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.

SOMMARIO: 1. Introduzione – 2. I sistemi di IA vietati e le ampie maglie del concetto di vulnerabilità – 3. I sistemi ad alto rischio e la violazione in concreto del principio di parità – 4. (segue ...) La valutazione di impatto ai sensi dell'art. 27 del Regolamento sull'IA – 5. L'altra faccia della medaglia: l'*alfabetizzazione* paritaria dell'IA.

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.*

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).

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».

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).

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).

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.).

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).

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).

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.