizione in cui gli individui in diverse fasi della vita potrebbero trovarsi, la disabilità. Questa particolare condizione di vulnerabilità costituisce infatti un banco di prova della effettività del diritto alla non discriminazione e degli strumenti a sua tutela.

Nei prossimi paragrafi a partire dalla definizione giuridica della condizione di disabilità si richiameranno i principali riferimenti normativi a tutela del diritto di non discriminazione delle *disabled persons*. Seguiranno alcuni cenni al contesto giuridico europeo e alle norme di maggiore impatto sul diritto alla non discriminazione dei disabili. Una digressione sul principio di uguaglianza condurrà alla definizione di un diritto alla diversità. Si tenterà infine di trarre alcune riflessioni dall'indagine svolta, individuando possibili ambiti di sviluppo dell'applicazione dell'Intelligenza Artificiale per una maggiore tutela dei

[1] «Digital technologies are profoundly changing our daily life, our way of working and doing business, and the way people travel, communicate and relate with each other. Digital communication, social media interaction, e-commerce, and digital enterprises are steadly transforming our world. They are generating an even-increasing amount of data, which, if pooled and used, can lead to a completely new means and levels of value creation. It is a transformation as fundamental as that caused by the industrial revolution*», cfr.* European Commission*, Communication: Shaping Europe's digital future,* 19 febbraio 2020.

[2] A. PAINO, L. VIOLANTE (a cura di), *Biopolitica, pandemia e democrazia. Rule of law nella società digitale*, Bologna, 2021; S. COCCHI, A. SIMONI (a cura di), *Freedom v. Risk. Social Control and the Idea of Law in the Covid-19 Emergency*, Torino, 2022.

[3] K. CUKIER, V. MAYER-SCHOENBERGER, *The Rise of Big Data: How It's Changing the Way We Think About the World*, in *Foreign Aff.*, 2013, 28, 28-40; *Big data: The next frontier for innovation, competition, and productivity,* http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation (ultima consultazione 02/12/2024). Autorità per le garanzie nelle comunicazioni, *Big Data. Interim report nell'ambito dell'indagine conoscitiva di cui alla delibera n. 217/17/CONS;* M. PALMIRANI*, Big Data e conoscenza, in Riv. fil. dir.,* 2020, 1, 73-92.

[4] Commissione Europea, *Libro bianco sull'intelligenza artificiale. Un approccio europeo all'eccellenza e alla fiducia,* COM(2020) 65, 19 febbraio 2020.

Downloaded from www.biodiritto.org.

ISSN 2284-4503

diritti dei disabili[5].

2. Disabilità e non discriminazione nella Convenzione ONU e nelle politiche europee

La Convenzione delle Nazioni Unite sui diritti delle persone con disabilità, adottata il 13 dicembre 2006, di cui l'Unione europea è parte dal 21 gennaio 2011 annovera tra le persone con disabilità «quanti hanno minorazioni fisiche, mentali, intellettuali o sensoriali a lungo termine che in interazione con varie barriere possono impedire la loro piena ed effettiva partecipazione nella società su una base di eguaglianza con gli altri»[6]. Si tratta di una definizione ormai ampiamente condivisa[7], che pone l'accento sulla presenza di barriere che, nella interazione con la particolare condizione della persona disabile, creano un impedimento al pieno godimento dei diritti. Il riconoscimento del diritto a godere dei diritti in condizioni di eguaglianza rispetto agli altri cittadini non disabili[8] rappresenta il fulcro dell'intero articolato, tanto da aver suscitato dubbi, in sede di adozione della Convenzione, sulla effettiva necessità di un nuovo decalogo che avrebbe avuto lo scopo di ribadire garanzie già riconosciute a tutti gli individui senza distinzioni.

Effettivamente la Convenzione ONU sui diritti delle persone con disabilità può essere vista come una versione aggiornata di precedenti dichiarazioni sui diritti[9], arricchita con misure specifiche mirate a garantire il massimo livello di tutela anche alle persone con disabilità[10].

La Carta dei diritti fondamentali dell'Unione europea riconosce all'Articolo 21 il diritto alla non discriminazione fondata «sul sesso, la razza, il colore della pelle o l'origine etnica o sociale, le caratteristiche genetiche, la lingua, la religione o le convinzioni personali, le opinioni politiche o di qualsiasi altra natura, l'appartenenza ad una minoranza nazionale, il patrimonio, la nascita, gli handicap, l'età o le tendenze sessuali», mentre l'Articolo 26 è dedicato alle misure concrete che dovrebbero garantire l'autonomia, l'inserimento sociale e professionale dei disabili, e la loro partecipazione alla vita della comunità[11].

---

[5] M. Whittaker, M. Alper, C.L. Bennett, S. Hendren, L. Kaziunas, M. Mills, M. Ringel Morris, J. Rankin, E. Rogers, M. Salas, S. Myers West, *Disability, Bias, and AI*, in *AI NOW*, November 2019; M. Wald, *AI Data-driven personalisation and Disability inclusion*, in *Frontiers in Artificial Intelligence*, 18 January 2021; P. Smith, L. Smith, *Artificial Intelligence and Disability: too much promise, yet too little substance?*, in *AI and Ethics*, 2021; M. Marks, *Algorithmic Disability discrimination*, in I. Glenn Cohen et al. (a cura di), *Disability, Health, Law and Bioethics,* Cambridge, 2020.

[6] Convenzione delle Nazioni Unite sui diritti delle persone con disabilità del 13 dicembre 2006, Art. 1 par. 2.

[7] La Direttiva n. 2019/882 sui requisiti di accessibilità dei prodotti e dei servizi, che sarà pienamente applicabile nel 2025, riprende integralmente la definizione contenuta nella Convenzione ONU.

[8] Convenzione delle Nazioni Unite sui diritti delle persone con disabilità del 13 dicembre 2006, Preambolo, lettera e): «[…]la disabilità è il risultato dell'interazione tra persone con menomazioni e barriere comportamentali ed ambientali, che impediscono la loro piena ed effettiva partecipazione alla società su base di uguaglianza con gli altri».

[9] Tra cui la Dichiarazione Universale dei Diritti Umani, la Convenzione europea per la salvaguardia dei diritti dell'uomo e delle libertà fondamentali, il Patto internazionale sui diritti civili e politici.

[10] V. Della Fina, R. Cera, G. Palmisano (a cura di*), The United Nations Convention on the Rights of Persons with Disabilities. A Commentary*, Cham, 2017; D. Piccione, *Costituzionalismo e disabilità. I diritti delle persone con disabilità tra Costituzione e Convenzione ONU*, Torino, 2023.

[11] Per un commento alla Carta dei diritti dell'Unione europea, ved. R. Bifulco, M. Cartabia, A. Celotto (a cura di), *L'Europa dei diritti: commento alla Carta dei diritti fondamentali dell'Unione europea*, Bologna, 2001.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Per rendere effettive queste dichiarazioni di principio, l'Unione europea ha elaborato nel corso degli anni delle strategie con l'obiettivo di garantire alle persone con disabilità in Europa il godimento dei loro diritti, con pari opportunità e piena partecipazione alla vita della società e all'economia, indipendentemente dal sesso, dall'origine razziale o etnica, dalla religione o credo, dall'età o dall'orientamento sessuale[12]. Nel marzo 2021, la Commissione ha adottato la *Strategia per i diritti delle persone con disabilità 2021-2030*[13], nella quale si affrontano anche i rischi di svantaggio multiplo cui sono soggette donne, bambini, anziani, rifugiati con disabilità e persone con difficoltà socioeconomiche. La Strategia promuove così una prospettiva intersezionale in linea con l'Agenda 2030 delle Nazioni Unite per lo Sviluppo Sostenibile e gli Obiettivi di Sviluppo Sostenibile[14]. In particolare, la partecipazione paritaria mira a proteggere le persone con disabilità da ogni forma di discriminazione e violenza, garantendo pari opportunità e accesso alla giustizia, all'istruzione, alla cultura, allo sport, al turismo e a tutti i servizi sanitari.

Vi è da dire che la produzione normativa europea sulla disabilità pare essere focalizzata principalmente sull'accessibilità dei disabili a prodotti e servizi[15]. Recentemente l'Unione Europea ha però introdotto una serie di regolamentazioni che, sebbene non riguardino direttamente l'accessibilità, hanno un impatto notevole sulla tutela del diritto alla non discriminazione. Queste norme, che saranno esaminate nel dettaglio nel prossimo paragrafo, sono fondamentali per garantire che le persone con disabilità possano godere pienamente dei loro diritti in una società inclusiva e paritaria.

[12] Nel 2010 la Commissione ha adottato la *Strategia europea sulla disabilità 2010-2020: un rinnovato impegno per un'Europa senza barriere*, COM(2010) 636 final.

[13] Commissione Europea, Comunicazione della Commissione al Parlamento europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle Regioni, Un'Unione dell'uguaglianza: strategia per i diritti delle persone con disabilità 2021-2030, COM(2021) 101 final, 3 marzo 2021.

[14] Si veda: https://www.agenziacoesione.gov.it/dossier_tematici/agenda-onu-2030-per-lo-sviluppo-sostenibile/ (ultima consultazione 25/07/2024). .

[15] Si veda la Direttiva UE 2019/882 del Parlamento europeo e del Consiglio del 17 aprile 2019 sui requisiti di accessibilità dei prodotti e dei servizi.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

3. Disabilità e non discriminazione nel GDPR e nell'AI Act

Nel contesto dei numerosi atti normativi adottati a livello europeo nell'ambito della strategia digitale[16], due testi fondamentali si distinguono per la governance dei dati e la regolazione dei sistemi di Intelligenza Artificiale: il GDPR (Regolamento 2016/679[17]) e l'AI Act (Regolamento 2024/1689[18]).

Il GDPR ha delineato un modello di governance dei dati che, per la sistematicità e completezza delle regole poste, nonché per l'importanza dei beni giuridici tutelati, si pone sia al vertice che alla base di tutte le altre normative sui dati. Questo regolamento ha stabilito un quadro giuridico solido e coerente, garantendo la protezione dei dati personali e il rispetto della *privacy* degli individui, elementi essenziali per la fiducia dei cittadini nell'economia digitale.

L'AI Act, d'altra parte, rappresenta il primo tentativo su scala globale di regolamentare un fenomeno tecnologico potente e dagli sviluppi spesso imprevedibili come l'Intelligenza Artificiale. Questo regolamento mira a stabilire un quadro normativo che assicuri l'uso sicuro e trasparente dell'IA, promuovendo al contempo l'innovazione e proteggendo i diritti fondamentali delle persone. In questo modo, l'Unione Europea si propone di essere all'avanguardia nella regolamentazione delle tecnologie emergenti, affrontando le sfide e le opportunità poste dall'Intelligenza Artificiale in modo proattivo e responsabile.

Entrambi questi articolati, seppure in ambiti di applicazione e con finalità differenti, contengono disposizioni volte a proteggere i soggetti o i gruppi vulnerabili da possibili rischi di discriminazione derivanti da trattamenti malevoli dei dati personali e da usi scorretti dei sistemi di Intelligenza Artificiale.

Nell'affrontare un'analisi incentrata esclusivamente sulla condizione delle persone con disabilità, è fondamentale evidenziare la principale differenza di approccio tra i due regolamenti. Infatti, nel Regolamento generale sulla protezione dei dati personali (GDPR), il termine "disabilità" appare soltanto due

[16] Cfr. Commissione europea, *Comunicazione della Commissione al Parlamento europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle Regioni "Costruire un'economia dei dati europea"*, COM(2017) 9 final, 10 gennaio 2017; Commissione europea, *Comunicazione della Commissione al Parlamento europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle Regioni "Una strategia europea per i dati"*, COM(2020) 66 final, 19 febbraio 2020.

[17] Regolamento (UE) 2016/679 del Parlamento e del Consiglio del 27 aprile 2016 relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (Regolamento generale sulla protezione dei dati). Per una introduzione: G. Finocchiaro, *Il quadro d'insieme sul regolamento europeo sulla protezione dei dati personali*, in G. Finocchiaro (a cura di), *La protezione dei dati personali in Italia. Regolamento UE n. 2016/679 e d.lgs. 10 agosto 2018, n. 101*, Torino, 2019, 1 ss.; E. Lucchini Guastalla, *Il nuovo regolamento europeo sul trattamento dei dati personali: i principi ispiratori*, in *Contr. Impr.*, 2018, 106 ss.

[18] C. Camardi (a cura di), *La via europea per l'Intelligenza artificiale*. *Atti del Convegno del Progetto Dottorale di Alta Formazione in Scienze Giuridiche – Ca' Foscari Venezia, 25-26 novembre 2021*, Milano, 2022; G. Cerrina Feroni, C. Fontana, E.C. Raffiotta (a cura di), *AI Anthology. Profili giuridici, economici e sociali dell'intelligenza artificiale*, Bologna, 2022; C. Casonato, B. Marchetti, *Prime osservazioni sulla Proposta di Regolamento dell'Unione Europea in materia di Intelligenza Artificiale*, in *Biolaw Journal – Rivista di BioDiritto*, 3, 2021, 415 ss.; A. Mantelero, *Sulle regole AI l'Europa sceglie approccio "industriale": luci e ombre*, in *AgendaDigitale*, 27 aprile 2021; A. Simoncini, *Verso la regolamentazione della Intelligenza Artificiale. Dimensioni e governo*, in *BioLaw Journal – Rivista di BioDiritto*, 2, 2021.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

volte nei Considerando[19]. Al contrario, nel Regolamento che disciplina l'Intelligenza Artificiale, la disabilità è menzionata diciannove volte[20].

La condizione di disabilità è considerata uno *status* particolare che può esporre le persone a potenziali discriminazioni o danni. Pertanto, l'AI Act riconosce e affronta in modo dettagliato le esigenze e le vulnerabilità delle persone con disabilità, cercando di garantire che i sistemi di IA siano sviluppati e utilizzati in modo inclusivo e non discriminatorio.

Nel GDPR, la condizione delle persone con disabilità è inclusa nella più ampia categoria dei dati relativi allo stato di salute[21]. La disciplina per il trattamento di questi dati è contenuta nell'articolo 9[22], che

[19] GDPR, Considerando 35 e 54.

[20] AI Act, Considerando 29, 32, 48, 54, 56, 57, 58, 80, 132, 142, 165, Articolo 5, Articolo 60, Articolo 95.

[21] Cfr. GDPR, Considerando 35.

[22] Articolo 9 - Trattamento di categorie particolari di dati personali**:**

«*1. È vietato trattare dati personali che rivelino l'origine razziale o etnica, le opinioni politiche, le convinzioni religiose o filosofiche, o l'appartenenza sindacale, nonché trattare dati genetici, dati biometrici intesi a identificare in modo univoco una persona fisica, dati relativi alla salute o alla vita sessuale o all'orientamento sessuale della persona.*

2. Il paragrafo 1 non si applica se si verifica uno dei seguenti casi:

a) l'interessato ha prestato il proprio consenso esplicito al trattamento di tali dati personali per una o più finalità specifiche, salvo nei casi in cui il diritto dell'Unione o degli Stati membri dispone che l'interessato non possa revocare il divieto di cui al paragrafo 1;

b) il trattamento è necessario per assolvere gli obblighi ed esercitare i diritti specifici del titolare del trattamento o dell'interessato in materia di diritto del lavoro e della sicurezza sociale e protezione sociale, nella misura in cui sia autorizzato dal diritto dell'Unione o degli Stati membri o da un contratto collettivo ai sensi del diritto degli Stati membri, in presenza di garanzie appropriate per i diritti fondamentali e gli interessi dell'interessato;

c) il trattamento è necessario per tutelare un interesse vitale dell'interessato o di un'altra persona fisica qualora l'interessato si trovi nell'incapacità fisica o giuridica di prestare il proprio consenso;

d) il trattamento è effettuato, nell'ambito delle sue legittime attività e con adeguate garanzie, da una fondazione, associazione o altro organismo senza scopo di lucro che persegua finalità politiche, filosofiche, religiose o sindacali, a condizione che il trattamento riguardi unicamente i membri, gli ex membri o le persone che hanno regolari contatti con la fondazione, l'associazione o l'organismo a motivo delle sue finalità e che i dati personali non siano comunicati all'esterno senza il consenso dell'interessato;

e) il trattamento riguarda dati personali resi manifestamente pubblici dall'interessato;

f) il trattamento è necessario per accertare, esercitare o difendere un diritto in sede giudiziaria o ogniqualvolta le autorità giurisdizionali esercitino le loro funzioni giurisdizionali;

g) il trattamento è necessario per motivi di interesse pubblico rilevante sulla base del diritto dell'Unione o degli Stati membri, che deve essere proporzionato alla finalità perseguita, rispettare l'essenza del diritto alla protezione dei dati e prevedere misure appropriate e specifiche per tutelare i diritti fondamentali e gli interessi dell'interessato;

h) il trattamento è necessario per finalità di medicina preventiva o di medicina del lavoro, valutazione della capacità lavorativa del dipendente, diagnosi, assistenza o terapia sanitaria o sociale ovvero gestione dei sistemi e servizi sanitari o sociali sulla base del diritto dell'Unione o degli Stati membri o conformemente al contratto con un professionista della sanità, fatte salve le condizioni e le garanzie di cui al paragrafo 3;

i) il trattamento è necessario per motivi di interesse pubblico nel settore della sanità pubblica, quali la protezione da gravi minacce per la salute a carattere transfrontaliero o la garanzia di parametri elevati di qualità e sicurezza dell'assistenza sanitaria e dei medicinali e dei dispositivi medici, sulla base del diritto dell'Unione o degli Stati membri che prevede misure appropriate e specifiche per tutelare i diritti e le libertà dell'interessato, in particolare il segreto professionale;

j) il trattamento è necessario a fini di archiviazione nel pubblico interesse, di ricerca scientifica o storica o a fini statistici in conformità dell'articolo 89, paragrafo 1, sulla base del diritto dell'Unione o nazionale, che è

Downloaded from www.biodiritto.org.
ISSN 2284-4503

stabilisce un generale divieto di trattamento delle categorie particolari di dati. Queste categorie comprendono informazioni la cui conoscenza e utilizzo potrebbero causare discriminazioni per gli interessati, come l'origine razziale, le scelte politiche e sindacali, le convinzioni religiose e filosofiche, l'orientamento sessuale e i dati relativi alla salute.

Il paragrafo 2 dello stesso articolo prevede una serie di eccezioni che consentono, in presenza di una base giuridica adeguata, di procedere al trattamento di questi dati. Da queste eccezioni emerge chiaramente come il legislatore europeo abbia voluto rafforzare le tutele per garantire il diritto delle persone fisiche a mantenere il controllo sui propri dati sensibili. Le eccezioni al divieto di trattamento sono ancorate a interessi di elevato valore, quali la tutela della vita umana, la disciplina del rapporto di lavoro e motivi di interesse pubblico rilevante.

Nonostante ciò, nel GDPR non compaiono misure specifiche che riguardano lo *status* delle persone con disabilità, tali da giustificare una separazione e un diverso regime di tutela per queste informazioni. Questa mancanza di specificità implica che le persone con disabilità non beneficiano di protezioni aggiuntive o distinte rispetto ad altre categorie di dati sensibili.

Per quanto riguarda l'oggetto della nostra ricerca, è di massimo interesse la disciplina prevista dall'articolo 22[23] del GDPR, che regola il trattamento dei dati personali effettuato attraverso algoritmi[24]. Questi algoritmi hanno il potenziale di incidere significativamente sui diritti e le libertà delle persone fisiche. L'articolo 22 stabilisce un divieto generale secondo il quale «*l'interessato ha il diritto di non essere sottoposto a una decisione basata unicamente sul trattamento automatizzato [...] che produca effetti giuridici che lo riguardano o che incida in modo analogo significativamente sulla persona*[25]».

Questa norma vieta che tali decisioni possano essere prese senza intervento umano, predisponendo delle garanzie per l'interessato. Tra queste garanzie vi è il diritto dell'interessato di esprimere la propria opinione e di contestare la decisione. In questo modo, il GDPR cerca di bilanciare l'uso delle tecnologie automatizzate con la necessità di proteggere i diritti fondamentali delle persone, assicurando che le

proporzionato alla finalità perseguita, rispetta l'essenza del diritto alla protezione dei dati e prevede misure appropriate e specifiche per tutelare i diritti fondamentali e gli interessi dell'interessato.

3. I dati personali di cui al paragrafo 1 possono essere trattati per le finalità di cui al paragrafo 2, lettera h), se tali dati sono trattati da o sotto la responsabilità di un professionista soggetto al segreto professionale conformemente al diritto dell'Unione o degli Stati membri o alle norme stabilite dagli organismi nazionali competenti o da altra persona anch'essa soggetta all'obbligo di segretezza conformemente al diritto dell'Unione o degli Stati membri o alle norme stabilite dagli organismi nazionali competenti.

4. Gli Stati membri possono mantenere o introdurre ulteriori condizioni, comprese limitazioni, con riguardo al trattamento di dati genetici, dati biometrici o dati relativi alla salute». Per un commento ved. L. BOLOGNINI, E. PELINO, (a cura di), *Codice della disciplina privacy*, Milano, 2019; G.M. RICCIO, G. SCORZA, E. BELISARIO (a cura di), *GDPR e normativa privacy. Commentario*, Milano, 2018.

[23] Tra i commenti all'art. 22 del Regolamento europeo n. 2016/679 si vedano, *ex multis*, G.M. RICCIO, G. SCORZA, E. BELISARIO (a cura di), *GDPR e normativa privacy. Commentario*, Milano, 2018, 219 ss.; L. BOLOGNINI L., E. PELINO, *Codice della disciplina privacy*, Milano, 2019, 181 ss.; G. FINOCCHIARO (opera diretta da), *La protezione dei dati personali in Italia. Regolamento UE n. 2016/679 e d.lgs. 10 agosto 2018, n. 101*, Bologna, 2019, 458 ss. Sia consentito rinviare anche a V. PAGNANELLI, *Decisioni algoritmiche e tutela dei dati personali. Riflessioni attorno al ruolo del Garante*, in *Osservatorio sulle fonti*, 2, 2021, 783 ss.

[24] *Processo decisionale automatizzato relativo alle persone fisiche, compresa la profilazione*, cfr. rubrica Art. 22 del GDPR.

[25] Regolamento UE n. 2016/679, Art. 22, par. 1.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

decisioni di impatto significativo non siano prese esclusivamente da algoritmi senza la supervisione e i trattamenti automatizzati vietati la disposizione cita espressamente la profilazione. Ancora una volta in assenza di un riferimento specifico alla condizione di disabilità, la disposizione prevede che i dati atti a rivelare caratteristiche fisiche e condizioni di salute possano essere utilizzati solo dietro prestazione del consenso esplicito dell'interessato o in ragione di un interesse pubblico rilevante. Torneremo però più avanti sulla applicazione di questa norma nei casi in cui i dati riguardino persone con disabilità, per esplorarne non solo i rischi ma anche le potenzialità, alla luce del diritto alla diversità.

Nel passare all'esame delle disposizioni in tema di tutela dei soggetti vulnerabili e non discriminazione che sono contenute nell'AI Act occorre per prima cosa osservare che le misure riguardano due profili di sviluppo ed utilizzo dei sistemi di Intelligenza Artificiale. Per un verso si tiene conto dei rischi connessi al processo di selezione dei dati utilizzati per l'addestramento dei sistemi; per l'altro non si sottovaluta la rischiosità che potrebbe essere insita nelle singole pratiche utilizzate.

Relativamente al primo profilo di cautela citato, ovverosia quello relativo alla tipologia di dati utilizzati nelle fasi di addestramento dei software di AI, rileva l'articolo 53 dell'AI Act, ove si pongono regole volte ad evitare che tali sistemi sviluppino dei *bias* dovuti alla tipologia di dati inseriti e che possano di conseguenza restituire risultati viziati e discriminatori[26]. La norma in esame elenca gli obblighi per i fornitori dei c.d. modelli di AI per finalità generali[27], tra i quali, oltre alla tenuta della documentazione tecnica e al rispetto dei diritti di proprietà intellettuale e dei segreti commerciali, compaiono gli obblighi di trasparenza, tra cui il dovere di fornire informazioni sui dati utilizzati per l'addestramento (Allegato XII, par.2, lett. c). Ancora più rilevante ai fini di una tutela nei confronti di utilizzi discriminatori dei sistemi di AI è l'obbligo di mettere a disposizione del pubblico una «*sintesi sufficientemente dettagliata*» dei contenuti utilizzati per l'addestramento al rispetto delle regole di trasparenza[28].

Questi obblighi rimandano alla duplice natura degli atti regolatori europei in tema di innovazione tecnologica, cioè lo sviluppo del mercato e la tutela dei diritti fondamentali, come ribadito nell'Articolo 1 dell'AI Act[29].

Come anticipato, nel Regolamento sull'Intelligenza Artificiale, lo *status* di persona disabile assume una rilevanza autonoma, collocando le persone disabili entro una sfera di protezione giuridica rafforzata riservata ai gruppi vulnerabili. Questa condizione emerge particolarmente in relazione alle cautele previste per l'utilizzo di determinate pratiche di IA.

---

[26] Per un commento al Regolamento UE 2024/1689 ved. G. CASSANO, E.M. TRIPODI (a cura di), *Il Regolamento europeo sull'Intelligenza artificiale. Commento al Reg. UE n. 1689/2024*, Milano, 2024.

[27] «Un modello di IA, anche laddove tale modello di IA sia addestrato con grandi quantità di dati utilizzando l'autosupervisione su larga scala, che sia caratterizzato una generalità significativa e sia in grado di svolgere con competenza un'ampia gamma di compiti distinti, indipendentemente dalle modalità con cui il modello è immesso sul mercato, e che può essere integrato in una varietà di sistemi o applicazioni a valle, ad eccezione dei modelli di IA utilizzati per attività di ricerca, sviluppo o prototipazione prima di essere immessi sul mercato», cfr. AI Act, Articolo 3 par. 1 n. 63.

[28] Regolamento UE 2024/1689, Art. 53 par. 1 lett. d.

[29] «Lo scopo del presente regolamento è migliorare il funzionamento del mercato interno e promuovere la diffusione di un'intelligenza artificiale (IA) antropocentrica e affidabile, garantendo nel contempo un livello elevato di protezione della salute, della sicurezza e dei diritti fondamentali sanciti dalla Carta dei diritti fondamentali dell'Unione europea, compresi la democrazia, lo Stato di diritto e la protezione dell'ambiente, contro gli effetti nocivi dei sistemi di IA nell'Unione, e promuovendo l'innovazione», cfr. AI Act, Articolo 1 par. 1.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Un esempio significativo è rappresentato dall'articolo 5[30] che elenca le pratiche di Intelligenza Artificiale vietate. In particolare, il paragrafo 1, lettera b), vieta l'immissione nel mercato di sistemi che sfruttano la vulnerabilità di persone o gruppi in modo da distorcerne il comportamento o causare danni significativi, facendo esplicitamente riferimento alle vulnerabilità dovute alla disabilità.
Questo approccio evidenzia l'attenzione del regolamento alla protezione delle persone disabili, riconoscendo che la loro condizione può renderle particolarmente suscettibili a forme di sfruttamento e manipolazione attraverso sistemi di IA. Pertanto, l'AI Act stabilisce un quadro normativo che mira a prevenire tali abusi, garantendo che i sistemi di IA siano sviluppati e utilizzati in modo etico e responsabile, con particolare attenzione ai diritti e alle esigenze dei gruppi vulnerabili, inclusi quelli con disabilità.
Proseguendo nell'analisi delle disposizioni rilevanti per le persone con disabilità, è importante citare il Recital 32. Questo descrive dettagliatamente e con la dovuta enfasi i possibili esiti dannosi della pratica di identificazione biometrica remota in tempo reale in spazi pubblici. Se tale attività fosse svolta per fini di contrasto, potrebbe far sentire la popolazione costantemente sotto sorveglianza, limitando così le principali manifestazioni di libertà e democrazia. Il Considerando 32 evidenzia inoltre come le inesattezze tecniche dei sistemi di Intelligenza Artificiale utilizzati possano causare gravi distorsioni, con effetti discriminatori che colpirebbero gruppi vulnerabili, tra cui le persone con disabilità. Il rischio aumentato di tale pratica è chiaramente correlato all'immediatezza dell'impatto e all'impossibilità di effettuare ulteriori controlli o correzioni in caso di errore.
In conclusione, rimandando l'analisi delle singole disposizioni ad altra sede, è fondamentale sottolineare come il Considerando 48 correli il livello di rischiosità di un sistema di Intelligenza Artificiale all'impatto negativo sui diritti fondamentali protetti dalla Carta di Nizza, richiamando esplicitamente anche i diritti delle persone con disabilità. La violazione di tali diritti diventa così un criterio per la valutazione dell'impatto sui diritti fondamentali per i sistemi di AI ad alto rischio, ex Articolo 27, paragrafo 1, lettera c) dell'AI Act.
Le disposizioni sopra menzionate mettono in luce, da un lato, l'attenzione alla qualità dei dati e alla riservatezza delle informazioni, e dall'altro, una costante valutazione della rischiosità di determinati sistemi di Intelligenza Artificiale, considerando i potenziali effetti lesivi su categorie vulnerabili come le persone con disabilità. Come si è evidenziato, l'approccio adottato mira principalmente, sebbene non esclusivamente, a evitare discriminazioni basate sulla vulnerabilità di questi soggetti.
Per approfondire l'impatto dell'utilizzo dell'intelligenza artificiale sulle persone con disabilità, è necessario ora fare una digressione sul principio di uguaglianza e sulle sue correlazioni con i diritti delle minoranze, fino ad arrivare all'affermazione del diritto alla diversità.

3. Principio di uguaglianza, minoranze e diritto alla diversità

Dal 1789 in poi, il principio di uguaglianza ha costituito la pietra angolare di qualsiasi conquista democratica, trovando la sua espressione in tutte le Costituzioni e nei principali cataloghi internazionali dei diritti fondamentali. Molte delle garanzie di tutela presenti nelle Costituzioni e nelle Convenzioni sui

[30] La lettera h) del primo comma lascia impregiudicato l'articolo 9 del regolamento (UE) 2016/679 per quanto riguarda il trattamento dei dati biometrici a fini diversi dall'attività di contrasto.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

diritti umani derivano chiaramente da questo principio; tra i diritti di più immediata derivazione si annoverano il diritto all'uguaglianza, il divieto di discriminazione e il diritto alla diversità (o alla differenza).

La nascita dello Stato sociale ha portato alla proclamazione del principio di uguaglianza sostanziale, che tiene conto delle diverse condizioni di vita degli individui per eliminare le situazioni di svantaggio e garantire a tutti uguali condizioni di partenza. Questo cambiamento ha fatto sì che lo Stato, inizialmente mantenuto a distanza dai rivoluzionari per tutelare l'esercizio delle libertà negative, intervenisse poi attivamente per assicurare a tutti i cittadini il pieno godimento dei diritti fondamentali. Un esempio compiuto e significativo dei risultati di questo percorso storico è rappresentato dalla Costituzione italiana, che accoglie entrambe le accezioni del principio di uguaglianza e le integra in una norma fondamentale.

La collocazione di tale norma, subito dopo l'enunciazione del principio di sovranità popolare e il riconoscimento dei diritti inviolabili, ne conferma il ruolo centrale. Nell'Articolo 3 della Costituzione trovano fondamento sia il principio di eguaglianza formale che quello di eguaglianza sostanziale. L'eguaglianza formale, affermatasi a partire dalla Rivoluzione francese, si concretizza nel divieto per il legislatore di adottare trattamenti irragionevolmente differenziati tra i cittadini[31]. Questo principio viene violato sia quando il legislatore tratta in modo ingiustificatamente uguale situazioni diverse, sia quando discrimina ingiustificatamente situazioni analoghe[32].

D'altro canto, l'eguaglianza sostanziale, sancita dal secondo comma dell'articolo, autorizza il Parlamento ad adottare leggi differenziate, a favore di specifiche categorie, con l'obiettivo di colmare, attraverso interventi mirati, situazioni di svantaggio e garantire la parità nelle condizioni di partenza[33]. I due principi si limitano e si completano reciprocamente: l'eguaglianza sostanziale attenua la rigidità dell'eguaglianza formale, che non ammetterebbe eccezioni, mentre l'applicazione del principio di eguaglianza formale evita che le misure di azione positiva, previste dall'Art. 3, secondo comma, possano diventare causa di discriminazioni a rovescio[34].

La Corte costituzionale ha definito l'eguaglianza «*un principio generale che condiziona tutto l'ordinamento nella sua obiettiva struttura*[35]». In effetti l'eguaglianza è un principio supremo, che non può essere messo in discussione o in alcun modo eliminato[36]. «*L'eguaglianza – in forza della sua tradizione storica, della strettissima connessione con l'essenza della democrazia, della vocazione generale, della*

[31] P. CARETTI, U. DE SIERVO, Istituzioni di diritto pubblico, Torino, 2006, 448.

[32] *Ibidem.*

[33] Una analisi dell'Art. 3 II comma in B. CARAVITA, *Oltre l'eguaglianza formale, Un'analisi dell'Art. 3 comma 2 della Costituzione*, Padova, 1984.

[34] L. AZZENA, *Divieto di discriminazione e posizione dei soggetti "deboli". Spunti per una teoria della "debolezza",* in C. CALVIERI (a cura di), *Divieto di discriminazione nella giurisprudenza costituzionale: atti del seminario di Perugia del 18 maggio 2005*, Torino, 2006, 46.

[35] Corte costituzionale, sentenza n. 25 del 17 marzo 1966, in www.giurcost.org.

[36] Così Celotto: «[…] possiamo dire che anche se fosse esplicitamente abrogata la previsione dell'Art. 3 I comma Cost., il principio di eguaglianza formale continuerebbe ad operare nel nostro ordinamento in maniera pressoché identica», in R. BIFULCO, A. CELOTTO, M. OLIVETTI (a cura di), Commentario *alla Costituzione,* Volume 1, Artt. 1-54, Torino, 2006, 69.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

presenza in tutte le Costituzioni contemporanee – costituisce un 'principio generalissimo', una 'super-norma', destinata ad operare come 'norma di chiusura' dell'ordinamento[37]».

L'Articolo 3, secondo comma, della Costituzione è stato al centro di un acceso dibattito all'interno dell'Assemblea costituente, in particolare riguardo al concetto di "rimozione degli ostacoli". Non tutti i costituenti condividevano l'idea di inserire questo concetto nel testo costituzionale. L'immagine di una Repubblica che si occupa di rimuovere gli ostacoli, paragonata a «*una squadra di operai intenti a levare dei massi, a togliere dalla strada qualcosa per far passare l'uomo*[38]», non era un'idea condivisa da tutti. I detrattori di questa formula sostenevano che la Repubblica dovesse occuparsi del pieno sviluppo della personalità umana, come concetto omnicomprensivo e «*meno materializzato*[39]», volto ad una estensione dell'azione positiva dello Stato, che andasse oltre la semplice rimozione delle barriere materiali.

Tuttavia, come dimostra il testo definitivo dell'Art. 3 della Costituzione, il concetto di "rimozione degli ostacoli" fu considerato il più appropriato per descrivere l'azione dello Stato, orientata a creare le condizioni per una reale uguaglianza sostanziale. Questa formula esprime in modo concreto il compito della Repubblica di intervenire attivamente per superare le disuguaglianze di fatto, garantendo a tutti i cittadini pari opportunità e il pieno sviluppo della persona umana: «Partiamo dalla constatazione della realtà, perché mentre con la rivoluzione dell' '89 è stata affermata l'eguaglianza giuridica dei cittadini membri di uno stesso Stato, lo studio della vita sociale in quest'ultimo secolo ci dimostra che questa semplice dichiarazione non è stata sufficiente a realizzare tale eguaglianza, e fa parte della nostra dottrina sociale una serie di rilievi e di constatazioni circa gli ostacoli che hanno impedito di fatto la realizzazione dei principi proclamati nell' '89[40]».

L'Articolo 3, nel suo primo comma, vieta tutte le discriminazioni basate su categorie specifiche, come sesso, razza, lingua, religione, opinioni politiche, e condizioni personali e sociali. Quest'ultima categoria comprende tutti gli atti che ledono la dignità dei singoli in situazioni particolari, inclusi i comportamenti discriminatori nei confronti delle persone con disabilità[41]. Il secondo comma dell'Articolo 3, invece, stabilisce le linee guida per garantire una concreta eguaglianza tra tutti gli individui, assicurando che anche coloro che convivono con una qualche forma di disabilità possano godere pienamente dei loro diritti, attraverso l'eliminazione delle barriere che ne ostacolano la partecipazione.

Proseguendo il percorso che dal principio di uguaglianza porterà al diritto alla diversità, occorre ora accostare la tematica della disabilità alla questione della tutela delle minoranze. Prima di procedere, è importante sottolineare che chi scrive è pienamente consapevole del fatto che la categoria delle persone disabili non può essere considerata come un insieme dai confini rigidi e immutabili. Al contrario la condizione di disabilità nella maggior parte dei casi non è insita nell'uomo, ma è collegata alle vicende della vita; per di più l'*handicap* che dalla disabilità può derivare è assolutamente condizionato

[37] *Ibidem*, 68.

[38] Intervento dell'On. Corbino nell' Assemblea Costituente, seduta del 24 marzo 1947, in Atti della Assemblea Costituente, Discussione sul progetto di Costituzione, Volume I, Tipografia della Camera dei Deputati, 1951, 2422.

[39] *Ibidem*.

[40] Intervento dell'On. Fanfani nell'Assemblea Costituente, seduta del 24 marzo 1947, in Atti della Assemblea Costituente, Discussione sul progetto di Costituzione, Volume I, Tipografia della Camera dei Deputati, 1951, 2424-2425.

[41] P. Caretti, U. De Siervo, *Istituzioni di diritto pubblico*, Torino, 2006, 453.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

dalle fattispecie del caso concreto e ben può variare a seconda delle soluzioni di volta in volta predisposte. Ciò nondimeno, considerare i disabili come categoria a carattere tendenzialmente permanente[42] può essere un utile espediente per ragionare sui diritti, in particolare su quelli che non ricevono una tutela adeguata e che risultano meglio difendibili se interpretati attraverso un'ottica collettiva.

Dopo queste necessarie premesse, si farà riferimento alla dottrina sulla tutela delle minoranze territoriali, etniche e religiose per tracciare un parallelo con la minoranza costituita dalle persone con disabilità. Tale confronto consentirà di esaminare i vari aspetti che influenzano l'applicazione del principio di uguaglianza.

È utile, innanzitutto, richiamare la distinzione tra minoranze discriminate e minoranze volontarie. Per le prime, l'obiettivo principale è quello di vedere rimosse le discriminazioni che sono attuate dalla maggioranza, e vedere affermato il pieno diritto a godere dei medesimi diritti del gruppo dominante. Obiettivo principale delle minoranze volontarie è invece la difesa e conservazione della propria specificità e la garanzia delle condizioni affinché ciò si possa realizzare. I differenti obiettivi delle due tipologie di minoranze appena descritte si prestano ad essere sovrapposti ai due profili, formale e sostanziale, del principio di eguaglianza. Se l'Articolo 3, comma 1, della Costituzione garantisce alle minoranze tutela e protezione contro le discriminazioni, d'altra parte, l'Articolo 3, comma 2, sembra costituire la base giuridica sulla quale le minoranze volontarie possano rivendicare il loro diritto alla diversità.

Autorevole dottrina sostiene che «con riferimento alle situazioni minoritarie a carattere tendenzialmente permanente è assolutamente normale che si realizzi una situazione di fatto tale da comportare la necessità che per dare applicazione al principio di eguaglianza debba ricorrersi a forme di tutela positiva. E infatti a tali situazioni è intrinseca l'esistenza di un rapporto di inferiorità della minoranza nei confronti della corrispondente maggioranza, tanto ove questo rapporto sia determinato dallo squilibrio puramente numerico, quanto ove esso sia realizzato altresì – come spesso avviene – da un divario di ordine economico o culturale.

Si può dire perciò che le applicazioni del principio di eguaglianza ai rapporti fra i gruppi che danno luogo a situazioni di tipo minoritario siano quasi inevitabilmente riconducibili a quelle che comportano l'impiego della nozione di "eguaglianza sostanziale" e di forme di tutela positiva[43]».

E' stata la stessa Corte costituzionale a definire le azioni positive come «il più potente strumento a disposizione del legislatore, che, nel rispetto della libertà e dell'autonomia dei singoli individui, tende a innalzare la soglia di partenza per le singole categorie di persone socialmente svantaggiate - fondamentalmente quelle riconducibili ai divieti di discriminazione espressi nel primo comma dello stesso art. 3 (sesso, razza, lingua, religione, opinioni politiche, condizioni personali e sociali) - al fine di assicurare alle categorie medesime uno statuto effettivo di pari opportunità di inserimento sociale, economico e politico[44]».

Il mondo della disabilità si offre come paradigma perfetto dell'unione delle due esigenze di tutela sopra richiamate e delle diverse modalità di realizzazione di tali garanzie.

Certamente, la persona con disabilità rivendicherà il suo diritto a vivere come qualsiasi altro cittadino, senza subire discriminazioni a causa delle proprie limitazioni funzionali. Al contempo, essa affermerà

[42] A. PIZZORUSSO, *Minoranze e maggioranze*, Torino, 1993, 83.
[43] *Ibidem*.
[44] Corte costituzionale, sentenza n. 109 del 24 marzo 1993, considerato in diritto, punto 2.2.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

la propria "originalità" come essere umano diversamente abile, richiedendo accomodamenti che le permettano di condurre una vita normale, senza che le barriere architettoniche, sociali e culturali compromettano le peculiarità psicofisiche che la contraddistinguono. Se il primo profilo appare già ampiamente implementato e garantito da convenzioni, direttive e leggi, il secondo appare ancora ricco di potenziale, specie nell'era dell'Intelligenza artificiale.

Anche in questo caso ci basterà compiere una rapida ricognizione.

Sul primo versante, l'articolo 21 della Carta dei diritti fondamentali dell'Unione europea, come si è visto, vieta qualunque forma di discriminazione, tra cui quelle basate sulla disabilità. Il Considerando 32 dell'AI Act, anch'esso già citato, in relazione ai sistemi di identificazione biometrica remota in tempo reale recita: «Le inesattezze di carattere tecnico dei sistemi di IA destinati all'identificazione biometrica remota delle persone fisiche possono determinare risultati distorti e comportare effetti discriminatori. Tali possibili risultati distorti ed effetti discriminatori sono particolarmente importanti per quanto riguarda […] le disabilità». Inoltre, nel Regolamento 2024/1689 si pone molta attenzione alle discriminazioni che potrebbero aver luogo in conseguenza dell'utilizzo di sistemi di AI nell'ambito dell'istruzione[45] e dei rapporti di lavoro[46] (cfr. Considerando 29,48, Allegato III).

Il diritto alla diversità invece è riconosciuto dalla Convenzione ONU sui diritti delle persone con disabilità[47], mentre il principio di diversità compare tra gli orientamenti etici per un'IA affidabile del 2019[48]. Il riferimento alla diversità in questo ultimo documento è volto a orientare lo sviluppo di sistemi di IA attraverso l'inclusione di soggetti diversi, con la promozione della parità di accesso, dell'uguaglianza di genere e della diversità culturale.

Salvi i due richiami alla diversità appena svolti, però, appare a chi scrive che la tutela di questo diritto fondamentale goda di minori garanzie e tutele giuridiche rispetto al diritto alla non discriminazione.

[45] «La diffusione dei sistemi di IA nell'istruzione è importante per promuovere un'istruzione e una formazione digitali di alta qualità e per consentire a tutti i discenti e gli insegnanti di acquisire e condividere le competenze e le abilità digitali necessarie, compresa l'alfabetizzazione mediatica, e il pensiero critico, per partecipare attivamente all'economia, alla società e ai processi democratici. […] Se progettati e utilizzati in modo inadeguato, tali sistemi possono essere particolarmente intrusivi e violare il diritto all'istruzione e alla formazione, nonché il diritto alla non discriminazione, e perpetuare modelli storici di discriminazione, ad esempio nei confronti delle donne, di talune fasce di età, delle persone con disabilità o delle persone aventi determinate origini razziali o etniche o un determinato orientamento sessuale», Regolamento UE 2024/1689, Considerando 29.

[46] «Durante tutto il processo di assunzione, nonché ai fini della valutazione e della promozione delle persone o del proseguimento dei rapporti contrattuali legati al lavoro, tali sistemi possono perpetuare modelli storici di discriminazione, ad esempio nei confronti delle donne, di talune fasce di età, delle persone con disabilità o delle persone aventi determinate origini razziali o etniche o un determinato orientamento sessuale. I sistemi di IA utilizzati per monitorare le prestazioni e il comportamento di tali persone possono inoltre comprometterne i diritti fondamentali in materia di protezione dei dati e vita privata», Regolamento UE 2024/1689, Considerando 48. Cfr. anche Allegato III.

[47] Convenzione delle Nazioni Unite sui diritti delle persone con disabilità del 13 dicembre 2006, Preambolo, lettera i).

[48] Documento elaborato dal gruppo di esperti ad alto livello sull'IA, ved. https://digital-strategy.ec.europa.eu/it/library/ethics-guidelines-trustworthy-ai (ultima consultazione 01/12**/2024). I sette principi etici volti a garantire che l'IA sia eticamente valida ed affidabile comprendono intervento e sorveglianza umani, robustezza tecnica e sicurezza, vita privata e governance dei dati, trasparenza, diversità, non discriminazione ed equità, benessere sociale e ambientale e responsabilità.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Molto si potrebbe ancora ottenere su questo fronte, anche attraverso un uso regolato e sapiente dei nuovi strumenti di Intelligenza Artificiale, come si cercherà di spiegare nell'ultimo paragrafo.

4. Dalla non discriminazione al diritto alla diversità: una nuova missione per l'AI?

La ricerca condotta ha evidenziato che l'apparato giuridico volto a garantire la non discriminazione delle persone con disabilità è attualmente ben strutturato. Questo diritto è sancito da una convenzione internazionale specifica, ribadito nelle principali carte dei diritti fondamentali, e trova attuazione concreta sia nella legislazione europea che nelle normative nazionali di recepimento delle direttive comunitarie.

Tuttavia, lo stesso non si può affermare per quanto riguarda il diritto alla diversità, che le persone con disabilità potrebbero legittimamente rivendicare. Come discusso nelle pagine precedenti, questo diritto emerge nel punto di intersezione tra la realizzazione del principio di uguaglianza sostanziale e l'affermazione dei diritti delle minoranze.

Lo Stato deve attivarsi per garantire il godimento dei diritti a tutti i cittadini, comprese le persone con disabilità, ma è necessario che tale azione positiva sia disegnata sulla specificità del singolo o di un gruppo omogeneo di individui, rispettando la loro diversità. Di fronte a queste esigenze di tutela, gli strumenti offerti dalle nuove tecnologie potrebbero offrire un grande supporto alle persone con disabilità e ancor prima allo Stato che ha il dovere di farsi carico delle situazioni di vulnerabilità. L'applicazione di nuove soluzioni altamente innovative dovrà però collocarsi all'intero del quadro normativo qui più volte richiamato.

Ad esempio, è interessante rileggere l'articolo 22 del GDPR, che riguarda i trattamenti automatizzati, inclusa la profilazione. Sebbene generalmente vietate, attività come la profilazione potrebbero, in questa prospettiva, diventare il preludio a prestazioni, servizi o sussidi altamente personalizzati. Si consideri l'utilizzo dei sistemi di intelligenza artificiale per supportare la comunicazione, in particolare il riconoscimento vocale: sebbene attualmente tali sistemi presentino delle lacune quando utilizzati da categorie di utenti meno rappresentate, un corretto addestramento potrebbe renderli un valido strumento di supporto per persone sorde o con disturbi specifici del linguaggio[49]. Il confine tra l'uso legittimo e quello improprio delle informazioni sensibili è estremamente sottile; tuttavia, con un consenso esplicito, informato e ottenuto in modo legittimo, tale utilizzo potrebbe non solo costituire un'eccezione, ma anche rappresentare la base per l'esercizio di diritti fondamentali. Inoltre, come ben espresso da autorevole dottrina[50], dal combinato disposto degli articoli 13, 14 e 22 e del Considerando 72 del GDPR, si possono trarre i principi di conoscibilità e comprensibilità degli algoritmi, il principio di non esclusività della decisione algoritmica e il principio di non discriminazione algoritmica[51], che costituiscono ulteriori garanzie a tutela dell'interessato, soprattutto, e ancor più, nel caso in cui la sua maggiore vulnerabilità sia dovuta alla presenza di una o più disabilità. Sempre in tema di trasparenza e

[49] Cfr. S. TREWIN, *AI Fairness for People with Disabilities: Point of View,* arXiv-1811, 2018.

[50] A. SIMONCINI, *L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà*, in *Biolaw Journal - Rivista di BioDiritto*, 1, 2019, 63 ss.

[51] Si veda A. SIMONCINI-S. SUWEIS, *Il cambio di paradigma nell'intelligenza artificiale e il suo impatto sul diritto costituzionale*, in *Rivista di filosofia del diritto*, I, giugno 2019, 86 ss., e A. SIMONCINI, *op. cit.,* 63 ss.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

conoscibilità, il Considerando 132 dell'AI Act contiene un esplicito invito a considerare le diversità delle persone con disabilità, al fine di tutelarle specificamente quando interagiscono con un sistema di IA che comporti rischi specifici, come impersonificazione o inganno. La trasparenza, in questa prospettiva, assume dunque una connotazione più personale, diventando un'espressione del diritto alla diversità.
Interessante, in questa prospettiva di rafforzamento del diritto alla diversità per le persone con disabilità, è anche l'articolo 95. Questo articolo assegna all'Ufficio per l'IA e agli Stati membri il compito di promuovere l'elaborazione di Codici di condotta da parte dei fornitori e degli utilizzatori di sistemi di Intelligenza Artificiale, indipendentemente dal loro livello di rischio. Tra gli obiettivi di questi Codici dovrebbero figurare la progettazione e lo sviluppo inclusivi e diversificati dei sistemi di IA, nonché la valutazione e prevenzione degli impatti negativi sull'accessibilità per le persone con disabilità. Occorrerà attendere ancora un po' di tempo per valutare l'efficacia di questa norma come incentivo alla produzione di tali Codici di condotta e per verificarne l'applicazione concreta.
Merita di ricordare, infine, che il Regolamento 2024/1689 contiene anche disposizioni che invitano gli Stati membri a sostenere attività di ricerca e sviluppo di soluzioni di IA che possano portare a risultati vantaggiosi dal punto di vista sociale e ambientale. In particolare, il regolamento promuove lo sviluppo di soluzioni basate sull'IA per migliorare l'accessibilità per le persone con disabilità[52]. Esempi virtuosi di progetti volti ad aumentare l'inclusione e l'accessibilità delle persone disabili negli spazi urbani confermano il grande potenziale dell'azione di soggetti pubblici e privati in ottica proattiva e di valorizzazione della specificità e della diversità delle persone con disabilità[53]. Va detto che la bocciatura da parte del Garante per la protezione dei dati personali di progetti che prevedevano la profilazione dei cittadini attraverso l'elaborazione algoritmica di dati sullo stato di salute[54] ha messo in evidenza la complessità e la delicatezza del contesto e dei valori in gioco. È auspicabile per il futuro la realizzazione interventi legislativi mirati che insieme ad una corretta architettura che integri la conformità al GDPR con il rispetto delle norme dell'AI Act potranno consentire concreti avanzamenti e una maggiore tutela del diritto alla diversità delle persone con disabilità.
In sintesi, l'evoluzione normativa e tecnologica rappresenta sia una sfida che una grande opportunità per la tutela dei diritti delle persone con disabilità. Se da un lato l'attuale quadro giuridico offre già strumenti efficaci per contrastare la discriminazione, dall'altro il consolidarsi del diritto alla diversità richiede un impegno costante e mirato. La profilazione e l'impiego dell'Intelligenza Artificiale, se gestiti

[52] Regolamento UE 2024/1689, Considerando 142: «Per garantire che l'IA porti a risultati vantaggiosi sul piano sociale e ambientale, gli Stati membri sono incoraggiati a sostenere e promuovere la ricerca e lo sviluppo di soluzioni di IA a sostegno di risultati vantaggiosi dal punto di vista sociale e ambientale, come le soluzioni basate sull'IA per aumentare l'accessibilità per le persone con disabilità, affrontare le disuguaglianze socioeconomiche o conseguire obiettivi in materia di ambiente, assegnando risorse sufficienti, compresi i finanziamenti pubblici e dell'Unione, e, se del caso e a condizione che siano soddisfatti i criteri di ammissibilità e selezione, prendendo in considerazione soprattutto i progetti che perseguono tali obiettivi. Tali progetti dovrebbero basarsi sul principio della cooperazione interdisciplinare tra sviluppatori dell'IA, esperti in materia di disuguaglianza e non discriminazione, accessibilità e diritti ambientali, digitali e dei consumatori, nonché personalità accademiche».
[53] W. Del Negro, M. Lazzati, *Promoting Accessibility in European Metros: The Power of Open Data*, in Data Valley Consulting White paper, *Use and Reuse of Urban Data in the Smart city domain*, 2024, 77 ss.
[54] Cfr. Garante per la protezione dei dati personali, *Parere al Consiglio di Stato sulle nuove modalità di ripartizione del fondo sanitario tra le regioni proposte dal Ministero della salute e basate sulla stratificazione della popolazione* - 5 marzo 2020, docweb n. 9304455.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

con trasparenza e in conformità con il GDPR e l'AI Act, possono trasformarsi da potenziali rischi a potenti alleati per l'inclusione e la personalizzazione dei servizi. Le raccomandazioni e le direttive contenute nelle normative più recenti, insieme alle decisioni delle autorità competenti, indicano il cammino verso un futuro in cui la diversità non solo è riconosciuta, ma diventa un valore fondamentale.
Pertanto, è fondamentale che istituzioni, legislatori e fornitori di tecnologie collaborino attivamente per garantire che i benefici dell'Intelligenza Artificiale siano accessibili a tutti, nel pieno rispetto e nella promozione dei diritti fondamentali delle persone con disabilità. Solo attraverso un impegno congiunto sarà possibile trasformare le potenzialità tecnologiche in strumenti concreti di inclusione e uguaglianza.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Protezione e *empowerment* dei minori nell'era dell'intelligenza artificiale: coordinate costituzionali

Nadia Maccabiani*

PROTECTION AND EMPOWERMENT OF CHILDREN IN THE AGE OF AI: CONSTITUTIONAL COORDINATES
ABSTRACT: In respect of the multidimensional implications of Artificial Intelligence systems, this writing intends to focus on relational aspects, particularly regarding minors who build their personal identity and develop their personality through social relationships. Recent European acts, such as the Digital Services Act and the Artificial Intelligence Act, offer some safeguards that need to be complemented by the traditional constitutional duties of education entrusted to families and schools. The objective of the paper is to "update" the understanding of the constitutional framework within which such duties are embedded, both for the protection and empowerment of minors in order to incorporate the "virtuality" of artificial intelligence systems.

KEYWORDS: Artificial Intelligence; Online Platform; Social Media; Children Protection; Children Empowerment.

ABSTRACT: Rispetto alle multidimensionali implicazioni dei sistemi di Intelligenza Artificiale, lo scritto si sofferma sugli aspetti di natura relazionale, con specifico riguardo ai minori di età che, attraverso i rapporti sociali, formano l'identità personale e sviluppano la personalità. Recenti atti europei, quali il *Digital Services Act* e l'*Artificial Intelligence Act*, offrono alcuni indirizzi destinati a trovare completamento nei tradizionali doveri costituzionali di educazione ed istruzione affidati a famiglia e scuola. L'obiettivo è quindi di "ri-attualizzare" la lettura delle coordinate costituzionali entro cui si incardinano tali doveri costituzionali, sia per la protezione che l'*empowerment* dei minori, inglobando le "virtualità" dei sistemi di intelligenza.

PAROLE CHIAVE: Intelligenza Artificiale; Piattaforme Online; Social Media; Protezione dei minori; Empowerment dei minori.

* *Professoressa Associata, Università di Brescia, e-mail: nadia.maccabiani@unibs.it.* *Contributo sottoposto a doppio referaggio anonimo.*

Downloaded from www.biodiritto.org.
ISSN 2284-4503

1. Introduzione

L'essere umano è connotato da una profonda dimensione relazionale, quale «individuo 'realizzato' nella società, *entelechia* dell'uomo 'animale sociale'»[1]. Al riguardo, gli articoli 2-3 della Costituzione sono icastici. Risulta altresì evocativo il noto ordine del giorno Dossetti, presentato in Assemblea Costituente, che, nel sancire il principio personalista, richiamava la presupposta, quanto necessaria, socialità delle persone, destinate a completarsi e perfezionarsi a vicenda mediante una reciproca solidarietà economica e spirituale, attraverso comunità intermedie, per poi giungere allo Stato. Siamo di fronte ad una vera e propria «pedagogia costituzionale» in grado di far «apprezzare la natura interrelata della vita e delle azioni di ciascuno rispetto al resto della comunità»[2]. La stessa Corte costituzionale non ha mancato di sottolineare la «primigenia vocazione sociale dell'uomo, derivante dall'originaria identificazione del singolo con le formazioni sociali in cui si svolge la sua personalità e dal conseguente vincolo di appartenenza attiva che lega l'individuo alla comunità degli uomini»[3]. Socialità come "ricchezza", quindi, poiché essenziale per la formazione e lo svolgimento della personalità.

Ma l'intrinseca "relazionalità" dell'uomo racchiude un'insidia, rivelatrice della "fragilità" della condizione umana, dovuta alla necessaria dipendenza dagli altri[4]. Fragilità accentuata per i minori, essendo il grado di dipendenza (non solo materiale, ma anche immateriale, spirituale) inversamente proporzionale alla maturità psico-emotiva dell'individuo. L'Istituto Superiore di Sanità ha evidenziato come scuola e famiglia siano luoghi privilegiati «in cui i bambini, le bambine, i ragazzi e le ragazze possono sviluppare la propria personalità, la coscienza critica e la conoscenza di sé, il senso di responsabilità e della propria autonomia individuale»[5]. La dottrina costituzionalistica ha in merito sottolineato la «dimensione fortemente relazionale»[6] di scuola e famiglia, all'interno della quale si esplicano quelli che, non a caso, sono stati definiti «diritti relazionali» del minore[7], posti alla base di uno «statuto costituzionale della persona minore di età»[8]. La stessa Corte costituzionale, sin dalla storica sentenza n. 11 del 1981, ha valorizzato il bisogno di relazioni sociali dei minori per lo svolgimento della loro personalità[9].

---

[1] F.D. Busnelli, V. Calederai, *Declinazioni della persona: un itinerario dal diritto privato al diritto internazionale (passando per il diritto costituzionale),* in *Giurisprudenza Italiana,* 2010, 2214.

[2] E. Rossi, *La doverosità dei diritti: analisi di un ossimoro costituzionale?,* in *Gruppo di Pisa*, 2018, 72.

[3] Corte cost. 17-28 febbraio 1992, n. 75, punto 2 del considerato in diritto, dove la Corte parla altresì di «profonda socialità che caratterizza la persona stessa».

[4] A. Macintyre, *Dependent Rational Animals: Why Human Beings Need the Virtues,* Londra, 1999.

[5] Cfr. C. Mortali, L. Mastrobattista, I.Palmi, R. Solimini, R. Pacifici, S. Pichini, A. Minutillo, *Dipendenze comportamentali nella Generazione Z: uno studio di prevalenza nella popolazione scolastica (11-17 anni) e focus sulle competenze genitoriali*, Rapporto ISTISAN 23/25, 2023, V.

[6] C. Di Costanzo, *La tutela costituzionale del minore: identità, salute e relazioni,* Torino, 2023, 10.

[7] C. Di Costanzo, *op. cit.*

[8] G. Matucci, *Lo statuto costituzionale del minore d'età*, Padova, 2015, 58.

[9] Corte cost. 10 febbraio 1981, n. 11, punto 5 considerato in diritto, dove viene posto in evidenza il «valore primario [del]la promozione della personalità del soggetto umano in formazione e la sua educazione nel luogo a ciò più idoneo» nonché il bisogno relazionale «avvertito con forza dal minore, che richiede per la sua crescita normale affetti individualizzati e continui, ambienti non precari, situazioni non conflittuali».

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Se quindi, come pare assodato da quanto brevemente ricordato ed esposto, il minore in fase di formazione realizza la propria personalità attraverso le relazioni intessute con l'esterno, ne deriva che i cambiamenti dell'ambiente entro tale "relazionalità" si sviluppa, producono riflessi sulla identità del minore, sulla costruzione del proprio sé. Materia per gli studi di filosofia, psicologia, sociologia, neuroscienza verrebbe da dire. Senz'altro è così. Ma anche materia costituzionale, se le dinamiche che si instaurano in tale ambiente relazionale interferiscono con le coordinate costituzionali che qualificano la profonda quanto essenziale socialità umana, volta a contribuire alla strutturazione della capacità di libera autodeterminazione, quindi allo sviluppo di identità e personalità nel rispetto della dignità umana[10].

Ora, nel presente scritto l'attenzione è posta sull'integrazione di tale "ambiente relazionale" con quelle *disruptive technologies* qualificate come sistemi di Intelligenza Artificiale, con specifico riguardo ai minori di età. Partendo da due premesse. Da un lato, per sistemi di IA, secondo la definizione data *dall'Artificial Intelligence Act* (AI Act), si intendono sistemi automatizzati (quindi qualificati da grandi capacità computazionali) che, secondo diversi livelli di autonomia, sulla base di *inputs* (grandi quantità di dati processati da algoritmi variamente sofisticati) generano *outputs* in termini di previsioni, decisioni, contenuti, raccomandazioni[11]. Dall'altro lato, particolare attenzione va posta su quei sistemi che sono dotati di capacità accentuatamente relazionali, in grado di comprendere preferenze, interessi, emozioni dell'interlocutore e quindi ad inferire valutazioni sul suo "pensiero", adeguando, conseguentemente, i *feedbacks* forniti[12]. Si tratta di sistemi talora caratterizzati da abilità linguistiche, in grado di processare e riprodurre linguaggio umano[13], siano essi "impersonati" in robot sociali o semplici applicazioni, quindi *software*[14]. Tali sistemi, restituendo risposte a richieste (c.d. *prompt*), "dialogano" proprio con quello strumento (il linguaggio) che ha sempre contraddistinto l'uomo dagli altri animali[15]; offrono inoltre il loro modo di pensare, la loro conoscenza e visione del mondo, con il rischio di contribuire ad "atrofizzare" la capacità di ricerca individuale, la valutazione e autonomia di giudizio della persona[16] (il

[10] A. SANTOSUOSSO, *About coevolution of humans and intelligent machines: preliminary notes*, in *BioLaw Journal,* S1, 2021, 7 ss.

[11] Cfr. art. 3, par. 1, punto 1.

[12] S. MCCARTHY-JONES, *The Autonomous Mind: The Right to Freedom of Thought in the Twenty-First Century,* in *Frontiers of Artificial Intelligence,* n. 2, 2019, 1: «The ability to think freely is so essential to our identity that to violate it is to deprive us "of personhood altogether"».

[13] I *large language models* rientrano in quella particolare tipologia di sistemi di AI che l'AI Act definisce Modelli di IA per finalità generali, cfr. Artt. 51 ss., e Allegato XIII.

[14] In merito aveva fatto notizia, anche per le implicazioni etiche, il caso di Alexa che leggeva una favola a un bambino con la voce, perfettamente riprodotta, della nonna morta poco tempo prima, cfr. D. SISTO, *Alexa dà voce ai morti? Così cambia il nostro rapporto col lutto*, in *agendadigitale.eu,* 28 giugno 2022.

[15] A. SIMONCINI, *Il linguaggio dell'intelligenza artificiale e la tutela costituzionale dei diritti,* in *Rivista AIC,* n. 2/2023, 23, osserva: «il mezzo che gli esseri umani utilizzano ordinariamente per comunicare – scambiare informazioni, domande, comandi - è proprio il linguaggio. Per questo nella prospettiva che abbiamo descritto, quella cioè di tecnologie progettate per interagire con le persone, è fondamentale il tema del linguaggio. Per poter "interloquire" e, dunque, per realizzare la propria funzione, queste "macchine sociali" debbono condividere la lingua con gli esseri umani».

[16] La nostra capacità di pensiero viene "catturata" dall'algoritmo, cfr. A. SIMONCINI, *L'algoritmo incostituzionale: intelligenza artificiale e future delle libertà,* in *BioLaw Journal,* 1, 2019, 69; nonché, con particolare riguardo ai minori, cfr. documento UNESCO del luglio 2023, *Generative AI and the future of education,* predisposto da S.GIANNINI.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

c.d. *effet moutonnier*)[17]. Con il conseguente rischio di contribuire a "rinchiudere" la persona all'interno di realtà che nella virtualità trovano la loro quintessenza, allontanandola quindi dalla sua innata (quanto umana) socialità[18]. Senza peraltro scordare gli effetti che il filtro algoritmico delle informazioni è in grado di produrre sul dibattito pubblico, incrinando pluralismo delle opinioni e delle idee, producendo le ben note *eco chambers* e *filter bubbles*, con relativa polarizzazione sociale[19].

Da qui il campanello d'allarme. Nello specifico, per dirla con le recenti parole del Garante privacy, i rischi sono significativi «specialmente per lo sviluppo intellettuale e culturale delle giovani generazioni…L'IA può influenzare il modo in cui apprendiamo, pensiamo e interagiamo gli uni con gli altri, potenzialmente limitando la nostra capacità di pensiero critico e la nostra creatività», quindi **«**tocca questioni fondamentali di autonomia e identità personale fino ad arrivare al libero arbitrio, alla capacità di discernimento fra ciò che è giusto e ciò che è sbagliato», con la conseguenza che «se le giovani generazioni crescono in un ambiente in cui l'IA fornisce risposte pronte e soluzioni facili, risultati senza fatica né consultazione di più fonti, potrebbero perdere la capacità di affrontare sfide complesse, di sviluppare il pensiero critico e di valutare le informazioni in modo consapevole, meditato, indipendente»[20].

Se questo è il terreno sul quale giocare la "partita", si tratta di capire quali sono le "coordinate costituzionali" da preservare nel *best interest of the child.* Si prenderanno le mosse dagli interventi del legislatore europeo, per poi approdare alle comunità principalmente incaricate, secondo le previsioni costituzionali, dell'educazione e dell'istruzione dei minori di età.

2. Protezione dei minori dall'IA: l'approccio europeo

Gli atti dell'UE ai quali si avrà riguardo sono il Regolamento Europeo sull'IA (AI Act) e quello sui Servizi Digitali (DSA)[21], con specifico riferimento alle disposizioni rivolte ai minori di età. Entrambi introducono essenzialmente cautele di ordine generale e "in negativo", volte a proteggere *dall'*intelligenza artificiale, mentre nulla specificamente prescrivono "in positivo", quanto ai requisiti che questi sistemi, tanto più i modelli di IA per finalità generali, dovrebbero soddisfare per assecondare e promuovere lo sviluppo, l'autodeterminazione e la personalità dei minori, in ottica di *empowerment.* Per queste ragioni non ci si soffermerà sulle protezioni approntate dal GDPR, in quanto necessariamente confinate entro un approccio "negativo", volto a limitare accesso e trattamento dati rispetto a persone con età inferiore ad una certa soglia, senza poter con ciò contribuire "in positivo" alle caratteristiche che i sistemi di trattamento dati devono assumere al fine di valorizzare lo sviluppo dei minori che interagiscono con essi.

[17] Cfr. E. Fronza, C. Caruso, *Ti faresti giudicare da un algoritmo? Intervista ad Antoine Garapon*, in *Questione Giustizia*, 4, 2018, 196 ss.

[18] K. Chaika, *Filterworld. How Algorithms flattened culture,* New York, 2024.

[19] Sono fenomeni ampiamente noti e ben descritti sia da C.R. Sunstein, *#republic. Divided Democracy in the age of Social Media*, Princeton-Oxford, 2018; che da E. Pariser, *The Filter Bubble: What the Internet is Hiding from You,* Neuilly-sur-Seine, 2012.

[20] Cfr. intervista a A. Ghiglia, *Una tassa alle big tech per l'educazione digitale di tutti*, in https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9986495, 20 febbraio 2024.

[21] Regolamento UE 2022/2065.

Downloaded from www.biodiritto.org.

ISSN 2284-4503

L'AI Act si dichiara consapevole delle specifiche vulnerabilità dei *children*[22], sicchè nel proibire pratiche in grado di limitare l'autodeterminazione e distorcere il comportamento, con tecniche subliminali o comunque intenzionalmente manipolative, suscettibili di causare un danno fisico o piscologico significativo (art. 5, par. 1, lett. a), tratta separatamente i casi in cui tali tecniche siano specificamente volte a sfruttare particolari situazioni di fragilità, tra l'altro per ragioni legate al fattore età (art. 5, par. 1, lett. b)[23]. Proibisce inoltre l'uso da parte degli istituti scolastici di sistemi di IA di riconoscimento delle emozioni degli studenti (art. 5, par. 1, lett. f). Qualifica come ad alto rischio l'impiego di pratiche di IA volte ad incidere sull'ammissione all'istruzione e formazione professionale, o destinate a valutare la preparazione degli studenti, ovvero monitorarne il comportamento scolastico (Allegato 3, par. 3). Nell'aggiornare l'elenco dei sistemi ad alto rischio, la Commissione è chiamata (art. 7, par. 2, lett. h) a tenere conto della situazione di particolare vulnerabilità dell'utilizzatore del sistema di IA, dovuta, tra l'altro, all' età; così come sono chiamati a tenerne conto i fornitori nel momento in cui predispongono il *risk management system* (art. 9, par. 9). Quando i sistemi di IA ad alto rischio sono testati in condizioni reali, ma al di fuori di ambienti controllati quali le *regulatory sandboxes*, i *providers* devono in ogni caso assicurare l'implementazione di adeguate protezioni per i minori (art. 60, par. 4, lett. g). Sempre seguendo l'indirizzo di una protezione rafforzata per i minori, l'AI Act sottolinea come i sistemi di riconoscimento biometrico remoto in tempo reale in spazi accessibili al pubblico, laddove consentiti, possano risultare portatori di *bias* dovuti, tra l'altro, all'età del soggetto[24], ai quali deve, quindi, essere prestata la dovuta attenzione.

Particolare significato, nell'ambito "relazionale", assumono poi i *large language models* e più in generale l'IA generativa, vista la capacità di interazione. I minori risultano senz'altro maggiormente esposti ai casi di generazione di contenuti sintetici, non distinguibili da quelli autentici (c.d. deep fake). Al riguardo l'AI Act si limita a sancire un generale obbligo di trasparenza volto a rendere evidente l'origine artificiale e non umana del contenuto cui si è posti di fronte (Art. 50, parr. 2 e 4). In merito, si dovranno attendere le specifiche dei codici di condotta e degli *implementing acts* della Commissione europea per capire in che termini i "marchi di riconoscibilità" verranno calibrati in base alle capacità di discernimento legate all'età[25]. Inoltre, rispetto ai requisiti di trasparenza ed informazione, volti a rendere nota l'interazione con un sistema artificiale o l'attivazione un sistema di riconoscimento delle emozioni (Art. 50, parr. 1 e 3), l'AI Act non si premura di prescrivere specificità relative allo *status* di minori. Neppure gradua la pericolosità dei sistemi di riconoscimento delle emozioni distinguendola a seconda che siano rivolti ad adulti o minori in generale (salvo lo specifico divieto rivolto dall'art. 5, par. 1, lett. f), all'ambito scolastico o lavorativo).

[22] Cfr. Considerando n. 48 dell'AI Act: «i minori godono di diritti specifici sanciti dall'articolo 24 della Carta e dalla Convenzione delle Nazioni Unite sui diritti dell'infanzia e dell'adolescenza, ulteriormente sviluppati nell'osservazione generale n. 25 della Convenzione delle Nazioni Unite dell'infanzia e dell'adolescenza per quanto riguarda l'ambiente digitale, che prevedono la necessità di tenere conto delle loro vulnerabilità e di fornire la protezione e l'assistenza necessarie al loro benessere».

[23] Come sottolinea il considerando n. 29 dell'AI Act, «i sistemi di IA possono inoltre sfruttare in altro modo le vulnerabilità di una persona o di uno specifico gruppo di persone dovute all'età, … che potrebbe rendere tali persone più vulnerabili allo sfruttamento».

[24] Considerando nn. 32, 54 e art. 5, par. 1, (b), AI Act.

[25] Cfr. AI Act, considerando 133, 134 e 135, nonché l'art. 50, par. 7, AI Act.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Quanto al *Digital Services Act*[26], è disposto il divieto, per le piattaforme online, di mostrare pubblicità basata sulla profilazione dell'utilizzatore del servizio quando siano consapevoli con ragionevole certezza che si tratti di un minore[27]. Al di là di questa proibizione, espressamente e specificamente rivolta alla tutela dei minori, le ulteriori previsioni, seguendo un approccio *risk-based*, non introducono dettagli prescrittivi, quanto piuttosto "direttive, indirizzi" in vista di finalità che unicamente i considerando qualificano e calibrano con particolare riguardo ai minori di età. A tal riguardo, risulta necessario che le piattaforme progettino le loro interfacce in modo da garantire il massimo livello di *privacy*, sicurezza e protezione dei minori per impostazione predefinita[28] e, per converso, risulta proibita la progettazione di interfacce online che sfruttano intenzionalmente o involontariamente le debolezze e l'inesperienza dei minori, o che possano causare comportamenti di dipendenza[29]. Nel valutare i rischi per i diritti dei minori, il DSA impone ai prestatori di piattaforme online di dimensioni molto grandi e di motori di ricerca online di dimensioni molto grandi di prendere in considerazione alcuni "indicatori"[30]. Ogni rischio che possa produrre ripercussioni negative, effettive o prevedibili, sui minori, viene qualificato come rischio sistemico, con la conseguenza di far scattare a carico del gestore della piattaforma tutta una serie di obblighi ed adempimenti aggiuntivi[31]. Le norme tecniche devono essere ponderate in ragione della tutela dei minori, così come i codici di condotta[32]. Infine, quale regola di chiusura, le misure che il fornitore di piattaforme online di dimensioni molto grandi e di motori di ricerca online di dimensioni molto grandi è chiamato ad adottare per la tutela del minore devono in ogni caso ispirarsi all'interesse superiore di quest'ultimo[33]. Si versa quindi in ipotesi di previsioni generali, forse un po' troppo superficiali nel bilanciare i rimedi previsti rispetto alla problematica dei *dark patterns* con specifico riguardo alla portata dei rischi per i minori[34] e, in ogni caso, ampiamente delegate alla autoregolamentazione.

---

[26] Regolamento (UE) 2022/2065 relativo ad un mercato unico dei servizi digitali e che modifica la direttiva 2000/31/CE.

[27] Considerando 71 e art. 28, par. 2.

[28] Cfr. Considerando n. 71. In merito va rammentato che l'AGCOM, in data 6 marzo 2024, facendo riferimento agli artt. 28 e 35 del DSA, ha avviato una «consultazione pubblica per l'approvazione di un provvedimento che disciplina le modalità tecniche e di processo che i soggetti individuati dalla norma [piattaforme *online*] sono tenuti ad adottare per l'accertamento della maggiore età degli utenti», cfr. delibera n. 61/24/CONS.

[29] Considerando nn. 67, 81 e art. 28, par. 1.

[30] Il considerando 81 esemplifica alcuni "indicatori": «quanto sia facile per i minori comprendere la progettazione e il funzionamento del servizio, come possano essere esposti tramite il servizio a contenuti suscettibili nuocere alla loro salute o al loro sviluppo fisico, mentale e morale».

[31] Considerando n. 83 e n. 89 ; art. 34, par. 1, b), d); art. 35, par. 1, j).

[32] Artt. 44, par. 1, j) e 45, par. 1.

[33] Considerando n. 89 e artt. 28; 35, par. 1, j); 44, par. 1, j).

[34] I *dark patterns* sono spesso utilizzati nei *games* per i minori. La dottrina parla in merito di «behavioural design», ossia «design choices that make the gamer do something that is in the game company's interest, usually because it allows them to make money, but not necessarily something the gamer wants to do themself or something that is in their best interests… Gaming, as a form of play, can make an important contribution to the well-being and development of children. This is not necessarily the case with games that are driven by behavioural design with a view to profit maximization. In fact, such commercial practices are usually not in the best interests of the child and, moreover, may interfere with, or even violate, other children's rights, such as their rights to health and protection against economic exploitation»: S. VAN DER HOF, S.R. VAN DER HILTEN, S.L. OUBURG, M.V. BIRK, A.J. VAN ROOIJ, *Don't gamble with children's rights": how behavioral design impacts the right of children to a playful and healthy*

Downloaded from www.biodiritto.org.
ISSN 2284-4503

3. IA come "assistive technology"?

Sulla scorta delle menzionate previsioni, sembra possa essere invocato un passo ulteriore, volto a prendere meglio in considerazione e mettere meglio a fuoco la portata degli impatti dei sistemi di intelligenza artificiale destinati ad interagire con i minori di età, facendo distinzione a seconda degli ambiti, degli usi, delle finalità delle modalità, della tipologia e della intensità di tale interazione. Servirebbe quindi un intervento addizionale, più specifico in quanto maggiormente approfondito, da realizzarsi con l'apporto multidisciplinare di psicologi[35], pedagogisti, sociologi e neuroscienziati[36], insomma con l'apporto di coloro che conoscono nel dettaglio la conformazione, le funzionalità e lo sviluppo dei processi cognitivi umani. Tanto più necessario quanto più vanno diffondendosi studi che pongono in evidenza gli effetti negativi sulla salute mentale dei minori prodotti dalla crescente interazione con "sistemi artificiali"; interazione virtuale che va a scapito di quella reale, volta a favorire lo sviluppo di capacità sociali e relazionali, componenti essenziali per la formazione ed evoluzione psico-fisica[37].

Un intervento quindi non "general purpose", secondo l'approccio seguito dal legislatore europeo con l'AI Act e il DSA, ma "purpose specific", targettizzato alle specifiche vulnerabilità dei minori ed ai particolari contesti di utilizzo.

Non si tratta solo di sicurezza, di protezione dei dati personali dei minori, di protezione dei medesimi da discriminazioni, da contenuti non adeguati o illeciti, nonché di tutela della loro più facile manipolabilità e suggestionabilità. Tutti aspetti, questi, già in sé esecrabili, ed infatti già considerati dalla normativa esistente.

Si tratta altresì di preservare e promuovere la naturale quanto essenziale "relazionalità" attraverso cui i minori vanno sviluppando identità e personalità. Di prendere la questione a monte, alla radice, quale condizione prioritaria. Si tratta quindi di approfondire anzitutto quali effetti sullo sviluppo psichico e psico-fisico può avere l'interazione del minore con sistemi di IA, spesso utilizzati come "palliativo" per colmare "assenze" o "carenze" umane, o comunque per integrarle[38]. Nel momento in cui la dimensione

game environment, in *Frontiers In Digital Health*, 4, 2022, 2. Peraltro, il 24 febbraio 2023, il Comitato Europeo per la Protezione dei Dati personali ha emesso linee guida volte al riconoscimento dei dark patterns, cfr. *https://www.edpb.europa.eu/our-work-tools/documents/public-consultations/2022/guidelines-32022-dark-patterns-social-media_en* (ultima consultazione 14/06/2024). In merito, va altresì rammentato che la città di New York ha fatto causa ad alcuni gestori di social media, nel febbraio 2024, accusandoli di danni alla salute e alla sicurezza pubblica, all'ordine sociale e al benessere psicofisico della cittadinanza, derivanti dall'uso di piattaforme deliberatamente progettate con il fine di capitalizzare la vulnerabilità di bambini e adolescenti, attraverso l'inserimento nei loro algoritmi di funzionalità destinate a generare uso compulsivo e dipendenza, omettendo peraltro di informare adeguatamente gli utenti (cfr. L. DAFFARRA, *Social, nuovi danni sui minori: nuove tutela allo studio,* 7 maggio 2024, in *https://www.agendadigitale.eu/cultura-digitale/i-danni-dei-social-sui-minori-un-problema-globale-che-esige-scelte-immediate/* ultima consultazione).

[35] Recenti tecniche psicometriche si propongono di misurare l'impatto dell'IA sulla sfera neuro-cognitiva, cfr. S. DI PLINIO, *Navigare il sé digitale: implicazioni delle neurotecnologie sull'autonomia cognitiva e la consapevolezza dei diritti,* in *federalismi.it,* 6, 2024, 111.

[36] U. MÄKI, A. WALSH, M. FERNÁNDEZ PINTO, *Scientific Imperialism: Exploring the Boundaries of Interdisciplinarity*, Routledge Studies in Science, Technology and Society, vol. 38, Londra, 2018.

[37] J. HAIDT, *The Anxious Generation: How the Great Rewiring of Childhood Is Causing an Epidemic of Mental Illness,* Penguin Press, 2024.

[38] Come sottolinea la *UN Convention on the Rights of the Child - General comment No. 25 (2021) on children's rights in relation to the digital environment*, punto 15: «States parties should pay specific attention to the effects

Downloaded from www.biodiritto.org.
ISSN 2284-4503

relazionale tipica della società ha ormai subito un'irreversibile integrazione con componenti "non umane", diventa necessario creare i presupposti affinché queste componenti siano funzionali alla piena realizzazione della persona, per un complessivo benessere e sviluppo psico-fisico sin dalla giovane età.

Avanzare simile richiesta sottende certo dei rischi. *In primis*, quello di cadere in posizioni paternalistiche[39]. Va pertanto trovato un punto di equilibrio, in ottica precauzionale, tra necessità, proporzionalità ed adeguatezza, rammentando che l'ordinamento è costellato da una pluralità di "soggetti deboli", rispetto ai quali ha diversamente modellato i propri interventi[40]. Tra di essi rientrano senz'altro i minori di età, dei quali va preservata non solo l'integrità psico-fisica in senso materiale, come ad esempio fatto dall'Unione Europea con la Direttiva sui giochi ed i relativi requisiti di sicurezza[41], ma altresì gli aspetti meno tangibili relativi alla loro socialità ed al conseguente sviluppo soggettivo, attraverso autodeterminazione, personalità e strutturarsi dell'identità[42].

Tale richiesta, risulta peraltro in linea con l'impostazione seguita dall'Organizzazione Mondiale della sanità che qualifica la salute come bene multidimensionale, non limitata all'assenza di malattia psico-fisica, ma estesa al pieno esplicarsi della persona nella società[43].

Se così è come pare, forse, i sistemi di intelligenza artificiale che interagiscono con i minori di età, quindi soggetti che non hanno ancora raggiunto la maturità cognitiva ed emotiva, andrebbero regolamentati come "assistive technologies" in quanto si traducono, in definitiva, in sistemi che integrano, completano e supportano processi cognitivi ancora "fragili"; sopperendo in sostanza, come fanno tutte le *assistive technologies,* ad *impairments.* In merito, giova rammentare che non tutte le "assistive technologies" sono *medical devices*, ai sensi e per gli effetti di cui all'art. 2 del Regolamento UE 2017/745 e che, nel caso di specie, la natura "assistive" dei sistemi di IA si giustificherebbe in ragione della vulnerabilità che specificamente qualifica i minori.

---

of technology in the earliest years of life, when brain plasticity is maximal and the social environment, in particular relationships with parents and caregivers, is crucial to shaping children's cognitive, emotional and social development. In the early years, precautions may be required, depending on the design, purpose and uses of technologies…, taking into account the research on the effects of digital technologies on children's development, especially during the critical neurological growth spurts of early childhood and adolescence»; ed ancora (punti 109-110): «Especially in their early years, children acquire language, coordination, social skills and emotional intelligence largely through play that involves physical movement and direct face-to-face interaction with other people. For older children, play and recreation that involve physical activities, team sports and other outdoor recreational activities can provide health benefits, as well as functional and social skills. Leisure time spent in the digital environment may expose children to risks of harm».

[39] Su «l'abbraccio soffocante» del diritto rispetto alla tutela dei soggetti deboli, cfr. M. AINIS, *I soggetti deboli nella giurisprudenza costituzionale,* in *Politica del Diritto,* 1, 1999, 52.

[40] D. POLETTI, *Soggetti deboli,* in *Enciclopedia del Diritto,* Annali VII, 2014, 962 ss.

[41] Cfr. art. 10 nonché l'allegato II, Direttiva 2009/48/EC del Parlamento europeo e del Consiglio del 18 giugno 2009 sulla sicurezza dei giocattoli.

[42] Sottolinea la necessità di una protezione ulteriore per i minori nel cyberspazio, in ragione della loro ontologica fragilità, N. BILIGOTTI, *La tutela dei minori nel cyberspazio. Parental Control di Stato e libera circolazione dei contenuti: un delicato equilibrio,* in *Media Laws,* 1, 2023, 35.

[43] Cfr. la Costituzione dell'OMS del 22 giugno 1946: «Health is a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity».

Downloaded from www.biodiritto.org.
ISSN 2284-4503

È forse il caso che l'Unione Europea, preso atto dello stretto rapporto che si va sempre più intessendo tra *assistive technologies* nel loro significato originario (destinate a soggetti affetti da *impairments*) e sistemi di intelligenza artificiale in generale, preso altresì atto di una lacuna normativa riguardante gli strumenti suscettibili di essere qualificati come "assistive technologies", introduca un'apposita disciplina avendo specifico riguardo ai minori *tout court* (che, in modo "atecnico", possono essere qualificati come affetti da *impairments,* sia pure di natura non patologica, ma "fisiologica", legati alla non ancora raggiunta maturità psico-fisica)[44].
Elevare i sistemi di IA destinati ad interagire con i minori di età ad "assistive technologies" significa specificare già a livello normativo requisiti più dettagliati (rispetto a quanto fatto dal DSA e dall'AI Act) proprio con riguardo agli aspetti di *risk management*, per assicurare che – *by design* – tali sistemi siano volti a promuovere al meglio la relazionalità, lo sviluppo delle capacità di libera autodeterminazione, della identità e della personalità: tappe essenziali del percorso formativo ed educativo del minore volte ad assecondare il suo *empowerment*.

4. Famiglia e scuola in azione: AI for children empowerment

Certo è che, alla luce della profonda relazionalità dell'essere umano, le evocate (§§ 2,3) previsioni ordinamentali non bastano, dovendo trovare naturale completamento nel ruolo tradizionalmente assegnato dalla nostra Costituzione a famiglia e scuola[45]: "comunità" in cui il minore compie i primi passi dello sviluppo psico-sociale, maturando la propria identità e personalità[46]. Istruzione ed educazione indicano quindi la progressività di un percorso lungo il quale il minore è accompagnato, come singolo e come essere sociale, per poter apprendere a muoversi in modo consapevole e autonomo[47].
Questa tradizionale "rete" relazionale (familiare e scolastica), si trova ora a giocare la partita a fianco di sistemi di intelligenza artificiale che sono in grado di decifrare preferenze e interessi, componenti essenziali della personalità del minore[48], di avanzare su tali basi suggerimenti, suggestioni, risposte[49]. Tali sistemi di IA hanno pertanto trasformato in un quadrilatero la relazione dialogica che in precedenza era "triangolare" (tra minore, famiglia e scuola), aggiungendo un lato di natura "sintetica"[50]. Hanno quindi penetrato ambiti, quali l'educare e l'istruire, che non solo presuppongono delicate

[44] Cfr. *L'in-depth analysis* sulle *assistive technologies* per le persone con disabilità elaborata *dal European Parliamentary Research Service* del gennaio 2018, ref. PE 603.218.
[45] In quanto persone in fase di formazione, i minori devono essere indirizzati, guidati nel loro percorso di crescita verso la completa autonomia con conseguente ruolo di vigilanza e controllo anzitutto da parte della famiglia, ma anche della scuola, cfr. G. MATUCCI, *La responsabilità educativa dei genitori fra scuola e dinamiche familiari,* in G. MATUCCI, F. RIGANO (a cura di), *Costituzione e istruzione,* Milano, 2016, 237.
[46] Come esplicitato da E.C. RAFFIOTTA, M. BARONI, *Intelligenza artificiale, strumenti di identificazione e tutela dell'identità*, in A: PAJNO, F: DONATI, A. PERRUCCI (a cura di), *Intelligenza Artificiale e diritto: una rivoluzione?,* vol. 1, Bologna, 2022, 364: «l'identità può indicarsi come elemento caratterizzante della persona, parte integrante – quasi colonna portante – di quest'ultima, strumentale al pieno sviluppo dell'individuo».
[47] In tali termini, G. MATUCCI, *La responsabilità educativa dei genitori fra scuola e dinamiche familiari,* cit., 234.
[48] Cfr. D.G. RUGGIERO, *Persona e identità digitale,* Napoli, 2023, 79 ss.
[49] C. PERLINGERI, *La tutela dei minori di età nei* social networks, in *Rassegna di diritto civile,* 4, 2016, 1329.
[50] Per i rischi e le opportunità delle ITs in ambito educativo e culturale, cfr. G. SARTOR, *Human Rights and Information Technologies,* in R. BROWNSWORD, E. SCOTFORD, K. YEUNG (a cura di), *The Oxford Handbook of Law, Regulation and Technology,* Oxford University Press, Oxford, 2017, 426.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

conoscenze e competenze di natura psicologica, sociologica e pedagogica, ma presuppongono altresì, dal punto di vista giuridico, l'assolvimento di un vero e proprio *officium,* tradizionalmente affidato a famiglia e scuola[51]: un dovere, finalizzato non solo alla protezione ma anche, in chiave positiva, alla promozione della personalità dei minori. Dovere destinato, in ultima analisi, a realizzare quella funzione culturale e democratica di «seminarium rei pubblicae» menzionata, in tempi non sospetti, da Calamandrei[52].

Ora più di prima diventa essenziale che queste comunità siano anzitutto in grado di assolvere a tale *officium,* supportando il minore nella strutturazione di una personalità attrezzata per autodeterminarsi in modo libero e critico. Dotandolo quindi di una basilare *soft skill*, volta alla «liberazione dell'intelligenza»[53] (umana) dalle "trappole" dell'intelligenza artificiale, affinché sia promossa e valorizzata al massimo la libertà di pensiero, e recuperata la kantiana dignità umana, intesa come capacità di scelte morali.

Affinché dunque la tecnologia (nella specie l'IA) possa interagire ed integrare l'intelligenza umana nel senso positivo teorizzato dai sostenitori del concetto di «mente estesa»[54], in una «profonda continuità tra umano e tecnologico»[55], è necessario che il primo (l'umano), sin dalla giovane età, sia adeguatamente "armato" per meglio esercitare le proprie capacità di discernimento[56].

Il ruolo della famiglia e della scuola diventa qui essenziale e va giocato tra la "protezione da" e la "promozione" della libera autodeterminazione, verso un equilibrio il più possibile funzionale all'*empowerment* del minore di età a fronte delle sfide poste da una tecnologia tanto complessa. Se è vero che spetta all' «individuo, con la sua autonomia e con i suoi connessi poteri di autodeterminazione… riempire di senso la formula astratta del vivere e dell'agire *cum dignitate*», è parimenti vero che proprio perché l'individuo, ed il minore *in primis*, è *homme situé*[57] (i.e. situato anzitutto nella formazione sociale familiare e scolastica), spetta a quest'ultime, a fronte delle fragilità tipiche della minore età, adempiere al dovere di adeguatamente indirizzarlo nella strutturazione e maturazione di tale capacità di autodeterminazione, fondamento di una piena dignità umana[58].

---

[51] S. Sileoni, *L'autodeterminazione del minore tra tutela della famiglia e tutela dalla famiglia***,** in *Quaderni costituzionali*, 3, 2014, 621 e 634.

[52] P. Calamandrei, Discorso pronunciato al III Congresso dell'Associazione a difesa della scuola nazionale (ADSN), Roma 11 febbraio 1950, pubblicato in *Scuola democratica*, suppl. al n. 2 del 20 marzo 1950, 1-5.

[53] F. Angelini, *«Generazione di adulti» e «generazioni di giovani» fra famiglia e scuola. valori, diritti e conflitti nel rapporto educativo*, in *costituzionalismo.it,* 3, 2021, 4.

[54] A. Clark, D. Chalmers, *The Extended Mind,* in *Analysis,* vol. 58, n. 1/1998, 8, osservano «In these cases, the human organism is linked with an external entity in a two-way interaction, creating a coupled system that can be seen as a cognitive system in its own right. All the components in the system play an active causal role, and they jointly govern behaviour in the same sort of way that cognition usually does».

[55] V. Gheno, B. Mastroianni, *Tienilo acceso,* Milano, 2018, 20.

[56] S. Tiribelli, *La dimensione etica e politica dell'algoritmo,* in A. Sterpa (a cura di), *L'ordine giuridico dell'algoritmo,* 33 ss.

[57] G. D'Amico, *La nascita del biodiritto come prodotto della costituzionalizzazione dell'ordinamento,* in *BioLaw Journal,* 2, 2019, 176-177.

[58] G. Sartor, *op.cit.*, 436 e 439, che rammenta lo stretto collegamento tra libertà di scelta morale, di *agency* fattuale, dignità come *empowerment* e lo «human flourishing in the IT-based society».

Downloaded from www.biodiritto.org.
ISSN 2284-4503

In questo percorso, viene in soccorso il criterio dei *best interests of the child* che, non a caso, nell'evoluzione seguita nell'ambiente nordamericano, ha oscillato tra autodeterminazione e protezione[59]. E, quali "indicatori" interni al nostro ordinamento, vengono in soccorso alcune pronunce della Corte costituzionale, "riflesso giuridico" di considerazioni di natura psicologica, pedagogica e sociologica.

Anzitutto, la sentenza che ha riconosciuto fondamentale valore costituzionale alla «sfera intima della coscienza individuale… [quale] riflesso giuridico più profondo dell'idea universale della dignità della persona umana… relazione intima e privilegiata dell'uomo con se stesso», dotata di « rilievo costituzionale quale principio creativo che rende possibile la realtà delle libertà fondamentali dell'uomo e quale regno delle virtualità di espressione dei diritti inviolabili del singolo nella vita di relazione»[60]. Quindi, la giurisprudenza costituzionale sulla identità personale, quale «bene per sé medesima» e diritto a che la propria «individualità sia preservata»[61] e possa, tanto più per i minori, connotarsi e sostanziarsi principalmente nell'ambito «relazioni di tipo socio-affettivo»[62]. Relazioni, quest'ultime, qualificate da «finalità di educazione e formazione», volte «a favorire l'espressione delle potenzialità cognitive, affettive e relazionali del bambino»[63]. Relazioni, tanto importanti al punto da integrare una concezione olistica del "bene fondamentale salute" che, secondo la giurisprudenza costituzionale va riferito «a tutte le attività, le situazioni e i rapporti in cui la persona esplica sé stessa nella propria vita: non soltanto, quindi, con riferimento alla sfera produttiva, ma anche con riferimento alla sfera spirituale, culturale, affettiva, sociale, sportiva e ad ogni altro ambito e modo in cui il soggetto svolge la sua personalità»[64].

L'obiettivo che fa da minimo comune denominatore a tali pronunce è volto a preservare la fondamentale relazionalità del minore di età non solo attraverso la "protezione da", ma anche per un suo *empowerment*, quindi per la costruzione di un "soggetto forte", nel determinarsi liberamente e pienamente[65].

Ciò è dovere della famiglia, ma anche della scuola, viste le professionalità che la compongono. Come già chiariva Pototschnig, l'istruzione non è indifferente ai contenuti di cui si nutre: «ciò richiede che ogni istituto o organizzazione scolastica si regga su strutture capaci di cogliere e interpretare prontamente e adeguatamente le esigenze e le attese della società civile in cui opera»[66]. Una scuola quindi "al passo con i tempi", alla quale la Convenzione delle Nazioni Unite sui diritti del bambino chiede l'insegnamento della «digital literacy» come parte dei «basic education curricula», sin dal «preschool level»[67]. Nello specifico, una "AI literacy" non limitata all'apprendimento delle modalità di utilizzo degli

[59] Per una ricostruzione dell'evoluzione del principio dei *best interests of the child* in ambiente nordamericano, tra, da un lato, *autonomy* e *self-determination* e, dall'altro, *protection, salvation* e *nurturance orientation*, cfr. E. LAMARQUE, *Diritti fondaemntali della persona minore di età e* best interests of the child, in *giustiziainsieme.it,* 6 febbraio 2023.

[60] Corte cost. 16 dicembre 1991, n. 467, punto 4 del considerato in diritto.

[61] Corte cost. 3 febbraio 1994, n. 13, punto 5.1 del considerato in diritto.

[62] Corte cost. 5 luglio 2023, n. 183.

[63] Corte cost. 20 dicembre 2002, n. 467, punto 2 del considerato in diritto.

[64] Corte cost. 18 luglio 1991, n. 356, punto 8 del considerato in diritto.

[65] D. POLETTI, *Soggetti deboli,* cit.

[66] U. POTOTSCHNIG, *Istruzione (diritto alla)*, in *Enciclopedia del Diritto,* vol. XXIII, Milano, 1973, 114.

[67] La *UN Convention on the Rights of the Child - General comment No. 25 (2021) on children's rights in relation to the digital environment*, precisa, infatti, ai punti 104-105, che i «Curricula should include the knowledge and skills

Downloaded from www.biodiritto.org.
ISSN 2284-4503

strumenti digitali, ma anche alla comprensione dei rischi implicati, benché di non immediata percezione[68]. In questa direzione *l'European Digital Education Hub* della Commissione Europea, nel Report 2024 sull'IA[69], prevede un «teaching for AI», «teaching with AI» nonché un «teaching about AI»[70].

5. Conclusioni

Sulla base di quado sinora si è cercato di argomentare, sembra necessario un intervento ulteriore, con specifico riguardo ai minori di età, rispetto a quanto previsto dal DSA e dall'AI Act. Un intervento che, sulla scorta di un apporto conoscitivo multidisciplinare, sia in grado di meglio definire i requisiti normativi dei sistemi di IA che interagiscono con i minori qualificandoli quali *assistive technologies,* in ragione dell'essenziale funzione che ormai vengono ad assolvere nell'accompagnare e plasmare, a fianco dei tradizionali ambiti sociali (famiglia e scuola), lo sviluppo dell'autodeterminazione e della personalità del minore di età (come del resto risulta sempre più suffragato da evidenze scientifiche). Sia pure assecondando una prospettiva di impostazione predefinita, di incorporazione delle regole nell'architettura della tecnologia, tipicamente definita *by design,* le prescrizioni normative necessitano di entrare meglio nel dettaglio, riducendo così il margine di discrezionalità lasciato ai "signori dell'algoritmo"[71], laddove – l'AI Act e il DSA – rinviano invece all'autoregolamentazione, ai codici di condotta ed alle specifiche tecniche. Così come è necessario che l'intervento normativo sia realizzato non solo in direzione "protettiva" (come fatto da AI Act e DSA) ma anche in ottica "positiva", funzionalizzata all'*empowerment* dei minori nella loro interazione con i sistemi di IA nell'ambiente relazionale di cui sono ormai parte. E ciò può realizzarsi unicamente con interventi normativi "targettizzati" alle peculiarità del percorso di sviluppo dei minori, tenendo conto dei loro "impairments", come tradizionalmente fatto in modo generale dalle "assistive technologies".

Al di là di questa rete di prescrizioni ordinamentali, rimane parimenti essenziale il ruolo assolto dalle "comunità" tradizionali (famiglia e scuola) che, mettendo a frutto le "coordinate costituzionali" entro cui si inscrive la dignità umana e lo sviluppo della personalità, si trovano a dover operare in sinergia perché i sistemi di IA fungano da opportunità ed *empowerment* per il minore. Tali "coordinate costituzionali" destinate a far da bussola a famiglia e scuola, emergono dalle accennate pronunce della Corte (§ 4), volte a connotare giuridicamente i bisogni sociali dell'individuo, sin dalla nascita, quali basi giuridiche per lo sviluppo dell'identità, dell'autodeterminazione e della personalità dei minori di età, in

to safely handle a wide range of digital tools and resources, including those relating to content, creation, collaboration, participation, socialization and civic engagement. It is of increasing importance that children gain an understanding of the digital environment, including its infrastructure, business practices, persuasive strategies and the uses of automated processing and personal data and surveillance, and of the possible negative effects of digitalization on societies».

[68] Come sottolinea G. PEDRAZZI, *Minori e social media: tutela dei dati personali, autoregolamentazione e privacy in Informatica e Diritto*, XVI, 2017, 438, risulta fallace «l'equazione che traduce la padronanza dello strumento con la consapevolezza delle conseguenze delle azioni connesse e le contestabili categorizzazioni dei "nativi digitali" o "millenials" ha talvolta condotto ad una sottovalutazione dei pericoli e dei rischi connessi all'uso degli strumenti».

[69] Disponibile su: *file:///C:/Users/Utente/Downloads/ai%20report-EC0623043ENN.pdf*.

[70] *Ibidem*, 9, 13 e 17.

[71] L. AMMANNATI, *I "signori" nell'era dell'algoritmo,* in *Diritto Pubblico,* 2, 2021, 381 ss.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

modo da realizzare pienamente il valore della dignità umana che trova, nel principio personalista, e quindi in un approccio umano-centrico, le sue radici. Famiglia e scuola sono pertanto protagoniste di questo umano-centrismo, chiamate a mantenere lo *human oversight* affinché l'utilizzo dei sistemi di IA da parte dei minori non sia un sostitutivo della essenziale relazionalità sociale, rendendolo semmai un complemento per l'*empowerment* in termini di autodeterminazione e sviluppo della persona. Famiglia e scuola condividono quindi un'essenziale responsabilità volta ad attualizzare il significato e la portata delle coordinate costituzionali che contribuiscono a definire il «ruolo che la società odierna e quella futura vorranno mantenere per gli esseri umani [v]ista la forte interdipendenza e complementarietà fra ambito dell'intelligenza artificiale e ambito dell'intelligenza umana»[72].

Downloaded from www.biodiritto.org.
ISSN 2284-4503

[72] C. CASONATO, *Potenzialità e sfide dell'intelligenza artificiale*, in *BioLaw Journal*, 1, 2019, 181.

I minori sulla rete: un problema di natura costituzionale

Bianca Pileggi*

CHILDREN ON THE INTERNET: A CONSTITUTIONAL PROBLEM

ABSTRACT: The article addresses the complex issue of protecting minors in the digital age, highlighting its constitutional implications. It examines how the Digital Services Act (DSA) has revitalized the issue of safeguarding children's rights online and how proceedings have been initiated against some digital giants for allegedly being harmful to minors. The discussion includes the historical context of digital maturity laws, comparing the Children's Online Privacy Protection Act (COPPA) with the European General Data Protection Regulation (GDPR). It concludes by invoking the doctrine of "constitutional precaution", which argues that the protection of minors' fundamental rights should be guaranteed in the design of new technologies, emphasizing the need to balance protection measures with the safeguarding of young users' rights.

KEYWORDS: Children protection; vulnerability; Digital Services Act; age verification systems; Constitutional Law.

ABSTRACT: L'articolo affronta la complessa questione della protezione dei minori nell'era digitale, sottolineandone le implicazioni costituzionali. Esamina come il Digital Services Act (DSA) abbia dato nuova linfa al tema della salvaguardia dei diritti dei bambini *online* e come siano stati avviati procedimenti contro i alcuni colossi del digitale proprio con l'accusa di essere dannosi per i minori. Viene inoltre delineato il contesto storico delle leggi sulla maturità digitale, confrontando il *Children's Online Privacy Protection Act* (*COPPA*) con il Regolamento Generale sulla Protezione dei Dati (*GDPR*) europeo. Si conclude facendo appello alla dottrina della "precauzione costituzionale", affinché la protezione dei diritti fondamentali dei minori possa essere garantita nella progettazione di nuove tecnologie, sottolineando la necessità di operare un bilanciamento tra istanze di tutela e garanzia dei diritti dei giovani utenti.

PAROLE CHIAVE: Protezione dei minori; vulnerabilità; Digital Services Act; sistemi di verifica dell'età; diritto costituzionale.

SOMMARIO: 1. Il caso: la Commissione europea apre i primi procedimenti ai sensi del DSA contro TikTok e Meta per violazione delle norme a tutela dei minori – 2. L'accesso a Internet dei minori: ovvero come i tredici anni sono diventati l'età della "maturità digitale" – 2.1. Il *Children's Online Privacy Protection Ac*t del 1998 – 2.2. Dal COPPA al GDPR – 3. Uno sguardo comparato: il contenzioso negli Stati Uniti – 4. Oltre la *privacy*: gli "*age verification*

* *Dottoranda di Diritto costituzionale, Università di Firenze. Mail: bianca.pileggi@unifi.it.* *Contributo sottoposto a doppio referaggio anonimo.*

Downloaded from www.biodiritto.org.
ISSN 2284-4503

system" – 5. I minori sulla rete: un problema di "precauzione" costituzionale.

1. Il caso: la Commissione europea apre i primi procedimenti ai sensi del DSA contro TikTok e Meta per violazione delle norme a tutela dei minori

A partire da agosto 2023 il *Digital Services Act*[1], il regolamento europeo sui servizi digitali, si applica alle piattaforme designate con oltre 45 milioni di utenti nell'UE[2] (il 10% della popolazione europea), vale a dire le piattaforme (VLOP) o i motori di ricerca *online* (VLOSE) di dimensioni molto grandi[3].

Con il nuovo regolamento sui servizi digitali il legislatore europeo ha dimostrato sensibilità al tema della tutela dei minori[4] nel contesto digitale prevedendo espressamente obblighi di protezione nei confronti degli utenti più giovani da parte dei fornitori di servizi digitali[5], nonché imponendo alle piattaforme e ai motori di ricerca *online* di dimensioni molto grandi obblighi più stringenti in relazione alla valutazione dei rischi[6] derivanti dall'utilizzo delle piattaforme tenendo in considerazione «qualsiasi

[1] Il Digital Services Act (DSA) è il Regolamento (UE) 2022/2065 del Parlamento europeo e del Consiglio del 19 ottobre 2022 relativo a un mercato unico dei servizi digitali e che modifica la direttiva 2000/31/CE (regolamento sui servizi digitali). L'applicazione generalizzata dal DSA a tutte le piattaforme è avvenuta a partire dal 17 febbraio 2024.
Il DSA, insieme al Digital Markets Act (DMA), il Regolamento (UE) 2022/1925 del Parlamento europeo e del Consiglio del 14 settembre 2022 relativo a mercati equi e contendibili nel settore digitale e che modifica le direttive (UE) 2019/1937 e (UE) 2020/1828 (regolamento sui mercati digitali), fa parte di un pacchetto di norme che l'Unione europea ha inteso adottare al fine di raggiungere due obiettivi principali: creare uno spazio digitale più sicuro in cui siano tutelati i diritti fondamentali di tutti gli utenti dei servizi digitali; stabilire condizioni di parità per promuovere l'innovazione, la crescita e la competitività, sia nel mercato unico europeo che a livello globale. Vd. https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package (ultima consultazione 22/07/2024).

[2] Articolo 33 del DSA.

[3] Per quanto riguarda il DSA, infatti, vengono distinte le regole applicabili alle cosiddette VLOP (Very Large Online Platforms) o VLOSE (Very Large Online Search Engines), ovverosia piattaforme che abbiamo più di quarantacinque milioni di utenti, da quelle applicabili a tutti gli altri prestatori di servizi intermediari. Sulla distinzione tra VLOP e VLOSE e le altre piattaforme s.v. J. VAN HOBOKEN ET AL, *Putting the Digital Services Act Into Practice: Enforcement, Access to Justice, and Global Implications*, in *Amsterdam Law School Research Paper*, 13, 2023, disponibile in: https://verfassungsblog.de/books/; E. LONGO, *Libertà di informazione e lotta alla disinformazione nel Digital Services Act*, in *Giornale di diritto amministrativo*, 6, 2024, 737-745.

[4] Il DSA fa parte della nuova strategia europea per un *Internet migliore per i ragazzi* adottata a maggio 2022 dalla Commissione europea, la c.d. "BIK+" perché si inserisce in linea di continuità con la precedente BIK (*Better Internet for Kids*) introdotta nel 2012. Vd. Considerando 71 del DSA.

[5] L'articolo 28, paragrafo 1, del DSA recita: «I fornitori di piattaforme online accessibili ai minori adottano misure adeguate e proporzionate per garantire un elevato livello di tutela della vita privata, di sicurezza e di protezione dei minori sul loro servizi».

[6] L'approccio del legislatore europeo, seppur diversamente declinato rispetto ad altre normative (Cfr. Regulation EU n. 679/2016 (*General Data Protection Regulation*) e Regulation EU n. 1689/2024 (*Artificial Intelligence Act*)) adottate nell'ambito della *Digital Strategy* (Commissione europea, Comunicazione: Shaping Europe's digital future, 19 febbraio 2020, disponibile al seguente link: https://ec.europa.eu/info/publications/communication-shaping-europes-digital-future_it ultima consultazione 22/07/2024), si è dimostrato ancora una volta di tipo *risk based* in relazione a quelli che sono i provvedimenti che i fornitori di servizi digitali devono adottare al fine di prevenire o arginare gli effetti dannosi derivanti dal verificarsi di rischi prevedibili a priori. Anziché limitarsi a stabilire nuovi diritti e garanzie, l'Unione ha cercato di regolamentare i pericoli aumentando la responsabilità

Downloaded from www.biodiritto.org.
ISSN 2284-4503

effetto negativo, attuale o prevedibile, in relazione alla violenza di genere, alla protezione della salute pubblica *e dei minori* e alle gravi conseguenze negative per il benessere fisico e mentale della persona»[7]. Ai fornitori di servizi digitali di grandi dimensioni spetta poi anche di provvedere alla attenuazione dei rischi così individuati, in particolare attraverso «l'adozione di misure mirate per *tutelare i diritti dei minori*, compresi strumenti di verifica dell'età e di controllo parentale, o strumenti volti ad aiutare i minori a segnalare abusi o ottenere sostegno, a seconda dei casi»[8].

Il legislatore europeo con la nuova normativa ha ritenuto di operare una distinzione tra piattaforme e motori di ricerca molto grandi e piattaforme più piccole in ragione dei diversi rischi che le prime possono comportare per la società in termini di portata ed effetti[9]. Data l'importanza che le VLOP e i VLOSE hanno in ragione del loro raggio d'azione, visto l'alto numero di destinatari dei servizi, il legislatore europeo ha ritenuto necessario di imporre ai fornitori di tali piattaforme obblighi specifici, in aggiunta agli obblighi previsti dal DSA e applicabili a tutte le piattaforme *online*. Tali obblighi supplementari per i fornitori di piattaforme *online* di dimensioni molto grandi e di motori di ricerca *online* di dimensioni molto grandi sono necessari per affrontare e prevenire il verificarsi di quelli che il legislatore ha definito "rischi sistemici"[10], rimettendo alle VLOP e ai VLOSE la responsabilità di individuare, analizzare e valutare con diligenza tali rischi[11] entro quattro mesi dalla notifica della designazione come piattaforme o motori di ricerca di dimensioni molto grandi ai sensi dell'articolo 33 del DSA.

Ad aprile 2023 la Commissione europea ha individuato un primo elenco di VLOP e VLOSE[12], designando (tra le altre) come piattaforme di dimensioni molto grandi *TikTok*, *Facebook* ed *Instagram*. Pertanto, ad agosto 2023 i colossi *social* hanno dovuto fornire ai sensi dell'articolo 34 del DSA una prima valutazione dei rischi sistemici derivanti dall'utilizzo dei loro servizi. Le informazioni così fornite da *TikTok* e da *Meta* non sono state ritenute del tutto esaustive da parte della Commissione europea che ha sollecitato le *big tech* a più riprese affinché fornissero ulteriori informazioni ai sensi dell'articolo 67 del DSA. In particolare, nei mesi successivi all'applicazione del DSA, la Commissione europea aveva chiesto a

degli attori pubblici e privati rispetto ai rischi e ai potenziali effetti collaterali derivanti dalle loro attività. Vd. G. De Gregorio, P. Dunn, *The European Risk-Based Approaches: Connecting Constitutional Dots in the Digital Age*, in *Common Market Law* Review, 59, 2, 2022, 473-500.

[7] Articolo 34 del DSA. Nostro il corsivo.

[8] Articolo 35 del DSA. Nostro il corsivo.

[9] Vd. Considerando 76 del DSA.

[10] I c.d. "rischi sistemici" sono suddivisi in quattro categorie, così individuate dall'articolo 34 del DSA: «a) la diffusione di contenuti illegali tramite i loro servizi; b) eventuali effetti negativi, attuali o prevedibili, per l'esercizio dei diritti fondamentali, in particolare i diritti fondamentali alla dignità umana sancito nell'articolo 1 della Carta, al rispetto della vita privata e familiare sancito nell'articolo 7 della Carta, alla tutela dei dati personali sancito nell'articolo 8 della Carta, alla libertà di espressione e di informazione, inclusi la libertà e il pluralismo dei media, sanciti nell'articolo 11 della Carta, e alla non discriminazione sancito nell'articolo 21 della Carta, al *rispetto dei diritti del minore sancito nell'articolo 24 della Carta*, così come all'elevata tutela dei consumatori, sancito nell'articolo 38 della Carta; c) eventuali effetti negativi, attuali o prevedibili, sul dibattito civico e sui processi elettorali, nonché sulla sicurezza pubblica; d) qualsiasi effetto negativo, attuale o prevedibile, in relazione alla violenza di genere, alla *protezione della salute pubblica e dei minori* e *alle gravi conseguenze negative per il benessere fisico e mentale della persona*». Nostro il corsivo.

[11] Vd. Articolo 34 del DSA.

[12] https://ec.europa.eu/commission/presscorner/detail/en/IP_23_2413 (ultima consultazione 22/07/2024).

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Meta e a *TikTok* di fornire maggiori informazioni[13] in merito alle misure adottate per ottemperare agli obblighi di protezione dei minori ai sensi del DSA, compresi quelli relativi alla valutazione e alle misure di attenuazione dei rischi per proteggere i minori *online*, specificatamente per quanto atteneva ai rischi per la salute mentale e fisica derivanti dall'utilizzo dei servizi da parte dei più giovani.

Nonostante le plurime richieste di informazioni e i solleciti della Commissione europea affinché *TikTok*, *Instagram* e *Facebook* provvedessero ad adeguarsi alle previsioni del DSA, in particolare con riferimento agli obblighi imposti in relazione al tema della protezione dei minori e alle conseguenze dannose che l'utilizzo di tali servizi potrebbe comportare per il loro benessere fisico e mentale, la Commissione ha considerato le risposte fornite da *TikTok* e *Meta* non soddisfacenti e a seguito di un'analisi preliminare delle informazioni ottenute ha ritenuto che non si fossero correttamente adeguate al DSA[14].

Pertanto, il 19 febbraio 2024 Commissione europea ha annunciato l'apertura di un procedimento ai sensi dell'articolo 66 del DSA contro il *social network* cinese *TikTok*[15], assumendo che tra gli articoli violati dalla piattaforma vi fossero quelli posti a tutela dei minori[16]. In seguito, il 16 maggio 2024 la Commissione europea ha annunciato di aver aperto un ulteriore procedimento contro *Meta*[17] per ragioni analoghe[18] a quelle contestate alla piattaforma *TikTok*.

In entrambi i casi la Commissione ha evidenziato il rischio che i sistemi di *TikTok*, *Facebook* e *Instagram*, compresi i loro algoritmi, potessero stimolare dipendenze comportamentali nei bambini e nei ragazzi e creare il cosiddetto effetto "*rabbit hole*"[19], cioè un fenomeno che avviene quando gli utenti vengono trascinati in una serie continua di contenuti correlati, portandoli ad esplorare argomenti sempre più lontani dalla loro ricerca iniziale. La Commissione ha inoltre rilevato una possibile illegittimità dei metodi di verifica dell'età predisposti sia da parte della società cinese che da parte di *Meta* in relazione

[13] Tra le richieste di informazioni *ex* art. 67 del DSA si veda https://ec.europa.eu/commission/presscorner/detail/en/mex_23_5145; https://digital-strategy.ec.europa.eu/en/news/commission-sends-requests-information-TikTok-and-youtube-under-digital-services-act; https://digital-strategy.ec.europa.eu/en/news/commission-sends-requests-information-Meta-and-snap-under-digital-services-act); https://digital-strategy.ec.europa.eu/en/news/commission-sends-request-information-Meta-under-digital-services-act; https://digital-strategy.ec.europa.eu/en/news/commission-sends-request-information-Meta-under-digital-services-act-1; https://digital-strategy.ec.europa.eu/en/news/commission-sends-requests-information-generative-ai-risks-6-very-large-online-platforms-and-2-very; https://digital-strategy.ec.europa.eu/en/news/commission-sends-request-information-TikTok-regarding-launch-TikTok-lite-france-and-spain (ultima consultazione dei link presenti in nota 22/07/2024).

[14] Per un commento sul tema si vd. M. FABBRI, *Moderating online platforms after the DSA: from designing rules to enabling enforcement*, https://digi-con.org/moderating-online-platforms-after-the-dsa-from-designing-rules-to-enabling-enforcement/ (ultima consultazione 22/07/2024).

[15] https://ec.europa.eu/commission/presscorner/detail/en/ip_24_926 (ultima consultazione 22/07/2024).

[16] Gli articoli che, se accertata la violazione, si assumono violati sono: artt. 34(1), 34(2), 35(1), 28(1), 39(1), e 40(12) del DSA. Vd. https://ec.europa.eu/commission/presscorner/detail/en/ip_24_926 (ultima consultazione 22/07/2024).

[17] https://digital-strategy.ec.europa.eu/it/news/commission-opens-formal-proceedings-against-Meta-under-digital-services-act-related-protection (7/07/2024).

[18] Gli articoli che, se accertata la violazione, si assumono violati sono: artt. 28, 34 e 35 del DSA.

[19] K. WOOLLEY, M.A. SHARIF, *Down a Rabbit Hole: How Prior Media Consumption Shapes Subsequent Media Consumption*, in *Journal of Marketing Research*, 3, 2022, 453-471. Così come Alice cadendo nella tana del coniglio si ritrova catapultata suo malgrado nel Paese delle Meraviglie, il giovane utente dei social media si trova risucchiato all'interno di un algoritmo che gli propone contenuti disturbanti e potenzialmente pericolosi da cui difficilmente riuscirà autonomamente ad uscire.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

all'accesso ai servizi forniti da parte di colossi del digitale, nonché la mancata adozione di misure adeguate e proporzionate per garantire un elevato livello di *privacy*, sicurezza e protezione dei minori.
Entrambi i procedimenti sono attualmente ancora in corso[20]. Laddove le violazioni contestate dovessero essere accertate in tema di protezione dei minori, le società saranno chiamate a rispondere ai sensi degli articoli 28 (*Protezione online dei minori*), 34 (*Valutazione del rischio*) e 35 (*Attenuazione dei rischi*) del DSA. L'eventuale accertamento da parte della Commissione delle violazioni contestate comporterà l'applicazione delle sanzioni pecuniarie previste dall'articolo 74 del DSA.
Con queste azioni di *enforcement* del DSA sembra chiaro che la protezione dei minori sia uno dei punti nevralgici dell'attuazione del regolamento europeo. La Commissione ha sin da subito invitato le piattaforme e i motori di ricerca di grandi dimensioni a presentare le informazioni relative ai possibili rischi derivanti all'utilizzo di tali servizi, mostrando una crescente preoccupazione per la salute e la tutela dei minori.
Quello però che appare necessario è di trovare risposta ad alcuni interrogativi: cosa intende il legislatore quando si riferisce alla tutela dei *minori*? Da cosa scaturiscono le odierne preoccupazioni per bambini e preadolescenti? Quali sono i prossimi passi per rafforzare i propositi del legislatore europeo?

2. L'accesso a Internet dei minori: ovvero come i tredici anni sono diventati l'età della "maturità digitale"

2.1. Il Children's Online Privacy Protection Act del 1998

Per comprendere l'ambito di tutela del DSA e delle norme previste a protezione del minore occorre innanzitutto interrogarsi su chi sia considerato *minore* nel contesto digitale.
Per molto tempo l'età minima per accedere ai servizi della società dell'informazione è stata fissata a tredici anni in virtù di una normativa statunitense del 1998 posta a tutela della *privacy* dei minori, il *Children's Online Privacy Protection Act* (COPPA)[21].
Il testo del COPPA è stato ampiamente discusso prima di essere stato approvato. L'idea iniziale del deputato Edward Markey, proponente della Carta dei diritti sulla *privacy* dei bambini, era quella di definire bambino, ai fini del COPPA, il minore di sedici anni. Questo limite trovò l'opposizione sia delle grandi società di *e-commerce* portatrici di interessi privati, sia dei gruppi per le libertà civili. I primi non volevano rinunciare ad una redditizia fetta di mercato che non avrebbe potuto beneficiare di beni e servizi della società dell'informazione, i secondi invece non volevano che molti ragazzi venissero esclusi dall'accesso ad Internet e dunque a una serie di informazioni come quelle legate all'utilizzo di contraccettivi, all'aborto o all'aiuto in situazioni di abuso[22].

[20] Il contributo è consegnato a luglio 2024.
[21] Il *Children's Online Privacy Protection Act* del 1998 (COPPA) è una legge federale degli Stati Uniti, situata al 15 USC §§ 6501 – 6506 (Pub. L.Tooltip Diritto pubblico (Stati Uniti) 105–277, 112 Stat. 2681-728, emanato il 21 ottobre 1998). 16 CFR Part 312.
[22] J. JARGON, *How 13 Became the Internet's Age of Adulthood*, in *The Wall Street Journal*, 18 giugno 2019, in https://www.wsj.com/articles/how-13-became-the-internets-age-of-adulthood-11560850201 (ultima consultazione 7/07/2024).

Downloaded from www.biodiritto.org.
ISSN 2284-4503

La soglia di legittimità del trattamento dei dati è stata quindi individua\ta nella regola pratica "sotto i dodici anni", utilizzata negli anni Settanta dai regolatori negli Stati Uniti e in altri Paesi per elaborare leggi sul *marketing* rivolto ai bambini. Ciò era stato supportato da una ricerca[23] che aveva dimostrato che i bambini dagli otto ai dodici anni erano in grado di distinguere la pubblicità da altri contenuti.
È stata dunque valutata adeguata la soglia di tredici anni perché il minore potesse esprimere liberamente il consenso al trattamento dei propri dati personali. L'età della maturità digitale trae dunque la propria origine da studi che hanno a che vedere con la capacità dei bambini di distinguere l'*advertising* dal programma televisivo che stanno guardando, anziché partire da studi sullo sviluppo cognitivo vero e proprio, sulla capacità di esprimere il consenso informato, sulla capacità di navigare sul *web* o di comprendere ed elaborare i contenuti cui hanno accesso.
Pertanto, il COPPA ha dato una definizione generale di "bambino"[24] dichiarando che ai fini dello stesso regolamento dovesse essere considerato tale il minore di tredici anni di età[25], determinando quelli che sono i limiti e le condizioni del trattamento dei suoi dati personali.
La legge aveva previsto poi il divieto di raccolta, conservazione e divulgazione dei dati personali dei bambini da parte di siti *web* o servizi *online* in violazione di quanto previsto dal COPPA[26]. Il trattamento dei dati personali dei minori di tredici anni poteva avvenire qualora fosse stato espresso il consenso del genitore[27] o di chi esercitasse la responsabilità genitoriale nei confronti del bambino, nei modi e nei termini[28] previsti dall'*Act* stesso.
Non vige dunque ad oggi un divieto assoluto di trattamento dei dati personali dei bambini negli Stati Uniti, ma un divieto eventualmente superabile nel caso in cui venga espresso il consenso di chi esercita la responsabilità genitoriale e qualora il consenso sia informato e carpito in maniera certa da parte del titolare del trattamento.
La ragione per cui la maggior parte delle piattaforme prevede un divieto di utilizzo da parte dei minori di tredici anni non è dovuto al fatto che i bambini e i preadolescenti debbano essere tutelati con maggiore forza rispetto alle eventuali conseguenze di un'esposizione prematura ai *social* e in generale ai prodotti della società dell'informazione, o al fatto che l'accesso a determinati contenuti presenti sulle piattaforme sia inappropriato se non addirittura rischioso, bensì è una scelta "commerciale" delle *big tech*. Come poc'anzi evidenziato, sarebbe infatti astrattamente possibile l'accesso e il ricorso ai servizi

[23] R.P. ADLER ET AL., *Research on the Effects of Television Advertising on Children; A Review of the Literature and Recommendations for Future Research*, Washington DC, 1975, in https://files.eric.ed.gov/fulltext/ED145499.pdf (ultima consultazione 07/07/2024).
[24] 16 CFR 312.2 «*"Child" means an individual under the age of 13*».
[25] J. JARGON, *op. cit.*, in https://www.wsj.com/articles/how-13-became-the-internets-age-of-adulthood-11560850201 (ultima consultazione 07/07/2024).
[26] 16 CFR 312.3 «*General requirements. It shall be unlawful for any operator of a Web site or online service directed to children, or any operator that has actual knowledge that it is collecting or maintaining personal information from a child, to collect personal information from a child in a manner that violates the regulations prescribed under this part*».
[27] 16 CFR 312.5(a)(1).
[28] 16 CFR 312.4(a) «*General principles of notice. It shall be the obligation of the operator to provide notice and obtain verifiable parental consent prior to collecting, using, or disclosing personal information from children. Such notice must be clearly and understandably written, complete, and must contain no unrelated, confusing, or contradictory materials*».

Downloaded from www.biodiritto.org.
ISSN 2284-4503

della società dell'informazione da parte dei bambini, qualora il consenso al trattamento dei loro dati venisse legittimamente espresso dagli esercenti la responsabilità genitoriale. Ad oggi però le società che forniscono questo genere di servizi anziché creare laboriosi sistemi di verifica del consenso dei genitori e di verifica dei dati, di per sé molto complessi e dispendiosi, preferiscono prevedere che possano registrarsi alle piattaforme solo coloro che abbiano compiuto tredici anni.
Peraltro, il COPPA prevede che le società siano soggette a sanzioni se raccolgono o divulgano dati di bambini in violazione delle previsioni della legge, ma solo laddove ne siano effettivamente a conoscenza[29]. Ciò significa che, laddove il minore di tredici anni acceda ad un sito o si registri a una piattaforma mentendo sulla propria età, il fornitore di servizi che a quel punto tratterà i suoi dati non sarà soggetto a sanzioni.

2.2. Dal COPPA al GDPR

In Europa la disciplina relativa alla protezione dei dati personali è contenuta nel *General Data Protection Regulation* UE/2016/679 (GDPR), entrato in vigore nel 2018, che all'articolo 8 rubricato «Condizioni applicabili al consenso dei minori in relazione ai servizi della società dell'informazione»[30] prevede una disciplina apposita per il trattamento dei dati dei minori. L'atto europeo, a differenza del corrispettivo americano, prevede come regola generale che l'età del consenso digitale al trattamento dei dati, da parte di coloro che offrono servizi, sia stabilita a sedici anni, con la facoltà però per i singoli Stati membri di abbassare tale soglia, purché non al di sotto dei tredici anni. Quest'ultimo è probabilmente un richiamo al contenuto del COPPA e alla definizione che esso fornisce di "bambino" ai fini dell'applicabilità della disciplina sulla protezione della *privacy* del minore. In Italia, il legislatore ha esercitato tale facoltà portando la soglia della "maturità digitale" a quattordici anni[31].
L'articolo 8 del GDPR prevede anche che «ove il minore abbia un'età inferiore ai sedici anni, tale trattamento è lecito soltanto se e nella misura in cui tale consenso è prestato o autorizzato dal titolare

[29] 16 CFR 312.3 "*General requirements*".
[30] L'articolo 8 del GDPR recita «1. Qualora si applichi l'articolo 6, paragrafo 1, lettera a), per quanto riguarda l'offerta diretta di servizi della società dell'informazione ai minori, il trattamento di dati personali del minore è lecito ove il minore abbia almeno 16 anni. Ove il minore abbia un'età inferiore ai 16 anni, tale trattamento è lecito soltanto se e nella misura in cui tale consenso è prestato o autorizzato dal titolare della responsabilità genitoriale. Gli Stati membri possono stabilire per legge un'età inferiore a tali fini purché non inferiore ai 13 anni. 2. Il titolare del trattamento si adopera in ogni modo ragionevole per verificare in tali casi che il consenso sia prestato o autorizzato dal titolare della responsabilità genitoriale sul minore, in considerazione delle tecnologie disponibili. 3. Il paragrafo 1 non pregiudica le disposizioni generali del diritto dei contratti degli Stati membri, quali le norme sulla validità, la formazione o l'efficacia di un contratto rispetto a un minore».
[31] Decreto legislativo 30 giugno 2003, n. 196 recante il "Codice in materia di protezione dei dati personali", Art. 2-*quinquies* (Consenso del minore in relazione ai servizi della società dell'informazione): «1. In attuazione dell'articolo 8, paragrafo 1, del Regolamento, il minore che ha compiuto i quattordici anni può esprimere il consenso al trattamento dei propri dati personali in relazione all'offerta diretta di servizi della società dell'informazione. Con riguardo a tali servizi, il trattamento dei dati personali del minore di età inferiore a quattordici anni, fondato sull'articolo 6, paragrafo 1, lettera a), del Regolamento, è lecito a condizione che sia prestato da chi esercita la responsabilità genitoriale. 2. In relazione all'offerta diretta ai minori dei servizi di cui al comma 1, il titolare del trattamento redige con linguaggio particolarmente chiaro e semplice, conciso ed esaustivo, facilmente accessibile e comprensibile dal minore, al fine di rendere significativo il consenso prestato da quest'ultimo, le informazioni e le comunicazioni relative al trattamento che lo riguardi».

Downloaded from www.biodiritto.org.
ISSN 2284-4503

della responsabilità genitoriale» e che «il titolare del trattamento si adopera in ogni modo ragionevole per verificare in tali casi che il consenso sia prestato o autorizzato dal titolare della responsabilità genitoriale sul minore, in considerazione delle tecnologie disponibili». Pertanto, il limite al trattamento dei dati dei minori infrasedicenni può essere teoricamente superato, in Europa come negli Stati Uniti, per mezzo del consenso informato espresso da colui che esercita la responsabilità genitoriale sul bambino. Come evidenziato in precedenza[32] buona parte dei fornitori di servizi della società dell'informazione hanno preferito porre un divieto assoluto di utilizzo per gli utenti di età inferiore ai tredici anni, così da risparmiare costi e rischi legati allo sviluppo e utilizzo di sistemi di verifica del consenso legittimamente rilasciato dagli esercenti la responsabilità genitoriale in vece degli infratredicenni.

Secondo il Considerando (38) del GDPR, «i minori meritano una specifica protezione relativamente ai loro dati personali, in quanto possono essere meno consapevoli dei rischi, delle conseguenze e delle misure di salvaguardia interessate nonché dei loro diritti in relazione al trattamento dei dati […]». I bambini godono di una protezione speciale ai sensi del Regolamento generale sulla protezione dei dati in quanto considerati vulnerabili[33]. Non hanno ancora raggiunto la maturità fisica e psicologica, quindi potrebbero essere meno consapevoli degli adulti dei rischi e delle conseguenze della condivisione dei loro dati personali quando si registrano a servizi *online* o utilizzano piattaforme connesse.

Recentemente il Garante per la protezione dei dati personali italiano ha dato applicazione al GDPR richiamando le norme poste a tutela dei minori al fine di oscurare due piattaforme: *TikTok*[34] e *ChatGPT*[35]. In entrambi i casi le *policy* delle società prevedevano un divieto di utilizzo delle piattaforme ai minori di tredici anni, senza però aver predisposto dei sistemi di verifica dell'età adeguati a impedire l'accesso a soggetti minori. Il Garante della *privacy* italiano, invocando le norme del GDPR poste a tutela del minore e il principio del *best interest of the child* di cui all'articolo 24, par. 2, della Carta dei diritti fondamentali dell'Unione europea[36], ha deciso di oscurare le piattaforme su tutto il territorio italiano, sottolineando in entrambi i casi come l'esposizione dei bambini a contenuti inidonei al loro grado di sviluppo, esponesse gli stessi a rischi intollerabili e pertanto fosse necessario vietare l'accesso a tali piattaforme su tutto il territorio italiano, a tutela dei minori.

[32] Cfr. par. 2.1.

[33] EUROPEAN DATA PROTECTION BOARD (EDPB), *Linee guida 5/2020 sul consenso ai sensi del regolamento (UE) 2016/679*, Versione 1.1, adottate il 4 maggio 2020, p. 28. «Rispetto alla direttiva attuale, il regolamento generale sulla protezione dei dati crea un ulteriore livello di protezione per il trattamento dei dati personali delle persone fisiche *vulnerabili*, in particolare i minori». Nostro il corsivo.

[34] Garante per la protezione dei dati personali, provvedimento del 22 gennaio 2021, in https://www.garanteprivacy.it/*web*/guest/home/doc*web*/-/doc*web*-display/doc*web*/9524194 (ultima consultazione 29/07/2024). Per un commento si vd. D. MARCELLO, *Circolazione dei dati del minore tra autonomia e controllo. Norme e prassi nel mercato digitale europeo*, Napoli, 2023, 62 ss.

[35] Garante per la protezione dei dati personali, Provvedimento del 30 marzo 2023, in https://www.garanteprivacy.it/home/doc*web*/-/doc*web*-display/doc*web*/9870832 (ultima consultazione 29/07/2024). Per un commento al provvedimento si vd. G. PISTORIO, *Chat GPT e la sfida della regolamentazione normativa*, in *Associazione italiana costituzionalisti – La Lettera,*5, 2023, in https://www.associazionedeicostituzionalisti.it/it/la-lettera/05-2023-costituzione-e-intelligenza-artificiale/chat-gpt-e-la-sfida-della-regolamentazione-normativa (ultima consultazione 29/07/2024).

[36] Carta dei diritti fondamentali dell'unione europea (2000/C 364/01).

Downloaded from www.biodiritto.org.
ISSN 2284-4503

3. Uno sguardo comparato: il contenzioso negli Stati Uniti

Per anni il fulcro della tutela dei minori nell'ambiente digitale è stata dunque la *privacy* dei bambini e la tutela dei loro dati personali, ma alcune inchieste e ricerche recenti sembrano aver dato nuova linfa al tema.

A settembre 2021 sono stati pubblicati i c.d. *Facebook files*[37] da parte del Wall Street Journal, un'inchiesta nata dalla collaborazione tra il famoso quotidiano americano e una ex dipendente di Meta, Frances Haugen, volta a dimostrare, tra le altre cose, come il colosso di Menlo Park fosse perfettamente a conoscenza degli effetti dannosi di *Instagram* sulla salute mentale di bambini e adolescenti[38].

A seguito della pubblicazione dei *Facebook files*, il *Surgeon General*[39] degli Stati Uniti, il Dr. Vivek Murthy, ha iniziato a pubblicare una serie di *advisory*[40] che hanno richiamato l'attenzione sulla crisi nazionale della salute mentale, del benessere dei giovani[41] e sulle profonde conseguenze sulla salute derivanti dalla c.d. "*social disconnection*" [42]. Infine, a maggio 2023 è stato pubblicato un parere dal titolo "*Social media and Youth Mental Health"*[43] che descrive l'impatto che i *social media* hanno sulla salute mentale di bambini e adolescenti sulla base dei dati ad oggi disponibili.

Quello che emerge chiaramente da quest'ultimo *advisory* è che, seppur gli studi non siano ancora del tutto conclusivi e vi sia necessità di approfondire ulteriormente l'impatto dei *social* sulla salute psicofisica dei minori, non è oggi possibile affermare con certezza che l'utilizzo delle piattaforme sia sufficientemente sicuro per il loro sviluppo ed il loro benessere[44]. Dai numerosi studi citati[45] emerge come

[37] Si vd. https://www.wsj.com/articles/the-Facebook-files-11631713039 (ultima consultazione 07/07/2024).

[38] G. WELLS, J. HORWITZ, D. SEETHARAMAN, Facebook *Knows* Instagram *Is Toxic for Teen Girls, Company Documents Show*, in *The Wall Street Journal*, 14 settembre 2021, in https://www.wsj.com/articles/Facebook-knows-Instagram-is-toxic-for-teen-girls-company-documents-show-11631620739 (ultima consultazione 07/07/2024).

[39] Il *Surgeon General* degli Stati Uniti (*Surgeon General of the United States*) è il capo esecutivo dello *United States Public Health Service Commissioned Corps* e il portavoce delle questioni di salute pubblica all'interno del governo federale. È il soggetto più autorevole e titolato in materia di sanità pubblica negli Stati Uniti e in quanto tale costituisce il principale consigliere della Casa Bianca sul tema.

[40] Per una definizione di "*advisory*" si vd. THE U.S. SURGEON GENERAL'S ADVISORY, *Social media and Youth Mental Health*, 2023, 3, in https://www.hhs.gov/surgeongeneral/priorities/youth-mental-health/social-media/index.html (ultima consultazione 07/07/2024): «*A Surgeon General's Advisory is a public statement that calls the American people's attention to an urgent public health issue and provides recommendations for how it should be addressed. Advisories are reserved for significant public health challenges that require the nation's immediate awareness and action*».

[41] THE U.S. SURGEON GENERAL'S ADVISORY, *Protecting Youth Mental Health*, 2021, in https://www.hhs.gov/sites/default/files/surgeon-general-youth-mental-health-advisory.pdf (ultima consultazione 07/07/2024).

[42] THE U.S. SURGEON GENERAL'S ADVISORY, *Our Epidemic of Lonliness and Isolation*, 2023, in https://www.hhs.gov/sites/default/files/surgeon-general-social-connection-advisory.pdf (ultima consultazione 07/07/2024).

[43] THE U.S. SURGEON GENERAL'S ADVISORY, *op. cit.*, in https://www.hhs.gov/sites/default/files/sg-youth-mental-health-social-media-advisory.pdf (ultima consultazione 7/07/2024).

[44] Per un commento si vd. S. CALZOLAIO, *Social media e minori. Il Safety-first approach. Nota a: U.S. Surgeon General, Social media and Youth Mental Health. The U.S. Surgeon General's Advisory, 2023*, in *Rivista di informatica e diritto*, 2, 2023, 292 ss.

[45] Fra i tanti cfr. G. FIORAVANTI, S. CASALE, S.B. BENUCCI, A. PROSTAMO, A. FALONE, V. RICCA, F. ROTELLA, *Fear of missing out and social networking sites use and abuse: A* Meta-*analysis*, in *Computers in Human Behavior*, 122, 2021, 1-12.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

vi sia una correlazione allarmante tra utilizzo dei *social* da parte di bambini e adolescenti e l'aumento delle malattie legate alla salute mentale come depressione, ansia, disturbi alimentari, disturbi dell'attenzione e della qualità del sonno.
A seguito della pubblicazione dell'*advisory* del *U.S. Surgeon General* e prima che la Commissione europea aprisse il procedimento contro *Meta* ai sensi del DSA, ad ottobre 2023 già oltre quaranta Stati americani e il Distretto di Columbia avevano citato il colosso dei *social media* in giudizio con l'accusa di aver intenzionalmente progettato dei prodotti che creano dipendenza e che sono dannosi per i giovani utenti di *Instagram* e *Facebook*[46]. I ricorrenti in questione hanno accusato *Meta* di aver ingannato i consumatori in merito agli effetti dannosi sui più giovani. Inoltre, la *big tech* è stata accusata di aver commercializzato dei prodotti a utenti di età inferiore ai tredici anni, contravvenendo alla legge federale sulla protezione della *privacy* in rete dei minori[47] e alla sua stessa *policy*[48]. Quello che emerge chiaramente dalle casistiche riportate è come il tema della *privacy* e della protezione dei dati personali finisca per diventare secondario, la maggiore preoccupazione non è più e non è solo la riservatezza del bambino, ma la tutela della sua persona.

4. Oltre la privacy: gli "age verification system"

Da quanto finora emerso, è possibile ravvisare una crescente preoccupazione da parte degli attori pubblici per quanto attiene alla tutela dei minori all'interno dello spazio digitale. Una tutela non più limitata al trattamento dei dati personali e al diritto alla *privacy* del minore, ma che si estende alla protezione del bambino.
Dal quadro delineato, infatti, il minore risulta un soggetto vulnerabile, che agisce nella società dell'informazione in maniera autonoma al pari di un adulto, seppur egli non sia dotato degli strumenti e della maturità di un adulto. La condizione di vulnerabilità del minore è da intendersi in maniera specifica in relazione alle sue caratteristiche intrinseche[49], poiché egli si trova in una fase della vita in cui necessita di una protezione rafforzata, proprio in virtù della sua incapacità di difendersi autonomamente dai danni che l'esposizione ai rischi del *web* potrebbe provocare[50].
Recentemente la Commissione europea, come evidenziato in apertura, si è adoperata sul tema, sostenendo e promuovendo l'attuazione di norme mirate alla tutela dei minori *online*: in particolare, l'articolo 28 del DSA richiede che tutti i fornitori di piattaforme *online* accessibili ai minori adottino misure

[46] https://www.washingtonpost.com/technology/2023/10/24/Meta-lawsuit-Facebook-Instagram-children-mental-health/ (ultima consultazione 7/07/2024).
[47] Cfr. *Children's Online Privacy Protection Act*, 1998, (COPPA).
[48] Nelle condizioni d'uso di Facebook si legge che «Il nostro obiettivo è rendere Facebook disponibile a tutti, ma il suo uso è proibito nei casi seguenti: per gli utenti che hanno meno di 13 anni [...]». Così come le Condizioni d'uso di Instagram prevedono «Chi può usare Instagram: desideriamo che il nostro Servizio sia quanto più aperto e inclusivo possibile, ma vogliamo che sia anche sicuro, protetto e conforme alla legge. Pertanto, l'utente è tenuto a rispettare alcune limitazioni legali per poter far parte della community di Instagram. L'utente deve avere almeno 13 anni [...]».
[49] In relazione alle dimensioni ontologica e specifica del concetto di vulnerabilità vds. L. BUSATTA, C. CASONATO, S. PENASA, M. TOMASI, *Le "maschere" della vulnerabilità nella cura della persona*, AA. VV. (a cura di), *Liber amicorum per Paolo Zatti*, Napoli, 2023, 651-652.
[50] Sulla nozione di vulnerabilità vds. *Ibidem*, 651.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

adeguate e proporzionate per garantire un elevato livello di tutela della vita privata, di sicurezza e di protezione dei minori, anzitutto mediante l'attivazione dei meccanismi di verifica dell'età. Inoltre, l'articolo 35, paragrafo 1, lettera j), del DSA, prevede che i fornitori di piattaforme *online* e di motori di ricerca *online* di dimensioni molto grandi adottino misure di attenuazione dei rischi sistemici, tra cui quelle «mirate per tutelare i diritti dei minori, compresi strumenti di verifica dell'età e di controllo parentale, o strumenti volti ad aiutare i minori a segnalare abusi o ottenere sostegno, a seconda dei casi».

I sistemi di *age verification* risultano pertanto essere il punto di partenza per poter rispondere efficacemente alla presenza illecita dei minori su Internet. In tal senso in Italia, a seguito della conversione del c.d. "Decreto Caivano"[51] recante specifiche disposizioni per la sicurezza dei minori in ambito digitale, è stata data attuazione all'articolo 49, comma 2, del DSA: l'Autorità per le garanzie nelle comunicazioni (AGCOM) è stata designata Coordinatore dei servizi digitali per l'Italia, ossia l'autorità preposta a garantire l'effettività dei diritti e l'efficacia degli obblighi stabiliti dal Regolamento, «nonché la relativa vigilanza e il conseguimento degli obiettivi previsti, anche con riguardo alla protezione dei minori in relazione ai contenuti pornografici disponibili *online*, nonché agli altri contenuti illegali o comunque vietati, veicolati da piattaforme *online* o altri gestori di servizi intermediari, e contribuire alla definizione di un ambiente digitale sicuro»[52].

L'AGCOM è stata incaricata di stabilire, previa consultazione del Garante per la protezione dei dati personali, le modalità tecniche e di processo che gestori e fornitori di servizi della società digitale devono adottare per l'accertamento della maggiore età degli utenti che accedano a siti a carattere pornografico, con un livello di sicurezza adeguato e il rispetto della minimizzazione dei dati raccolti[53]. Il 6 marzo 2024 l'AGCOM ha pertanto avviato una consultazione pubblica[54] per l'approvazione di un provvedimento che disciplini le modalità tecniche e di processo per l'accertamento della maggiore età degli utenti[55]. L'allegato B di tale provvedimento riporta quella che è un'analisi dei principali sistemi di verifica dell'età ad oggi esistenti, mettendone in luce pregi e difetti[56] e sottolineando poi quali devono essere i requisiti generali che un sistema di verifica dell'età deve rispettare.

[51] Decreto-legge 15 settembre 2023, n. 123, coordinato con la legge di conversione 13 novembre 2023, n., recante: «Misure urgenti di contrasto al disagio giovanile, alla povertà educativa e alla criminalità minorile, nonché per la sicurezza dei minori in ambito digitale».

[52] Art. 15, comma 1, d.l. n. 123/2023 (c.d. Decreto Caivano).

[53] Art. 13-bis, d.l. n. 123/2023 (c.d. Decreto Caivano).

[54] L'Autorità Garante per le Comunicazioni con Delibera del 6 marzo 2024, n. 61/24/CONS. ha dato «Avvio della consultazione pubblica di cui all'art. 1, comma 4, della delibera n. 9/24/CONS volta all'adozione di un provvedimento sulle modalità tecniche e di processo per l'accertamento della maggiore età degli utenti in attuazione della dalla legge 13 novembre 2023, n. 159».

[55] Legge 13 novembre 2023, n. 159, Conversione in legge, con modificazioni, del decreto-legge 15 settembre 2023, n. 123, recante misure urgenti di contrasto al disagio giovanile, alla povertà educativa e alla criminalità minorile, nonché per la sicurezza dei minori in ambito digitale.

[56] Secondo il report dell'European Parliamentary Research Service del febbraio 2023 i metodi più diffusi di age verification sono: autodichiarazione; inserimento della carta di credito; utilizzo della biometria; analisi dei comportamenti su internet; verifiche online e offline dei documenti di identità; consenso dei genitori; vouching; identificazione digitale (es. SPID); portafoglio per l'identità digitale; utilizzo di app specifiche; verifica tramite sms o e-mail; open banking.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Dal report dell'AGCOM è possibile individuare principalmente tre diverse macrocategorie di sistemi di verifica dell'età: l'autodichiarazione, cioè la dichiarazione semplice dell'utente in merito al possesso del requisito dell'età o meno, evidentemente il metodo meno attendibile in quanto facile da aggirare; i sistemi di certificazione da parte di terzi, che costituiscono il sistema più attendibile però più invasivo; i sistemi di riconoscimento, che fanno ricorso all'intelligenza artificiale per verificare l'età del soggetto e che utilizzano un vasto numero di dati e presentano il più alto grado di insidie in quanto maggiormente a rischio di commettere errori[57].

I sistemi di verifica dell'età assumono un ruolo particolarmente rilevante poiché sono gli strumenti che consentono o vietano l'accesso ai servizi della società dell'informazione per la cui fruizione è richiesta un'età minima. Di fatto, la predisposizione di tali sistemi e la loro effettività sono in grado di incidere sulla libertà dell'utente di accedervi o meno e di svolgervi la propria personalità. Non sembra un caso il fatto che l'individuazione di *age verification system* sia stata demandata proprio all'AGCOM, cioè l'autorità posta a garanzia, tra le altre cose, della libertà di comunicazione, di informazione e di espressione, anziché al Garante delle protezione dei dati personali, deputato, per l'appunto, alla tutela della *privacy*. Risulta evidente uno spostamento del problema dell'accesso dei minori alle piattaforme *online* da una logica *privacy*, incentrata sul consenso, a una logica costituzionale di bilanciamento delle libertà costituzionali e delle istanze di tutela del soggetto minore che potrebbe essere realizzato tramite il ricorso ai sistemi di *age verification*.

5. I minori sulla rete: un problema di "precauzione" costituzionale

Quello che sembra emergere dallo scenario finora delineato è una progressiva espansione dell'ambito di tutela del minore nell'ambiente digitale. Per anni gli attori pubblici hanno operato una tutela secondo la logica della *privacy*, dando preminente importanza al consenso quale strumento per la liceità del trattamento dei dati personali: da un lato, infatti, il COPPA e il GDPR stabiliscono un'età della maturità digitale diversa e non coincidente con l'età della capacità di agire, che consente all'infradiciottenne di disporre dei propri dati prestando il proprio consenso; dall'altro lato sia la normativa statunitense che quella europea consentono astrattamente la liceità del trattamento dei dati personali del minore anche qualora il consenso sia espresso da colui che esercenti la responsabilità genitoriale.

La tutela apprestata alla *privacy* è pertanto una tutela relativamente debole in quanto superabile attraverso l'esercizio dello strumento negoziale per eccellenza, ovverosia il consenso di colui che sia legittimato dalla legge a disporre dei dati personali di un soggetto minore o da parte del minorenne stesso. Ne deriva che fino ad oggi è stata solo la mancanza di consenso dell'esercente la responsabilità genitoriale a frapporsi tra il minore e la sua presenza *online*. Il divieto di accesso ad alcuni servizi della società dell'informazione è dovuto esclusivamente a una scelta di *policy* delle *big tech*, che hanno preferito vietare l'accesso ai minori anziché predisporre complicati sistemi di verifica del consenso

[57] Con il paradosso per cui lo stesso sistema di *age verification* integra un trattamento automatizzato di dati personali. Di conseguenza, laddove un minore di 14 anni tenti di accedere ad una piattaforma, non vi può essere un consenso validamente espresso neppure allo stesso trattamento di *age verification*.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

dell'esercente la responsabilità genitoriale [58]. L'unico fondamento della tutela, dunque, è solo l'autoregolamentazione delle piattaforme[59].

Non vi è dubbio che l'accesso alle piattaforme *online* costituisca oggi uno dei mezzi attraverso il quale un soggetto può svolgere la sua personalità, esercitare diritti e libertà costituzionalmente garantiti quali la libertà di comunicazione[60], di informazione e di espressione[61]. Nel caso del minore però si pone una questione ulteriore, legata al fatto che l'esposizione prematura dello stesso all'ambiente digitale potrebbe comportare dei danni per la sua salute psico-fisica e comprometterne persino lo sviluppo, come emerge dai più recenti risvolti scientifici[62].

La Costituzione italiana nel riconoscere i diritti fondamentali non opera alcuna distinzione basata sull'età, ciò significa che il minore può godere degli stessi a prescindere dalla sua condizione personale. Tuttavia, la complessità della posizione del minore, quale persona in divenire, pone non pochi problemi nel godimento dei diritti che attengono alla sua sfera personale. Sono due le finalità di tutela assunte come prioritarie per il minore: da una parte, le istanze di autodeterminazione; dall'altra, le esigenze di cura[63].

Se da un lato la presenza del minore *online* consente allo stesso di esercitare diritti e libertà costituzionalmente garantiti, dall'altro lato si pone la necessità di operare un bilanciamento con le istanze di tutela che nascono dai rischi che possono insorgere a seguito di questa esposizione prematura. Il minore, infatti, nella Costituzione riemerge come destinatario di una tutela più ampia e distinta rispetto a quella dell'adulto in quanto persona in formazione[64], pertanto si pone l'esigenza di predisporre strumenti adeguati a tale scopo.

Senza voler scadere in conclusioni paternalistiche e anacronistiche, alla luce delle incertezze e delle conseguenze negative che la presenza prematura del minore *online* può avere sullo stesso, sembrerebbe opportuno ricorrere al *principio di precauzione*[65], elaborato nell'area del diritto dell'ambiente,

[58] Questo è ciò che accade ogni volta che un genitore fornisce il proprio consenso alla creazione di un *account* del proprio bambino per registrarsi ad un'applicazione riservata ai minori, per esempio *YouTube Kids* richiede al genitore il consenso al trattamento dei dati del bambino che poi potrà utilizzare autonomamente la piattaforma. Vd. la *privacy policy* di *YouTube Kids* https://kids.youtube.com/t/privacynotice (ultima consultazione 30/07/2024).

[59] Sul tema si vd. E. CREMONA, *I poteri privati nell'era digitale. Libertà costituzionali, regolazione del mercato, tutela dei diritti*, Napoli, 2023, 46 ss.

[60] Art. 15, Cost.

[61] Art. 21, Cost.

[62] Vd. par. 3.

[63] G. MATUCCI, *Lo statuto costituzionale del minore d'età*, Padova, 2015.

[64] C. DI COSTANZO, *La tutela del minore: identità, salute e relazion*i, Torino, 2023, 18. Cfr. anche G. MATUCCI, *op. cit*.; G. DE MINICO, *Il favor constitutionis e il minore: realtà o fantasia?* in A. CIANCIO, G. DE MINICO, G. DEMURO, F. DONATI, M. VILLONE (a cura di), *Nuovi mezzi di comunicazione e identità. Omologazione o diversità?* Roma, 2012, 162; F. MODUGNO, *Breve discorso intorno all'uguaglianza. Studio di una casistica: i minori e i nuovi media*, in *Osservatorio costituzionale*, 1, 2014, 1-14.

[65] Seppur desumibile implicitamente anche da convenzioni precedenti, il principio di precauzione trova il suo esplicito riconoscimento internazionale nel 1992 nella Dichiarazione di Rio su ambiente e sviluppo. Il principio ha fatto il suo ingresso a livello comunitario con il Trattato di Maastricht ed è attualmente richiamato dall'art. 191 TFUE senza che ne venga fornita una sua definizione. S.GRASSI, A.GRAGNANI, *Il principio di precauzione nella giurisprudenza costituzionale*, in L.CHIEFFI (a cura di), *Biotecnologie e tutela del valore ambientale*, Torino, 2003, 149-169; G. GALASSO, *Il principio di precauzione nella disciplina degli OGM,* Torino, 2006; F. DE LEONARDIS, *Il principio di*

Downloaded from www.biodiritto.org.
ISSN 2284-4503

per riconsiderare le misure volte a limitare l'accesso ai servizi della società dell'informazione da parte dei minori. In base a tale principio, la condizione di incertezza a riguardo dei possibili effetti negativi dell'impiego di una tecnologia non può essere utilizzata come una ragione legittima per non regolare e limitare tale sviluppo[66].

Trasponendo tale principio nell'ambiente digitale, sembrerebbe opportuno ricorrere a un progressivo rafforzamento delle tutele disposte dall'ordinamento nei confronti del minore. Il ricorso a sistemi di *age verification* effettivi potrebbe essere un punto di partenza per operare il bilanciamento sopra auspicato. Occorrerà seguire con attenzione la concreta definizione di questi sistemi perché rappresenta un tema dall'indiscusso tono costituzionale.

precauzione nell'amministrazione del rischio, Milano, 2005; a livello internazionale non mancano autorevoli posizioni critiche nell'applicazione forte del principio di precauzione cfr. C.R. SUNSTEIN, *Laws of fear: beyond the precautionary principle*, Cambridge, 2005 (trad. it. *Il diritto della paura: oltre il principio di precauzione*, Bologna, 2010).

[66] A. SIMONCINI, *L'algoritmo incostituzionale: l'intelligenza artificiale e il futuro delle libertà*, in *BioLaw Journal – Rivista di BioDiritto*, 1, 2019, 86 ss.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

L'uso dell'intelligenza artificiale in ambito sanitario: riflessioni a partire da una sperimentazione per lo sviluppo di un SAMD per la diagnosi di autismo infantile

Chiara Vadalà *

THE USE OF ARTIFICIAL INTELLIGENCE IN THE DIAGNOSIS OF CHILDHOOD AUTISM

ABSTRACT: Starting from the experience of a research group at Stanford University, who is experimenting an AI medical device to analyse the voice and identify typical alterations, thus arriving at a possible diagnosis of autism spectrum disorder, the present research proposed a brief overview of the risks, benefits and of the regulatory corpus applicable today in the European territory and some considerations regarding bias and diagnostic error and the protection tools to guarantee ethical and reliable AI, also in light of the debate on product liability with artificial intelligence.

KEY WORDS: Autism; access to care; medical device; bias; diagnostic error.

ABSTRACT: Partendo dall'esperienza di un gruppo di ricerca dell'Università di Stanford, che sta sperimentando un dispositivo medico di IA per analizzare la voce e identificare le alterazioni tipiche, arrivando così a una possibile diagnosi di disturbo dello spettro autistico, la presente ricerca ha proposto una breve panoramica dei rischi, dei benefici e del corpus normativo oggi applicabile nel territorio europeo e alcune considerazioni su bias ed errori diagnostici e sugli strumenti di protezione per garantire un'IA etica e affidabile, anche alla luce del dibattito sulla responsabilità del prodotto con l'intelligenza artificiale.

PAROLE CHIAVE: Autismo; accesso alle cure; dispositivi medici; bias; errore diagnostico.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

* *Dottoranda di ricerca, Università Uninettuno. Mail: chiara.vadala@gmail.com.* *Contributo sottoposto a doppio referaggio anonimo.*

1. Inquadramento sistematico

Negli ultimi anni, lo sviluppo dei sistemi di Intelligenza Artificiale ha raggiunto un livello tale da permettere di apprezzarne l'impatto nella vita quotidiana e da lasciar intravedere i profili dello sviluppo futuro, sia nel campo prettamente privatistico – contrattuale, societario - sia nella pubblica amministrazione – digitalizzazione dei procedimenti amministrativi; utilizzo delle nuove tecnologie nei processi legislativi, decisionali, giudiziari -.

Quella che è stata definita la quarta rivoluzione industriale[1] si caratterizza per la capacità di trascendere i confini classici di pubblico e privato e attraversare trasversalmente le aree di interesse del diritto, per il tramite della chiave di lettura dei diritti e della loro tutela, in ambito sia privato che pubblico[2].

Il diritto sanitario, con la sua trasversalità tra pubblico e privato, permette di osservare in modo privilegiato l'enorme potenziale e gli enormi rischi[3] dell'utilizzo dell'intelligenza artificiale.

E difatti, in ambito sanitario risaltano, in modo quasi ovvio, una serie di benefici dell'utilizzo dell'IA, particolarmente in termini di costi, tempi ed efficienza. Ai grandi benefici si accompagnano i notevoli rischi, che lo sforzo regolatorio, attualmente in atto in modo globale, pur nelle diversità degli approcci, cerca di contenere. Entrambi gli aspetti emergono con chiarezza nello specifico settore dell'*AI Health*.

Più in particolare, la presente riflessione muove le mosse dalla sperimentazione di un dispositivo medico con IA per la diagnosi dell'autismo, indirizzato quindi ad una categoria di soggetti vulnerabili sotto il duplice profilo dell'età e della disabilità, quale paradigma per una più ampia riflessione sull'IA in ambito sanitario, attraverso alcune considerazioni in ordine agli strumenti di tutela profilati e alla loro astratta idoneità a garantire una IA etica ed affidabile, anche alla luce del dibattito in tema di responsabilità da prodotto con intelligenza artificiale.

2. Il potenziale dell'impiego dell'Intelligenza Artificiale nel settore sanitario

Le opportunità offerte dai sistemi artificiali attingono numerosi ambiti e, già tutt'oggi, questa tecnologia ha pervaso molteplici settori della nostra quotidianità[4]. In questi ultimi anni, i sistemi di IA hanno avuto uno sviluppo esponenziale[5], e ci si attende un ulteriore e rapido sviluppo futuro, in particolare per quanto concerne l'intelligenza artificiale generativa. Tale circostanza ha catalizzato l'attenzione sui benefici attesi per le persone e la società[6], evidenti quanto meno sotto tre profili: riduzione dei tempi, riduzione dei costi, miglioramento dell'efficienza.

---

[1] K. SCHWAB, *La quarta rivoluzione industriale*, Milano, 2016.

[2] S. RODOTÀ, *Il diritto di avere diritti*, Bari, 2012.

[3] *AI in health: huge potential, huge risks*, OECD, 2024 in https://www.oecd.org/en/publications/ai-in-health_2f709270-en.html (ultima consultazione 02/12/2024)

[4] Attualmente i sistemi di intelligenza artificiale sono di uso comune in molti settori della medicina. Per una disamina dell'utilizzo nella diagnostica per immagini, si veda SIRM, *Intelligenza artificiale in radiologia*, 2020 in https://sirm.org/wp-content/uploads/2021/04/317-Documento-SIRM-2020-Intelligenza-artificale-in-radiologia.pdf (ultima consultazione 02/12/2024).

[5] Il progresso tecnologico è descritto come una funzione esponenziale. Si veda R. KURZWEIL, *La singolarità è vicina*, Milano, 2008.

[6] M. FASAN, *I principi costituzionali nella disciplina dell'intelligenza artificiale. Nuove prospettive interpretative*, in *DPCE online*, 1, 2022, 184.

Downloaded from www.biodiritto.org.

ISSN 2284-4503

I sistemi di IA posseggono infatti una potenza computazionale di analisi delle informazioni e una capacità di individuare correlazioni rilevanti tra i dati esaminati[7], tale da poter abbattere i tempi normalmente richiesti per processare una mole ingente di dati e pervenire all'indicazione di una decisione. In secondo luogo l'AI aumenta la possibile efficacia delle soluzioni proposte, riducendo gli errori e permettendo un affinamento sempre più personalizzato delle valutazioni.

Queste due capacità, di velocità di processamento dei big data, e di efficientamento del processo valutativo, hanno, come ricaduta conseguenziale, la riduzione dei costi, anche sotto il profilo del migliore utilizzo delle risorse.

Queste doti emergono, con specifiche caratterizzazioni, se si osserva l'applicazione dell'intelligenza artificiale all'attività sanitaria; un utilizzo per cui è previsto, sia a livello nazionale[8] che globale[9], un incremento, con l'attesa di notevoli e rapidi benefici[10], ulteriori rispetto a quelli già oggi verificabili[11], idonea ad incidere in tutte e tre le aree (costi, tempi ed efficienza) nei profili del management, della comunicazione, della diagnosi, della personalizzazione della terapia, della ricerca, della prevenzione, della cybersicurezza[12].

Innanzitutto, l'IA può essere impiegata per la gestione dei dati sanitari, sotto diversi profili.

La digitalizzazione del fascicolo sanitario e la conseguente semplificazione dell'accesso per l'utenza, ma anche l'alleggerimento del carico burocratico sul professionista medico, hanno un immediato riscontro in termini di abbattimento dei costi ed incremento dell'efficienza dell'azienda sanitaria. Ben il 36% delle attività nel settore sanitario e sociale potrebbe essere automatizzato IA[13]. Questi incrementi di produttività ridurrebbero il deficit previsto di 3,5 milioni professionisti sanitari richiesti entro il 2030 in tutta l'OCSE[14].

L'IA può supportare questi aspetti non solo per la veloce e corretta gestione di una grande mole di dati, ma anche per migliorare la sicurezza della tenuta dei dati informatici con appositi sistemi a protezione dell'infrastruttura sanitaria digitale dalle minacce della criminalità informatica[15], per individuarle e contribuire a prevenirle.

---

[7] I sistemi di IA procedono infatti non secondo una logica di causalità, ma secondo una logica di correlazione. Sul punto P. Traverso, *Breve introduzione tecnica all'intelligenza artificiale*, in *DPCE online*, 1, 2022, 155-167.

[8] Si veda il DDL 23 aprile 2024.

[9] Si veda COM(2020)65final, *Libro bianco sull'intelligenza artificiale – un approccio europeo all'eccellenza e alla fiducia*, dove si afferma: «l'intelligenza artificiale si sta sviluppando rapidamente. Cambierà le nostre vite migliorando l'assistenza sanitaria (ad esempio rendendo le diagnosi più precisa e consentendo una migliore prevenzione delle malattie».

[10] Il suo utilizzo in ambito medico, ad esempio, se opportunamente regolato, permetterà di incrementare qualità e quantità delle prestazioni erogate dal sistema sanitario nazionale e aumentare la garanzia in concreto del diritto alla salute (L. Rinaldi, *Intelligenza artificiale, diritti e doveri nella Costituzione italiana*, in *DPCE online*, 1, 2022, 205).

[11] OEDC, *op. cit*., dichiara che l'evidenza empirica suggerisce che solo nel 2023 alcuni 163.000 persone potrebbero essere morte in Europa a causa di errori medici.

[12] L. Scaffardi, *La medicina alla prova dell'intelligenza artificiale*, in *DPCE online*, 1, 2022, 349- 359.

[13] R.K. Chebrolu, *Smart use of artificial intelligence in health care*, Deloitte, 2020.

[14] OECD (2023), *Ready for the Next Crisis? Investing in Health System Resilience*, OECD Health Policy Studies, OECD Publishing, Paris, https://doi.org/10.1787/1e53cf80-en (ultima consultazione 23/02/2024).

[15] B. Aiyer, *New survey reveals $2 trillion market opportunity for cybersecurity technology and service providers*, 2022, McKinsey & Company.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

Il migliore trattamento dei dati, poi, è in grado di apportare benefici alla comunicazione, permettendo di avere *le informazioni giuste al momento giusto alle persone giuste e per il giusto contesto, consentendo la prevenzione degli errori medici*[16].
L'intelligenza artificiale può aiutare il settore sanitario a sbloccare il valore del 97% delle risorse di dati sanitari[17] e, utilizzando grandi quantità di prove cliniche (ad esempio, imaging, storie di pazienti) potrebbe espandere esponenzialmente la medicina basata sull'evidenza per migliorare i risultati sanitari e l'assistenza centrata sulle persone, personalizzata, aperta a soluzioni nuove e nuovi approcci alla cura delle malattie rare, grazie alla capacità di individuare correlazioni insolite, proprie della modalità di "ragionamento" dei sistemi di *deep machine learning*.
La capacità computazionale permette di leggere i dati con un approccio predittivo, non solo sul singolo paziente, ampliando la sfera dell'intervento preventivo (che si risolve sempre in un ventaglio di vantaggi: personale, sociale, economico), ma anche nel rilevare i primi segnali di nuove patologie, come accaduto nella pandemia COVID-19, e nell'accelerare la scoperta di vaccini[18].
La diffusione di dispositivi medici con IA ha la potenzialità anche di delocalizzare, in tutto o in parte, la diagnosi e la terapia, permettendo di affidare l'esecuzione al sistema macchina e lasciare al personale medico la supervisione e il monitoraggio, in modo da attuare maggiormente una medicina territoriale, anche domiciliare, e, quindi, capillare, già implementata con le forme di telemedicina e che, integrando i software di IA nei dispositivi medici, può aumentare vertiginosamente i risultati di questo approccio.

2. I rischi dell'AI in ambito sanitario

Le attività ad alta complessità tecnica ed intrinsecamente pericolose espongono a dei rischi, ma apportano un benessere sociale, in virtù del quale vengono consentite[19].
Ci sono rischi che devono essere affrontati in modo efficace e, per far ciò, devono essere innanzitutto individuati e definiti.
L'IA ha strutturalmente delle criticità legate al livello di affidabilità del sistema e alla qualità del risultato, condizionate da due principali elementi: la qualità dei dati di addestramento e la trasparenza del processo computazionale. L'adeguatezza dei dati è stata oggetto di attenzione del legislatore europeo[20], che ne ha identificato i requisiti costitutivi: essere pertinenti, rappresentativi, esenti da errori e completi e tenere conto, in relazione all'uso cui sono destinati, delle caratteristiche o degli elementi particolari dello specifico contesto geografico, comportamentale o funzionale. Un set completo permette di minimizzare il rischio di bias algoritmico, frutto di una correlazione scorretta da imputare ad

[16] European Alliance for Access to Safe Medicines, *Medication Errors – the Most Common Adverse Event in Hospitals Threatens Patient Safety and Causes 16/1/2 Deaths per Year*, European Alliance for Access to Safe Medicines, https://eaasm.eu/engb/ (ultima consultazione 09/12/2024), riferisce che il 30% degli errori medici è dovuto a carenze nella comunicazione.
[17] J. Thomason, *Big tech, big data and the new world of digital health*, in *Global Health Journal*, Vol. 5/4, 2021, pp. 165-168.
[18] A. Sharma, *Artificial Intelligence-Based Data-Driven Strategy to Accelerate Research, Development, and Clinical Trials of COVID Vaccine.*, in *BioMed research international*, Vol. 2022.
[19] U. Beck, *La società del rischio. Verso una seconda modernità*, Francoforte, 1986.
[20] Art. 10 AI Act.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

una lacuna di dati.

Tutti i sistemi di IA hanno una determinata percentuale di errore sistemico, legato al modo stesso in cui il processo di analisi è condotto, ma tale rischio è minimizzato dall'accuratezza dell'addestramento[21], che dipende, appunto, dalla qualità di dati, non in senso avulso dal contesto di futuro utilizzo dello strumento, ma proprio in stretta correlazione con il bacino della futura utenza, onde evitare bias discriminatori[22] per contesto geografico, linguistico, di genere, di età. Questi bias si riverberano in una scelta errata, perché fondata su presupposti svianti e, nell'attività sanitaria, l'errore colpisce direttamente la salute e la vita delle persone.

La quantità e qualità dei dati necessari all'addestramento del sistema pone problemi in relazione alla violazione del diritto alla privacy, in fase di raccolta o di esecuzione degli algoritmi, per possibile violazione del principio di minimizzazione[23], del diritto all'oblio[24], *data brench* per carenza di cybersecurity[25]; ma anche per la natura stessa dei dati necessari ad esempio per un dispositivo sanitario, che saranno necessariamente dati sensibili, ai sensi dell'art. 9 GDPR, ma il cui utilizzo «è necessario a garantire parametri elevati di qualità e sicurezza dell'assistenza sanitaria e dei medicinali e dei dispositivi medici».

L'altro tema nevralgico è costituito dall'opacità del sistema che, entro un certo margine, è tutt'oggi strutturale per i sistemi di apprendimento automatico[26]. L'opacità è dovuta al fatto che l'algoritmo può identificare le correlazioni tra migliaia di variabili, ma non individua il nesso di causalità e, quindi, non ragiona per deduzione causale, ma per inferenza statistica[27]. Ciò rende non sempre identificabile e ripercorribile il processo logico seguito, nonostante il sistema conservi traccia di ogni processo[28].

L'argine a queste criticità viene individuato, nell'AI Act, nella previsione del controllo umano[29] cui compete l'ultima fase decisionale o, comunque, la verifica sull'iter logico decisionale, secondo lo schema

[21] Nei sistemi di Machine Learning e Deep Machine Learning l'utilizzo dei dati segue la logica GIGO (*garbage – in – garbage -out*). La quantità di dati costituisce il primo presupposto per la qualità. E' infatti nell'immensa quantità di elementi processati che l'IA è in grado di addestrarsi alle correlazioni in modo sempre più raffinato.

[22] Bias discriminatori legati ad incompletezza od inadeguatezza dei dati dell'addestramento sono stati analizzati in diversi campi da A. GUPTA, V.Y. WU, H. WEBLEY-BROWN, J. KING, D.E. HO, *The Privacy-Bias Trade-Off*, in *Policy Brief HAI*, Stanford, 2023. Per uno specifico studio sui bias di genere nei dispositivi medici, si veda M.C. PAGANI, A. ORICCHIO, F. LO IACONO, *L'intelligenza artificiale in campo biomedico: medicina, AI e genere*, in *T4F Series*, 2022.

[23] G. MOBILIO, *L'intelligenza artificiale e le regole giuridiche alla prova: il caso paradigmatico del GDPR*, in *Federalismi*, 16, 2020, 266 – 298.

[24] F. DI CIOMMO, R. PARDOLESI, *Dal diritto all'oblio in Internet alla tutela dell'identità dinamica. È la Rete, bellezza!*, in *Danno e responsabilità*, n. 7, 2012, 706.

[25] Sul punto l'AI Act interviene, all'art. 15, imponendo ai fornitori di sistemi ad alto rischio uno specifico obbligo in termini di accuratezza, robustezza e cibersicurezza.

[26] Sulla differenza tra l'intelligenza artificiale basata sui modelli e quella basata sull'apprendimento e sulle conseguenze in termini di effetto scatola nera, si veda P. TRAVERSO, *op. cit.,* 158.

[27] F.G. MURONE, *Responsabilità medica e Intelligenza Artificiale nel diritto unionale e italiano*, in *IusItinere*, 4 ottobre 2021.

[28] L'art. 20, comma 1, dell'AI Act impone termini e periodi di conservazione dei log generati automaticamente.

[29] Art. 14 AI Act impone ai fornitori di sistemi di IA ad alto rischio di progettare e sviluppare sistemi sorvegliabili dall'uomo durante l'uso. La norma specifica quali sono le misure che garantiscono il controllo umano, indicando plurimi livelli di intervento dell'uomo, a partire dal monitoraggio e fino alla possibilità di interrompere il funzionamento del sistema ad alto rischio.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

del *human in the loop* o del *human in command*[30]. Questa soluzione ha, però, delle difficoltà attuative, che si accentuano nei sistemi ad alto rischio e, tra tutti, nei sistemi impiegati nell'attività sanitaria. E difatti, il controllo umano è astrattamente ipotizzabile ma, all'aumentare della complessità e potenza computazionale, può divenire poco attuabile.

Il medico, dinanzi alle considerazioni in termini di opacità e rischio di bias potrebbe/dovrebbe, nel rispetto dell'art. 14 AI Act, valutare nel caso concreto di non affidarsi alla macchina e discostarsi dall'esito della stessa, facendo prevalere il vaglio umano sull'analisi artificiale. Il pericolo, non troppo nascosto, è ravvisabile in una rischiosa attribuzione, unicamente in capo al medico, dell'onere e della responsabilità della decisione finale[31].

L'altro argine offerto dalla normativa unionale è quello del vaglio di conformità prima dell'immissione nel mercato, secondo una logica di certificazione della sicurezza, in base al principio di prevenzione. La regolazione punta ad una sicurezza *by design*, in un approccio, appunto, preventivo, in cui si valorizza l'*accountability* in capo a fornitore ed utilizzatore. Il sistema che viene immesso nel mercato ha, quindi, i migliori livelli di affidabilità possibili.

Si crea così un doppio binario di protezione avverso i rischi su descritti: certificazione di conformità e sorveglianza umana.

Gli effetti nella pratica della correlazione uomo-macchina, in particolare nella specificità dell'attività sanitaria, si risolvono, allora, in una scelta tra l'approccio che vede l'uomo ultimo decisore, e che apre scenari in tema di responsabilità e, in particolare, di imputabilità da un lato e, dall'altro, di uomo che tende, all'atto pratico, ad affidarsi ad una macchina sicura e, tanto più complessa, tanto meno "sorvegliabile", dando vita, nello specifico campo medico, oltre che a problemi di responsabilità, a due ben noti fenomeni: il *deskilling*[32] e la disumanizzazione del rapporto terapeutico[33].

Accanto ai rischi legati all'implementazione tecnologica nella pratica medica, però, è stato osservato come un disomogeneo utilizzo dell'IA o una scelta radicale di esclusione o limitazione delle applicazioni tecnologiche, è in grado di dar vita a rischi ugualmente rilevanti.

Primo tra tutti: l'espansione del divario digitale, laddove le soluzioni di intelligenza artificiale divengano disponibili solo per sottoinsiemi limitati di pubblico, probabilmente quello che già ha risorse economiche sufficienti per accedere alle cure. In questo modo, il *digital divide* andrebbe solo ad ampliare una già esistente frattura, aumentando le diseguaglianze sociali, in particolare in settori sensibili e vitali come quello del *health care*. Inoltre, l'incapacità di approcciare le potenzialità tecnologiche in modo cauto ma positivo impedisce la diffusione dei benefici derivanti dal progresso, a fronte di una crisi

[30] L. RINALDI, *op. cit.*, p. 214 per una disamina dei diversi livelli di sorveglianza umana.

[31] Si vedano le riflessioni di C. CANULLO, *Chi decide? Intelligenza artificiale e trasformazione del soggetto nella riflessione filosofica*, in E. CALZOLAIO (a cura di), *La decisione nel prisma dell'intelligenza artificiale*, Torino, 2022, 25- 35.

[32] Se la tecnologia venisse percepita come oggettiva e in grado di limitare il margine di errore umano, allora il rischio diventerebbe quello dell'appiattimento della funzione del medico sul risultato dell'AI, in grado di portare al fenomeno del *deskilling*, cioè di perdita di capacità critica e dequalificazione legata ad una *over-reliance* sul risultato meccanico e di una conseguente riduzione della sensibilità diagnostica. Si veda F. PASQUALE, *New laws of robotics. Defending human expertise in the age of AI*, Boston, 2020.

[33] È necessario scongiurare la disumanizzazione del rapporto terapeutico, cioè di quella perdita di quei connotati di dedizione all'unicità del paziente e alla sua persona, prima ancora che alla sua malattia, che costituiscono il cardine della relazione di cura. Si veda P. SOMMAGGIO, *Biodiritto, società, salute*, Torino, 2023.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

globale dei sistemi sanitari nazionali[34], che non potrà avere una efficace svolta se non con l'utilizzo dei sistemi tecnologici, che permettano l'efficientamento delle risorse e la capacità di far fronte a sfide sempre nuove.
Il ruolo della regolazione dell'IA deve quindi tenere presente i rischi d'uso, ma non sottostimare i rischi del non-uso, che impedirebbero la diffusione di benefici ormai non più rinunciabili e che rendono centrale l'attività di policy[35], fondata sulla individuazione di principi condivisi e sovranazionali, idonei ad essere substrato teorico per la costruzione di argini capaci di canalizzare la tecnologia nell'alveo di uno sviluppo equo, etico, sostenibile, conscio dei rischi e capace di affrontarli.

3 Breve premessa sulla fattispecie concreta in esame

L'autismo comprende uno spettro di disturbi caratterizzati da ritardo nello sviluppo linguistico, deficit di interazione sociale e disturbi comportamentali. La prevalenza dell'autismo è aumentata rapidamente negli ultimi anni: le stime aggiornate si attestano su circa 1 su 54 tra i bambini di 8 anni negli Stati Uniti, 1 su160 in Danimarca e in Svezia, 1 su 86 in Gran Bretagna.
In età adulta pochi studi sono stati effettuati e segnalano una prevalenza di 1 su 100 in Inghilterra. In Italia, si stima che circa 1 bambino su 77 (età 7-9 anni) presenti un disturbo dello spettro autistico con una prevalenza maggiore nei maschi: i maschi sono 4,4 volte in più rispetto alle femmine.
Nonostante la crescente prevalenza dell'autismo, l'accesso alle risorse per la diagnosi continua ad essere limitato, sia a livello nazionale che mondiale. Queste inadeguatezze nelle risorse per l'autismo sono aggravate dalla naturale lunghezza dell'iter diagnostico. In media, dal momento delle prime consultazioni con gli operatori sanitari fino al momento della diagnosi decorrono circa 2 anni.
Ritardi così estesi spesso causano la diagnosi in età più avanzata (di solito ≥ 4 anni), che potrebbe comportare impatti maggiori per tutta la vita, incluso una maggiore probabilità di utilizzo di farmaci psicotropi, punteggi di QI più bassi e linguaggio ridotto attitudine. Dato che l'identificazione e l'intervento tempestivi dell'autismo hanno dimostrato di migliorare il successo del trattamento e le capacità sociali, la ricerca si è focalizzata sulla sua individuazione precoce.
Sebbene i sintomi varino da individuo a individuo, le anomalie della prosodia costituiscono un segno notevolmente diffuso di disturbo dello spettro autistico, e numerosi studi confermano la particolare incidenza delle peculiarità del parlato, tra cui ecolalia, intonazione monotona e tono e linguaggio atipici, nei bambini con disturbo dello spettro autistico.
Questa osservazione ha portato i ricercatori a sviluppare un sistema di IA in grado di classificare i suoni

[34] Per citarne alcuni: il *burn out* degli operatori sanitari, la carenza di personale, i tempi di attesa per le cure mediche nei sistemi sanitari nazionali. La tecnologia apre scenari di maggiore efficienza, riduzione del carico di lavoro amministrativo oggi gravante sul personale di cura, con conseguente aumento della disponibilità dei sanitari da impiegare nell'attività professionale specifica. Si veda in proposito AA. VV., *Navigating the clinician shortage crisis,* in *NEJM*, 2024.

[35] Tutti gli stati, ciascuno con un diverso approccio, si sono interrogati sulla policy regolatoria da adottare ed alcuni principi comuni sono gli stessi consacrati nelle sedi sovranazionali. Il 17 maggio 2024 il Comitato dei Ministri del Consiglio d'Europa ha adottato la Convenzione quadro sull'intelligenza artificiale, i diritti umani, la democrazia e lo Stato di diritto. Il 21 Marzo 2024 l'Assemblea generale delle Nazioni Unite ha adottato una risoluzione sulla promozione di sistemi di intelligenza artificiale (IA) sicuri, protetti e affidabili. Il 22 maggio 2019 ben 42 paesi hanno sottoscritto i 5 principi OCSE sull'intelligenza artificiale.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

all'esito di un addestramento sulle anomalie del parlato.
Ricerche precedenti hanno studiato i disturbi prosodici nei bambini con autismo, a vari livelli di successo e hanno sviluppato modelli che, però, utilizzavano dati raccolti in luoghi centralizzati, con apparecchiature di registrazione di alta qualità. Queste ricerche, concretamente, non accelerano in modo significativo il processo di *screening* per l'autismo, perché richiedono ancora l'uso di attrezzature specializzate e di luoghi di registrazione centralizzati per essere forniti di qualità audio costante. Inoltre, l'interazione vocale avviene con estranei e fuori dall'ambito domestico.
Partendo da questi presupposti, un *pool* di ricercatori della Standford University[36] ha messo a punto un sistema di intelligenza artificiale con apprendimento automatico per la rilevazione e classificazione delle anomalie della prosodia, finalizzata allo *screening* per l'autismo infantile attraverso l'utilizzo di una *app* di gioco vocale, in ambito domestico.
A differenza di altri studi, questo approccio non richiede apparecchiature di registrazione specializzate ad alta fedeltà e utilizza modelli linguistici naturalistici, registrando i bambini che giocano con giochi educativi con i genitori, in un ambiente domestico a basso stress. Infine, il sistema non richiede la presenza di personale medico qualificato durante l'utilizzo della *app*.

4. La disciplina del dispositivo sanitario con AI nel Reg. 2017/745

L'immissione sul mercato, la messa a disposizione e la messa in servizio nel territorio italiano dei dispositivi medici e dei dispositivi medico-diagnostici in vitro è subordinata alla conformità ai requisiti applicabili del Regolamento (UE) 2017/745, per i dispositivi medici, e ai requisiti applicabili del Regolamento (UE) 2017/746, per i dispositivi medico-diagnostici in vitro, che hanno modificato le norme previgenti[37], tenendo conto degli sviluppi del settore negli ultimi vent'anni.
I due regolamenti prescrivono requisiti di sicurezza e prestazione applicabili ai prodotti, e permettono di presumere il possesso di tali requisiti attraverso il rispetto di norme tecniche armonizzate[38], di successiva redazione, estranee al testo normativo, che risulta così tecnologicamente neutro ed in grado di sfuggire all'obsolescenza cagionata dal rapido e continuo avanzamento tecnologico.
Il Regolamento introduce la definizione di *software as medical device* con l'art. 2 lett. a), che qualifica come dispositivi medici «tutti i *software*[39] aventi finalità di diagnosi e cura e altresì i *software* che supportano l'operatore sanitario ad assumere una decisione terapeutica oppure aiutano l'erogazione della prestazione stessa». Quel che rileva per la classificazione come dispositivo medico è quindi la

[36] N.A. Chi, P. Washington, A. Kline, A. Husic, C. Hou, C. He, K. Dunlap, D.P. Wall, *Classifying Autism from Crowdsourced Semi-Structured Speech Recordings: A Machine Learning System*, in *JMIR*, 2021.
[37] I regolamenti andranno progressivamente a sostituire le direttive 93/42/CEE, 90/385/CE e 98/79/CE in vigore da oltre 20 anni.
[38] Le norme tecniche possono essere suddivise in tre categorie: norme internazionali, elaborate dall'ISO o, per il settore elettrico, dall'IEC; norme europee, elaborate da CEN o CENELEC; norme nazionali, elaborate, in Italia, da UNI o CEI. Le norme tecniche diventano "armonizzate" quando sono adottate a livello europeo, con la pubblicazione in Gazzetta Ufficiale dell'Unione Europea (OJEU); solitamente costituiscono l'adozione in campo europeo di norme internazionali (ISO o IEC), talora con eventuali adattamenti. La disciplina delle norme armonizzate è dettata dal Reg. (UE) 1025/2012.
[39] Questa definizione ricomprende sia i software che operano autonomamente (cd. *stand-alone*), sia i *software embedded*, ossia inclusi in un altro dispositivo medico *hardware*.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

destinazione d'uso e la funzionalità, per come stabilite dal produttore.
I dispositivi sono suddivisi in quattro classi di rischio, I, IIa, IIb e III, in funzione della destinazione d'uso e dei rischi che questa comporta, valutati[40] in modo crescente, da quelli meno critici, non attivi e non invasivi a quelli ad alto rischio e che interagiscono sulle funzioni di organi vitali. La regola 11 dell'allegato VIII prevede però specifici criteri per la classificazione del rischio del SAMD.
La procedura di valutazione della conformità può richiedere l'intervento di un Organismo notificato e, come già osservato, il rispetto della normativa armonizzata opera come presunzione di conformità.
Allo stato attuale, però, le norme tecniche armonizzate, approvate in attuazione della precedente Dir. 93/42/Cee, non sono ancora state aggiornate[41] e, pertanto, l'effetto presuntivo delle norme tecniche armonizzate non è effettivo.
E difatti l'art. 8, Reg(EU) 2017/745, prevede l'ultrattività delle vecchie norme tecniche armonizzate; ma l'art. 3 della Decisione di Esecuzione 2020/437 Comm EU del 24 marzo 2020, ne vieta l'utilizzo al di fuori dei cd. *dispositivi legacy*[42].
Altro aspetto rilevante della normativa è costituito dall'art. 2 comma 53, che impone, affinché sia autorizzata la commercializzazione di un dispositivo medico, che ne sia dimostrato il cd. beneficio clinico, ossia il prodotto, oltre che sicuro, deve risultare clinicamente efficace.
In ambito nazionale, l'entrata in vigore dei Regolamenti ha portato all'adozione dei decreti legislativi 137/2022 e 138/2022, che all'art. 22 e all'art. 18 individuano obiettivi, condizioni e ricadute per la valutazione delle tecnologie sanitarie, attraverso il programma nazionale di HTA.
La valutazione delle tecnologie sanitarie (Health Technology Assessment - HTA) è, infatti, un processo multidisciplinare che sintetizza le informazioni sulle questioni cliniche, economiche, sociali ed etiche connesse all'uso di una tecnologia sanitaria, in modo sistematico, trasparente, imparziale e solido.
Il processo di HTA si basa su evidenze scientifiche tratte da studi, che vengono considerate per specifiche tipologie di intervento sanitario su determinate popolazioni di pazienti, confrontando gli esiti e i risultati con quelli di tecnologie sanitarie di altro genere o con lo standard di cura corrente.
Per ciascuna tecnologia oggetto di valutazione (rapporto di HTA), la Cabina di regia adotta un giudizio di *appraisal*, contenente preliminari raccomandazioni di utilizzo nell'ambito del Servizio Sanitario Nazionale (utilizzo, non utilizzo, utilizzo in ricerca, utilizzo condizionato)[43].
Si deve poi sottolineare che la novella normativa ha modificato lo status della struttura sanitaria che

[40] La classificazione è effettuata dal fabbricante secondo i criteri dell'Allegato VIII del Regolamento (UE) 2017/745.
[41] Il 14 aprile 2021 la Commissione, con la decisione M/575, ha prorogato sino al 31 dicembre 2024 il mandato al CEN e al CENELEC per l'armonizzazione delle norme tecniche per la conformità dei dispositivi medici ai sensi del Reg. 745/2019. https://ec.europa.eu/transparency/documents-register/detail?ref=C(2021)2406&lang=en (ultima consultazione 2/12/2024).
[42] Si tratta di dispositivi di classe I ai sensi della Direttiva 93/42/CEE, per i quali è stata redatta una dichiarazione di conformità prima del 26 maggio 2021 e per i quali la procedura di valutazione della conformità in base al Regolamento (UE) 2017/745 richiede il coinvolgimento di un organismo notificato oppure di dispositivi con un certificato valido rilasciato ai sensi della Direttiva 90/385/CEE o della Direttiva 93/42/CEE, per i quali sussistono due condizioni: i dispositivi in questione continuano a essere conformi alla direttiva pertinente e non sono stati introdotti cambiamenti significativi nella progettazione e nella destinazione d'uso.
[43] Per una disamina della procedura HTA si veda A. Di Martino, *Intelligenza artificiale e responsabilità civile in ambito sanitario*, Milano, 2022.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

realizza in modo autonomo un dispositivo medico destinato alla pratica clinica interna. Essa viene infatti assimilata ad un produttore, assoggettata a tutti gli obblighi del regolamento europeo e con l'obbligo di utilizzare una soluzione realizzata da un fabbricante esterno e di poter ricorrere ad un dispositivo di propria creazione solo dimostrando l'esistenza di specifiche esigenze in capo ad un gruppo di pazienti, tali da non poter essere soddisfatte se non con un dispositivo medico creato *ad hoc*.
In conclusione, quindi, si può osservare che il Reg. (UE) 745/2017 non prevede alcuna specifica norma in merito a conformità, standard tecnici, HTA e valutazioni post-market che tengano conto delle specificità dei dispositivi medici in cui è impiegato un sistema di intelligenza artificiale. Le uniche innovazioni hanno riguardato: la definizione di dispositivo medico – per ricomprendere il *software* con destinazione d'uso medica - e la classificazione del rischio – ricalcolata in aumento, per i SAMD -, lasciando poi, a tutt'oggi, la normativa indifferente rispetto alle peculiarità dello strumento di AI, caratterizzato, rispetto ad altri strumenti, dalla adattatività sulla base di un apprendimento automatico.

5. La disciplina del SAMD alla luce dell'AI ACT

Il dispositivo medico dotato di intelligenza artificiale soggiace, oltre alla disciplina del Reg. 745/2017, anche alle normative unionali in quanto prodotto e in quanto bene munito di AI. Entrano quindi in gioco altre discipline, verticali e trasversali, che dovranno essere oggetto di raccordo da parte del legislatore europeo e che, al contempo, richiederanno agli interpreti una lettura armonica ed integrata[44].
Il cd. AI ACT costituisce il primo atto normativo, a livello globale, che tenti di regolare in modo trasversale i sistemi di intelligenza artificiale, definendo loro e gli operatori della filiera, categorizzandoli in base ai livelli di rischio, sulla scorta del principio di precauzione, e disciplinando la certificazione e messa in commercio dei prodotti di intelligenza artificiale.
Il Regolamento, infatti, individua quattro classi rischio: inaccettabile, elevato, medio/basso, sistemico nei modelli generali. I prodotti con rischio inaccettabile sono vietati. I prodotti ad alto rischio possono essere immessi in commercio ed utilizzati solo a seguito di una valutazione di conformità, che nasce al momento dell'immissione nel mercato e prosegue con una verifica post-market costante, necessitata dalla modificabilità nel tempo, dovuta alla capacità evolutiva e di (auto)apprendimento del sistema. I sistemi a basso rischio sono soggetti a doveri di trasparenza, a vigilanza post market e se ne incoraggia la disciplina attraverso codici di condotta. I cd modelli generali vengono regolati nei soli casi di rischio sistemico.
La normativa persegue lo scopo di migliorare il funzionamento del mercato interno, promuovere l'adozione di una intelligenza artificiale antropocentrica ed affidabile, supportare l'innovazione ed assicurare alti livelli di protezione di salute, sicurezza e diritti fondamentali previsti dalla Carta dei Diritti Fondamentali, tra cui: democrazia, stato di diritto, protezione dell'ambiente, dagli effetti dannosi dei sistemi di AI.
Accanto a tali norme vi sono specifiche misure per sostenere l'innovazione, in specie per le PMI e le *start-up*.

[44] In particolare, si fa riferimento al quadro normativo unionale in materia: sicurezza del prodotto, conformità del prodotto, intelligenza artificiale e privacy. Si veda G.F. SIMONINI, *La responsabilità del fabbricante nei prodotti con sistemi di intelligenza artificiale*, in *Danno e Responsabilità*, 4, 2023, 435.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

L'ambito di applicazione si estende sia ai produttori e distributori con sede nell'Unione Europea, sia a soggetti con sede extra Unione, ma il cui prodotto è destinato al mercato dell'Unione nonché a tutti i casi in cui i danneggiati sono collocati nell'Unione Europea.
Il Regolamento non si applica ai sistemi per scopo esclusivo di ricerca scientifica e ai sistemi in *free and open source licences*, salvo che non siano sistemi classificati ad alto rischio o rientranti tra gli usi vietati, e ad eccezione degli obblighi del produttore, previsti dall'art. 50 del Regolamento medesimo.
L'art. 2 prevede espressamente che il Regolamento non infici le altre norme unionali relative alla tutela dei consumatori, alla sicurezza del prodotto e alla privacy, con ciò espressamente imponendo un raccordo tra i vari testi normativi.
Allo scopo, poi, di compiere questo raccordo, all'allegato I vengono elencate una serie di normative e l'art. 6 del regolamento stabilisce che i prodotti utilizzati come componente di un prodotto disciplinato da una delle norme armonizzate di cui all'allegato I sono da considerare sistemi ad alto rischio.
L'art. 8 prevede, poi, che laddove queste normative armonizzate assoggettino i prodotti a procedure autorizzative, i fornitori possono scegliere di integrare la documentazione necessaria ai fini dell'AI Act all'interno della procedura autorizzativa specifica della propria normativa di settore.
Il Reg. 745/2017 è una delle norme armonizzate richiamate dall'allegato I e, pertanto, i dispositivi medici dotati di intelligenza artificiale ricevono due diverse classificazioni di rischio: una in base al Reg. 745/2017 e una in base all'AI Act, che li qualifica tutti e necessariamente ad alto rischio. Ugualmente, si sdoppiano i percorsi di valutazione della conformità, che, basandosi su criteri non necessariamente equiparabili, vengono raccordati attraendo la competenza a valutare la conformità in capo all'Organismo notificato della eventuale normativa speciale.
La disciplina dell'AI Act espressamente richiama il GDPR e sostiene la sua piena applicabilità anche ai prodotti con intelligenza artificiale, ma non detta specifiche norme di raccordo in proposito.
Le problematiche del difetto del prodotto e della conseguente responsabilità restano totalmente demandate all'attuale direttiva sul prodotto difettoso, la cui revisione è in corso, e ad una futura direttiva relativa all'adeguamento delle norme in materia di responsabilità civile extracontrattuale all'intelligenza artificiale.
Un dispositivo medico dotato di IA è quindi sottoposto, per il profilo della valutazione di sicurezza e conformità e la conseguente marcatura CE, agli specifici Regolamenti UE, alle norme nazionali di attuazione di tali regolamenti, e – nel prossimo futuro -alla disciplina dell'AI Act; quanto ai profili di responsabilità civile è invece soggetto alla direttiva sui prodotti difettosi, alle singole discipline nazionali in tema di responsabilità civile e nel futuro, alle norme dell'elaboranda direttiva, tese ad armonizzare i regimi nazionali in modo da apprestare una tutela comune minima.

6. Capacità dell'attuale normativa di contenere i rischi di utilizzo dell'AI in ambito sanitario

Alla luce della breve panoramica normativa tratteggiata, il dispositivo medico attualmente oggetto di ricerca per la diagnosi di autismo infantile sembra incarnare un modello che attualizza le riflessioni sui rischi e benefici dell'AI applicata all'*health care* e che, pertanto, può aiutare a verificare l'efficacia della cornice normativa.
Questo strumento di diagnosi, infatti, incide in un settore caratterizzato da barriere all'accesso alla

Downloaded from www.biodiritto.org.
ISSN 2284-4503

diagnosi e eccessiva durata dell'iter diagnostico, con riflessi negativi sul decorso della terapia, ed aggravamento di una disabilità pervasiva.

L'*app* attualmente allo studio dei ricercatori di Stanford ha il plurimo pregio di: abbattere tempi e costi della diagnosi, rendendola domiciliare. Inoltre, permette di far interagire il bambino in un contesto familiare e con soggetti familiari, eliminando i bias dovuti alla difficoltà, tipica dei soggetti con autismo, a generalizzare comportamenti in contesti nuovi e con soggetti estranei.

Non solo, quindi, una facilitazione dell'accesso alla diagnosi, ma anche la strutturazione di un contesto più favorevole ad una diagnosi corretta.

L'*app*, inoltre, affronta il problema della carenza di personale medico specializzato, perché elimina la necessità della supervisione specialistica nel momento dell'utilizzo del gioco educativo e la riduce alla fase di valutazione dei dati raccolti ed elaborati.

Come è paradigmatico dei benefici, così il SAMD in esame è paradigmatico dei rischi.

In primo luogo, il set di dati con cui la macchina deve essere addestrata è estremamente sensibile al contesto di futuro uso, per area geografica, lingua, genere, età. Un set non completo e non rappresentativo produrrà senza dubbio dei bias discriminatori, falsando i risultati e sviando il medico che dovrà, infine, formulare la diagnosi.

La necessità di trattare dati biometrici e sensibili pone poi il problema della loro conservazione, che potrebbe però ritenersi superato attraverso il processo di anonimizzazione, anche alla luce dei recenti arresti della Corte di Giustizia UE[45].

Si concretizza il rischio legato alla fallibilità sistemica dello strumento, difficile da valutare a causa del cd. effetto *black box*, ed idonea a provocare un errore diagnostico su un soggetto già particolarmente vulnerabile.

La strada per ridurre l'incidenza dannosa e permettere l'effettività del controllo umano appare essere allora quella di non fondare la diagnosi unicamente sulla valutazione effettuata dal sistema artificiale, ma utilizzarlo come una sorta di primo screening, per concentrare le risorse mediche solo sui soggetti che, ad una prima analisi, ricevono una diagnosi positiva.

In tal modo, il dispositivo sarebbe solo un facilitatore, ma non sostituirebbe l'iter diagnostico ordinario.

La normativa regolatoria attualmente esistente sembrerebbe supportare questa impostazione, giacché in tal modo lo strumento sarebbe rispettoso sia del divieto di profilazione automatica, di cui all'art. 22 GDPR, sia dell'obbligo di sorveglianza, di cui all'art. 14 AI Act.

Difatti, il dispositivo, per essere immesso nel mercato ed applicato, deve essere certificato conforme in base al Reg. 745/2017, perché integra le caratteristiche del *software* come dispositivo medico, ai sensi dell'art. 2 citato; questa conformità, in ossequio al disposto dell'art. 8 AI Act, non dovrà "limitarsi" ai requisiti previsti per la specifica classe di dispositivo medico, ma dovrà certificare anche la conformità alla Sezione 2 del Titolo 2 dell'AI Act.

La coesistenza delle due norme permette di conciliare la specificità delle verifiche effettuate in sede di HTA, necessarie per l'immissione in commercio di qualsiasi dispositivo medico, con le verifiche proprie di un *software* di IA, che deve essere esaminato anche per quel che riguarda l'adeguatezza dei dati di

[45] Con sentenza del 26 aprile 2023 (causa T-557/20), la Corte di Giustizia dell'Unione Europea ha stabilito che un dato pseudonimizzato trasmesso ad un destinatario che non ha i mezzi per poter identificare l'interessato non è un dato personale.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

addestramento, come previsto dall' art. 10 AI Act.
In mancanza di una revisione del Reg (UE) 745/2017, l'AI Act appare allora fornire una cornice normativa che introduce dei principi e dei controlli che, altrimenti, non si sarebbero effettuati.
V'è da dire che il raccordo tra le due discipline non appare sviluppato.
E difatti, oltre ad inserire il medesimo apparecchio in due categorie di rischio diverse, i due regolamenti fanno anche riferimento a norme armonizzate diverse e, a tutt'oggi, assenti.
Le norme armonizzate, a ben vedere, sono però il punto nevralgico della disciplina[46], perché fungono da schema su cui i singoli fornitori del bene possono strutturare i sistemi di prevenzione e controllo, conformandosi a dei livelli tecnici di *accountability* che concretizzano il concetto di sicurezza.
La standardizzazione, ad opera di organismi di normazione, ha inoltre il pregio di essere sovranazionale e, quindi, di poter diffondere in modo globale un modello di livelli di tutela, massimizzando gli effetti della norma regolatoria di principio.
E' quindi tutto in divenire il rapporto che si creerà tra norme armonizzate sui dispositivi medici e norme armonizzate in tema di sicurezza del prodotto di IA e se le prime saranno in grado di assorbire le seconde, rendendo più lineare il processo autorizzativo del dispositivo.
Quanto all'idoneità della cornice normativa a contenere i rischi dell'utilizzo dell'IA nel settore sanitario e, in specie, in quello dei dispositivi medici, sorgono alcune considerazioni.
Sotto il profilo della sicurezza del prodotto, la norma recepisce dei principi ormai consolidati, che si concretizzeranno solo con l'attuazione della delega alla Commissione e l'elaborazione delle norme armonizzate. La particolarità dello sforzo regolatorio unionale è nel partire dalla centralità dei diritti fondamentali e renderli un parametro essenziale della valutazione del rischio. E' un approccio generale, trasversale, in cui innovazione e mercato non sono valori assoluti, ma sono informati da una regolazione che tutela il minimo etico[47].
Nel concreto, però, questi criteri dovranno poi essere attuati nelle norme armonizzate che, forse, saranno sia strumento pratico sia banco di prova della bontà dell'approccio regolatorio prescelto dal legislatore unionale e della sua effettività.
Il punto che, però, appare nodale, è quello della responsabilità, non solo in caso di prodotto difettoso, ma anche in caso di prodotto sicuro e conforme[48], ma, allo stato dell'arte attuale, non infallibile, al di là (e nonostante) la diligenza del professionista.

7. La disciplina della responsabilità de iure condendo. Riflessioni conclusive

Il tema della responsabilità, che il dibattito attualmente riconduce, in sede europea, alla direttiva sui prodotti difettosi e, in sede nazionale, alle figure codicistiche tipiche di responsabilità speciale, è il punto di snodo per l'effettiva diffusione dei SAMD nella pratica quotidiana dell'attività sanitaria.
Da più parti è maturata e sottolineata la necessità di un intervento normativo *ad hoc*, che raccolga la specificità del mezzo tecnologico e l'ulteriore particolarità di dover valutare una condotta che è frutto

[46] E. Bellisario, *La rilevanza del criterio presuntivo della conformità alle norme armonizzate*, in *Persona e Mercato – Saggi*, 156.
[47] E. Bocchini, *La regolazione giuridica dell'intelligenza artificiale,* Torino, 2024.
[48] G. Guerra, *Il concetto di difettosità nella realtà che cambia. Un esercizio di microcomparazione*, Napoli, 2019.

Downloaded from www.biodiritto.org.
ISSN 2284-4503

della relazione tra uomo e macchina[49].

Il dibattito è aperto e vede la propria principale sede nella proposta di Direttiva sulla responsabilità del prodotto con IA.

Molteplici sono i profili di discussione: l'imputabilità, il nesso di causalità in relazione all'onere della prova, resa diabolica dall'opacità sistemica dello strumento, l'oggettivizzazione della responsabilità. La specifica fattispecie qui brevemente in esame, pur essendo toccata da tutte le sfaccettature delle problematiche inerenti la responsabilità, pone particolarmente in rilievo il profilo della sorvegliabilità umana e dell'imputabilità dell'errore diagnostico.

Si tratta, infatti, di un dispositivo in grado di formulare una diagnosi, sia anche come primo screening e, quindi, di commettere errori, anche omissivi. La capacità dello strumento di velocizzare i tempi dell'accertamento sanitario e di abbattere le barriere all'accesso alla diagnosi risiede proprio nella sua possibilità di utilizzo domiciliare, non sorvegliato durante l'esecuzione.

Pur se il completamento del percorso resta affidato al professionista sanitario, l'iter diagnostico risulta davvero realizzato attraverso una correlazione tra macchina ed uomo, tant'è che la decisione del secondo si fonda sui presupposti elaborati dal primo.

Le scelte normative in tema di responsabilità diventano allora nodali, perché l'eccessivo carico di responsabilità sul professionista[50], in virtù di una sorveglianza astrattamente possibile, potrebbe spingere verso il non uso dello strumento, che costituisce una scelta non esente da rischi, perché comunque foriera di danni, solo di diversa specie, probabilmente più a lungo termine e, quindi, meno impattanti *prima facie*. Al contempo, la deresponsabilizzazione del sanitario apre scenari inaccettabili di diminuzione di tutela del diritto alla salute, e proprio a discapito di soggetti più vulnerabili, e innesca un circuito di *deskilling* e disumanizzazione delle cure, in grado di disperdere i benefici potenziali dello strumento tecnologico.

La proposta di direttiva 2022 concentra la propria attenzione soltanto sui sistemi ad alto rischio, tra cui rientrano tutti i dispositivi medici, e tratteggia un affievolimento dell'onere probatorio del danneggiato, con presunzione del nesso di causalità in caso di colpa del produttore o dell'utilizzatore del sistema di intelligenza artificiale, già prospettando una possibile revisione in senso di oggettivizzazione della responsabilità, dopo cinque anni dall'entrata in vigore del testo normativo.

È una soluzione, qualora si dovesse convergere su un testo del genere, che sembrerebbe rimandare il confronto su alcuni punti nevralgici, tra cui, appunto, il margine di errore sistemico e la attuabilità di una totale sorveglianza umana, che possa arginare tale errore.

È probabilmente su questo fronte che si dovrà misurare il bilanciamento degli interessi in gioco per sbloccare effettivamente il potenziale di sviluppo dell'IA nell'attività sanitaria, perché è attraverso il sistema rimediale che i principi di tutela della salute e dei diritti fondamentali diventano effettivi.

[49] G. Teubner, *Soggetti giuridici digitali? Sullo status privatistico degli agenti software autonomi*, Napoli, 2019.
[50] M.M. Mello, N. Guha, *Understanding Liability Risk from Using Health Care Artificial Intelligence Tools*, in *NEJM*, 2024.

Downloaded from www.biodiritto.org.
ISSN 2284-4503