Essays

# Big brother is «*gendering*» you. Il diritto antidiscriminatorio alla prova dell'intelligenza artificiale: quale tutela per il corpo digitale?

***Luca Giacomelli****

BIG BROTHER IS «GENDERING» YOU. ANTI-DISCRIMINATION LAW AND THE CHALLENGES OF ARTIFICIAL INTELLIGENCE: WHAT PROTECTION FOR THE DIGITAL BODY?

ABSTRACT: Recent advancements in artificial intelligence are revolutionizing how easily and readily organizations can collect data and perform «data-driven» decisions across institutional contexts. Companies and institutions can now link a great variety of data sources, sometimes innocuous on their own but not in the aggregate, to inform an increasingly broad range of decisions tied to activities like credit reporting, advertising, hiring, judging. In this article, I analyse how data-driven decisions can discriminate by explaining how even unprejudiced algorithms and decision-makers can generate biased decisions and I try to verify the effectiveness of the anti-discrimination law categories in the face of discriminatory results of automated decisions. Although many risks of big data are well-known, other problems can arise from the refusal to acknowledge or collect certain data. In fact, under the idea of AI neutrality, we end to ignore or hide, rather than prevent, discrimination, because decisions can be biased even in the absence of socially disadvantaged groups data. This leads us asking the following questions: What are the legal remedies to unmask the discrimination of an algorithmic decision? How can we protect our privacy and fundamental rights?

KEYWORDS: Anti-discrimination law; privacy; algorithms; big data; comparative law

SOMMARIO: 1. Premessa: ma gli androidi sognano pecore elettriche? – 2. *Big data*, *big concerns*: ha ancora senso parlare di privacy? – 3. Dietro il velo della neutralità: algoritmi e profilazione – 4. Il diritto antidiscriminatorio alla prova dell'intelligenza artificiale: uno sguardo comparato – 4.1 I paradossi del *disparate impact* nell'esperienza americana – 4.2 La "recidività" del pregiudizio: la lezione canadese – 4.3 GDPR e algoritmi discriminatori: un'arma spuntata? – 5. Osservazioni conclusive.

* *Assegnista di Ricerca in Diritto comparato, Università di Firenze. E-mail: luca.giacomelli@unifi.it.* *Contributo sottoposto a doppio referaggio anonimo.*


ISSN 2284-4503

Essays

*La civiltà sta producendo macchine che si comportano come uomini*
*e uomini che si comportano come macchine.*
(Erich Fromm)

## 1. Premessa: ma gli androidi sognano pecore elettriche?

Se gli androidi di Philip Dick sognavano pecore elettriche, gli attuali algoritmi di intelligenza artificiale che cosa sognano? E, prima ancora, sognano? A chiederselo è il sociologo Dominique Cardon e la sua risposta forse non piacerà agli irriducibili amanti del fantascientifico: «i clic degli internauti producono popolarità, le citazioni ipertestuali autorità, gli scambi tra cerchie affini reputazione, le tracce dei comportamenti predizioni personalizzate ed efficaci»[1]. Di questo si tratta. Più che pensare, o addirittura provare emozioni, le odierne intelligenze artificiali sono infaticabili *concierge*, capaci di gestire e processare archivi giganteschi di dati (c.d. *big data*) per trasformarli in maniera ultrarapida in informazioni utili a migliorare la nostra vita. Attraverso la raccolta e la classificazione delle informazioni, la personalizzazione della pubblicità, i suggerimenti negli acquisti, la profilatura dei comportamenti, i computer si immischiano, sempre di più, nel quotidiano delle persone fino a diventare incontestabili e persino invisibili.

Ma, mentre siamo subito in grado di cogliere i lati positivi di una vita quotidiana che si fa sempre più *smart*, lo siamo molto meno nel domandarci quale prezzo stiamo pagando per tutto questo. Il lato oscuro della tecnologia informatica include per esempio la rinuncia a una fetta della nostra privacy, della nostra libertà, della nostra identità. Ma non soltanto: l'utilizzo sempre più diffuso degli algoritmi di intelligenza artificiale, a vari livelli e in diversi settori, presenta il rischio di decisioni discriminatorie e irragionevoli, con delicate implicazioni di carattere sociale, economico e anche politico, laddove non si conoscano e non si riescano a disciplinare i meccanismi posti alla base dell'effettivo funzionamento della "scatola nera".

Naturalmente alla base di tutto ci sono i dati sui quali operano questi algoritmi che non sono affatto la formula magica che produce automaticamente benefici in assenza di un quadro accurato di regole, anche di carattere etico. Al pari di ogni decisione umana, anche la decisione informatizzata può essere condizionata da errori che comportano esiti imprevedibili e talora discriminatori, soprattutto nei confronti di determinate categorie sociali. Il problema è che di questo non ce ne rendiamo conto e, anche quando accade, non abbiamo gli adeguati strumenti giuridici in grado di correggere tale malfunzionamento. Succede, pertanto, che digitando su *Google* «ragazze nere», anziché «ragazze bianche», i risultati della ricerca cambino radicalmente, suggerendo siti pornografici nel primo caso[2]. Allo stesso modo, se quando digitiamo il nome di alcune celebrità, il motore di ricerca ci suggerisce di ag-

[1] D. Cardon, *Cosa sognano gli algoritmi. Le nostre vite ai tempi dei big data*, Milano, 2016, 84.

[2] «*All search results are not created equal*». Questa è la premessa dell'interessante studio empirico di S. U. Noble, *Algorithms of oppression. How search engines reinforce racism*, New York, 2018, 4, in cui l'Autrice evidenzia come la combinazione di interessi privati ed economici nel promuovere certe attività e certi siti internet e il quasi monopolio dei dati da parte di un ristretto numero di motori di ricerca alimenti la discriminazione e il pregiudizio nei confronti di certe categorie di persone, specialmente nei confronti delle donne di colore. Dietro il falso mito che gli algoritmi di intelligenza artificiale siano neutrali e diano eguale importanza a ogni tipo di idea, identità e attività, si nasconde una cultura di razzismo, sessismo e discriminazione che viene quotidianamente riproposta e rinforzata.

giungere la parola «ebreo» oppure «gay», è perché molti altri utenti lo hanno già fatto. Si può allora dire che il motore di ricerca è discriminatorio? Usa uno schedario che divide le categorie «bianchi/neri», «eterosessuali/omosessuali» e via dicendo? L'algoritmo non ha bisogno di avere un'intenzione discriminatoria per produrre quel genere di risultati. Non contiene norme che gli chiedono di individuare le persone nere e le persone bianche. Si accontenta di lasciar fare alle regolarità statistiche che dicono che cognomi e nomi di persone nere sono statisticamente più spesso legati a clic verso ricerche sulle fedine penali. Lasciato a se stesso, il calcolatore si basa sui comportamenti degli altri internauti e contribuisce, innocentemente se così si può dire, alla riproduzione della struttura sociale, delle disparità e delle discriminazioni.

Gli algoritmi non si sognano di discriminare. Però l'apprendimento automatico, alla base di questi *software* intelligenti, impara dai dati. E quelli oggi a disposizione non rappresentano ugualmente la popolazione: da un lato, perché l'organizzazione e l'elaborazione delle informazioni avvengono attraverso la categorizzazione dei dati rispetto a standard e categorie predeterminate; dall'altro lato, perché chi crea i *software* non è rappresentativo delle diversità presenti nella società e quindi non è in grado di accorgersi del problema. Pertanto, se i set di dati sfruttati dalle intelligenze artificiali sono distorti per quanto riguarda informazioni sensibili come genere o razza, potrebbero derivarne decisioni discriminatorie, sia di tipo diretto qualora la decisione sia presa sulla base di attributi sensibili stigmatizzati, sia di tipo indiretto qualora la decisione sia presa sulla base di attributi non sensibili ma fortemente correlati con attributi sensibili stigmatizzati. Un *software* di apprendimento alimentato con *input* intrisi di pregiudizio non può che consegnare risultati anch'essi intrisi di pregiudizio.

Ben lontani da essere semplici strumenti tecnici neutrali, gli algoritmi sono portatori di pregiudizi. Sottovalutare o, peggio, mascherare dietro la presunta neutralità delle macchine gli effetti della discriminazione digitale (*abstract or data discrimination*) significa creare una società dove decisioni automatiche apparentemente oggettive limitano l'accesso alle opportunità e ai benefici dell'intelligenza artificiale a pochi privilegiati. Comprendere la logica, i valori e il tipo di società che promuovono significa fornire agli utenti di internet i mezzi di riconquistare potere nella società digitale.

## 2. Big data, big concerns: ha ancora senso parlare di privacy?

Viviamo in una società digitale[3]. Con l'avvento di internet e l'evolversi repentino delle tecnologie informatiche si è giunti a una perfetta convergenza tra mondo fisico e mondo virtuale, con mutamenti

[3] Cfr., tra gli altri, D. LYON, *Surveillance Society. Monitoring Everyday Life*, Londra, 2001, per il quale non soltanto viviamo in una società digitale ma anche in una società della sorveglianza e S. RODOTÀ, *Il mondo nella rete. Quali i diritti, quali i vincoli*, Roma-Bari, 2014, il quale invece parla della necessità di una «cittadinanza digitale» che tuteli il nostro accesso alla rete e il nostro «corpo elettronico», così come del bisogno di una tutela un tempo impensabile, il diritto all'oblio e alla cancellazione dei dati personali, per esempio. Questa è una consapevolezza che già da tempo hanno maturato gli scienziati dell'informatica: si veda, per esempio, D. PEDRESCHI, F. GIANNOTTI (a cura di), *Mobility Data and Privacy: Geographic Knowledge Discovery*, New York, 2007, 53, in cui Pedreschi parla di una società misurabile, in cui ogni attività umana lascia tracce digitali e può essere studiata in modo quantitativo e qualitativo. «Tutti noi siamo "pollicini digitali": ogni giorno lasciamo una quantità enorme di informazioni che, se aggregate, possono offrire una rappresentazione al microscopio della persona e della società. Questi dati rappresentano un tesoro informativo che permette di misurare la complessità del mondo interconnesso che viviamo oggi».

radicali nella società e nell'economia globale, costrette *obtorto collo* a ridefinire completamente i propri schemi e il proprio funzionamento. Il facile accesso a ingenti masse di dati, con il passaggio dalla "società della comunicazione" alla "società dell'informazione"[4], ha infatti condotto a una rimodulazione non solo delle modalità di informazione, di comunicazione e delle relazioni sociali e politiche, ma anche dei tradizionali modelli di *business* del mercato. Lo dimostra, per esempio, lo sviluppo del c.d. "*Internet of Things*", ossia la progressiva digitalizzazione della maggioranza dei dispositivi utilizzati nella vita quotidiana – dal frigo che suggerisce quali cibi consumare per primi e quali da acquistare al supermercato agli orologi che monitorano i nostri passi segnalandoci eventuali problemi di salute – che di fatto raccolgono, immagazzinano e processano dati, direttamente o mediante ulteriori dispositivi[5]. Chiamati da alcuni «i giacimenti petroliferi del Terzo Millennio»[6], questi giganteschi agglomerati di dati (*big data* o mega dati) possono essere considerati come frammenti elementari di informazioni, molto spesso di carattere personale, che per poter essere letti e analizzati richiedono il ricorso a potenti processori, gli algoritmi informatici. Questa nuova categoria di "beni", pertanto, acquista valore soltanto grazie all'utilizzo di sistemi algoritmici computazionali altamente predittivi che consentono di ricavarne in tempo reale "conoscenza", offrendo molteplici opportunità sia per le autorità pubbliche, sia per le imprese[7]. Le prime possono usufruire delle tecnologie connesse ai mega dati per finalità legate alla sicurezza pubblica e alla prevenzione del crimine[8], per il miglioramento dei servizi pubblici, per il potenziamento della salute, dell'istruzione e della cultura ma anche – a seconda del contesto politico – per finalità meno nobili di controllo sociale; le seconde ne fanno uso per migliorare i propri prodotti, aumentare la concorrenza, per scopi di *marketing*, per orientare il consumatore verso offerte sempre più personalizzate. La "contropartita" è però la cessione (in varia misura e vario titolo) dei propri dati (relativi tanto alla sfera personale quanto, più in generale, alle inclina-

[4] Similmente, fra gli altri, J.B. BENIGER, *Le origini della società dell'informazione. La rivoluzione del controllo*, Torino, 1995; V. ZENO-ZENCOVICH, *La libertà d'espressione: media, mercato, potere nella società dell'informazione*, Bologna, 2004; M. PIETRANGELO, *La società dell'informazione tra realtà e norma*, Milano, 2007; L. SARTORI, *La società dell'informazione*, Bologna, 2012 e E. FLORINDI (a cura di), *Computer e diritto: l'informatica giuridica nella società dell'informazione e della conoscenza*, Milano, 2012; G. PITRUZZELLA, O. POLLICINO, S. QUINTARELLI, *Parole e potere*, Milano, 2017.

[5] Sul punto si consiglia R. RUGGERI, *«Internet delle cose» e problematiche giuridiche: alcune considerazioni*, in *Ciberspazio e diritto*, 1-2, 2016, 3-22; E.C. PALLONE, *«Internet of Things» e l'importanza del diritto alla privacy tra opportunità e rischi*, in *Ciberspazio e diritto*, 1-2, 2016, 178 ss.

[6] Così, per esempio, M. BOGNI, A. DEFANT, *Big data: diritti IP e problemi di privacy*, in *Il diritto industriale*, 2, 2015, 117-126; F. DI PORTO, *La rivoluzione big data. Un'introduzione*, in *Concorrenza e mercato*, 1, 2016, 5-14; V. ZENO-ZENCOVICH, G. G. CODIGLIONE, *Ten legal perspectives on the big data revolution*, in *Concorrenza e mercato*, 1, 2016, 29-57.

[7] Cfr., tra gli altri, A. MANTELERO, *Big data: i rischi della concentrazione del potere informativo digitale e gli strumenti di controllo*, in *Il Diritto dell'informazione e dell'informatica*, 1, 2012, 135-144.

[8] Approfondiscono questo delicato profilo, fra gli altri, M. LUCIANI, *La decisione giudiziaria robotica*, in *Rivista AIC*, 3, 2018, 872-893; E. RULLI, *Giustizia predittiva, intelligenza artificiale e modelli probabilistici. Chi ha paura degli algoritmi?*, in *Analisi Giuridica dell'Economia*, 2, 2018, 533-546; A. BONFANTI, *Big data e polizia predittiva: riflessioni in tema di protezione del diritto alla privacy e dei dati personali*, in *Rivista di diritto dei media*, 3, 2018, 206-218; A. M. BERRY-JESTER, *The New Science of Sentencing: Should Prison Sentences Be Based on Crimes That Haven't Been Committed Yet?*, in *The Marshall Project*, 2015, disponibile su https://www.themarshallproject.org/2015/08/04/the-new-science-of-sentencing#.kmIWuDJLS (ultima consultazione 23/04/2019).

zioni, gusti e preferenze), esponendo l'individuo che naviga su internet, utilizza dispositivi *smart* o, semplicemente, guarda Netflix al rischio di subire pregiudizi della propria sfera personale, economica e giuridica senza che tali pericoli siano neppure immaginati dall'utente, principale artefice della lesione alla sua privacy.

I corpi diventano dati e la rappresentazione sociale del soggetto è sempre più affidata ad algoritmi che elaborano le informazioni raccolte e ai profili che su questa base vengono costruiti. L'*alter ego* digitale che deriva dalla scia di informazioni che ciascun individuo lascia ogni giorno dietro di sé non è certo immune da forme di sfruttamento, discriminazione e controllo sociale. Le tecnologie informatiche sono tutt'altro che neutrali, bensì tendono a riprodurre e perpetrare diseguaglianza ed esclusione. Dall'accesso a un prestito alla selezione per un posto di lavoro, i sistemi automatizzati prendono delle decisioni sulla base di ampie raccolte di informazioni personali, spesso senza rivelare esattamente quali esse siano. Alcune informazioni riguardano perfino gli amici, i familiari e i conoscenti, attraverso cui si cerca di capire il carattere di una persona. Le persone che vivono in aree dove il reddito è basso, per esempio, probabilmente hanno amici e familiari con entrate simili alle loro. È più probabile che tra i loro conoscenti ci sia qualcuno che in passato non abbia ripagato un debito. Se un algoritmo tenesse conto di questa variabile, potrebbe escludere la persona solo per colpa dell'ambiente in cui vive[9].

La percezione dei *big data* quale "minaccia" alla libertà degli individui ha assunto contorni più nitidi negli ultimi anni, anche dopo lo scoppio di scandali globali come il caso Snowden[10], *ex* dipendente

[9] Cfr., per esempio, l'interessante articolo di B. SAETTA, *Algoritmi, intelligenza artificiale, profilazione dei dati: cosa rischiamo davvero come cittadini?*, disponibile su https://www.valigiablu.it/algoritmi-dati-rischi/ (ultima consultazione 23/04/2019).

[10] È interessante segnalare la condanna del Regno Unito da parte della Corte europea per i diritti dell'uomo nel caso *Big Brother e altri c. Regno Unito* del 13 settembre 2018 (ricorsi n. 58170/13, 62322/14 e 24960/15). La decisione relativa al c.d. "*Datagate*", nato dalle rivelazioni di Edward Snowden in merito all'esistenza e al funzionamento di programmi di sorveglianza di massa condotti dal governo statunitense e da alcuni Stati europei. I ricorrenti – giornalisti e attivisti dei diritti umani – ritenevano che, in ragione della loro attività, le loro comunicazioni elettroniche fossero state intercettate dai servizi di *intelligence*, trasmesse al governo del Regno Unito dopo essere state captate dalle autorità straniere o, comunque, ottenute dai *Communications Service Providers.* La Corte di Strasburgo, per quanto riguarda la normativa inglese in tema *mass surveillance*, ha ritenuto come, seppur essa individui in modo chiaro e preciso i presupposti per disporre le intercettazioni di massa, la durata delle operazioni e le procedure riguardanti l'accesso, la conservazione e la trasmissione dei relativi dati, debba tuttavia ritenersi incompatibile con l'art. 8 Cedu, non essendo prevista alcuna supervisione da parte di un'autorità indipendente in merito all'utilizzo dei filtri e dei criteri di ricerca delle comunicazioni da intercettare. Inoltre, anche con riferimento all'acquisizione dei *communications data* da parte dei fornitori dei servizi di comunicazione, i giudici – richiamando anche la giurisprudenza della Corte di Giustizia nel caso *Digital Rights Ireland* (8 aprile 2014, cause n. C-293/12 e C-594/12) – hanno affermato l'incompatibilità della legge inglese con l'art. 8 Cedu: per un verso, l'accesso ai dati in possesso delle autorità nazionali non è limitato allo scopo di combattere forme particolarmente gravi di crimine e, per l'altro, non è sottoposto ad una preventiva autorizzazione a opera di un organo amministrativo indipendente. Per quanto volta a garantire un interesse meritevole di tutela, la legislazione nazionale non è dunque in grado di assicurare una protezione efficace contro il rischio di abusi, eccedendo i limiti imposti dal principio di proporzionalità (art. 8, comma 2). Per un approfondimento sulle conseguenze giuridiche del c.d. "*Datagate*" si suggeriscono, fra gli altri, M. NINO, *Il caso Datagate: i problemi di compatibilità del programma di sorveglianza PRISM con la normativa europea sulla protezione dei dati personali e della privacy*, in *Diritti umani e diritto internazionale*, 3, 2013, 727-746; G. RESTA, *La sorveglianza elettronica di massa e il conflitto regolatorio USA/UE*, 23-48; V. ZENO-ZENCOVICH, *Intorno alla*

della CIA, il quale ha reso noti i programmi di sorveglianza di massa utilizzati dalle agenzie di informazione e sicurezza americane successivamente all'attacco terroristico alle torri gemelle. Il Garante europeo per la protezione dei dati personali ha pubblicato nel 2014 un'indagine in materia di *big data*, algoritmi e *profiling*, rilevando una lunga serie di criticità connesse alla tutela dei diritti dei cittadini europei[11] e ha accelerato la riforma, già avviata nel 2012, finalizzata alla predisposizione di strumenti giuridici adeguati alla protezione del consumatore, alla sicurezza dei servizi informatici e alla tutela della privacy, essendo la direttiva 95/46/CE un documento ormai obsoleto e non più in grado di fronteggiare le sfide imposte dall'avanzamento tecnologico globale e dallo sviluppo del mercato digitale. Accanto al tema del consenso al trattamento dei dati personali, nel nuovo Regolamento europeo n. 679 del 2016 – divenuto vincolante per gli Stati a partire dal 25 maggio 2018 – il concetto della "correttezza e trasparenza" dei dati e degli algoritmi che li processano è diventato centrale[12], vale a dire la disponibilità di strumenti in grado di verificarne l'affidabilità, quali per esempio accurati sistemi di *audit* per gli algoritmi e di certificazione per i loro sviluppatori.

La vera novità è proprio il riconoscimento dei diritti legati alla profilazione e alla portabilità dei dati personali, specialmente di fronte alla possibilità che, raccogliendo le "briciole" digitali lasciate dopo la navigazione *online*, l'utilizzo di dispositivi *smart*, il pagamento tramite carte di credito, l'accesso ad *App* quali *Uber* o *Instagram* e via dicendo, il soggetto sia identificato o diventi identificabile con il rischio di determinare situazioni discriminatorie, nonché la violazione dei diritti e della dignità della persona[13]. Come una sorta di sarto invisibile, l'algoritmo è in grado di ricollegare le centinaia di migliaia di tracce lasciate e registrate dai *cookie* dopo la navigazione e di cucirle insieme per creare l'abito virtuale dell'utente, un abito sempre più su misura e personalizzato. I *cookie* agiscono infatti come marcatori elettronici: righe di testo generate dai siti visitati dall'utente e che memorizzano la data, la posizione, la durata dell'ultimo collegamento, le pagine consultate, gli acquisti fatti e più in generale le preferenze di ciascun internauta. La conseguenza? Alla ricerca della parola «ansia» sul motore di ricerca, seguirà l'installazione di centinaia di *cookie* sul dispositivo utilizzato e non dovrà sorprendere se il giorno dopo saranno suggerite pubblicità di rimedi e prodotti proprio contro l'ansia[14].

L'avvento dei *big data* offre certamente grandi possibilità, dal campo medico a quello della sostenibilità energetica, e grandi benefici in termini di miglioramento dei servizi e della qualità della vita;

---

*decisione nel caso Schrems: la sovranità digitale e il governo internazionale delle reti di telecomunicazione*, 7-22; P. PIRODDI, *I trasferimenti di dati personali verso Paesi terzi dopo la sentenza Schrems e nel nuovo regolamento generale sulla protezione dei dati*, 169-214, tutti in G. RESTA, V. ZENO-ZENCOVICH (a cura di), *La protezione transnazionale dei dati personali. Dai "Safe Harbour Principles" al "Privacy Shield"*, Roma, 2016.

[11] Si veda, in particolare, il parere del Garante europeo per la protezione dei dati personali rilasciato nel 2014 e consultabili all'indirizzo: https://edps.europa.eu/sites/edp/files/publication/17-01-13_big_data_ex_summ_it.pdf (ultima consultazione 23/04/2019).

[12] 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), artt. 5 e ss.

[13] Per un approfondimento si veda J. CHESTER, *Cookie Wars: How New Data Profiling and Targeting Techniques Threaten Citizens and Consumers in the "Big Data" Era*, in S. GUTWIRTH, R. LEENES, P. DE HERTYVES POULLET (a cura di), *European Data Protection: In Good Health?*, Berlino, 2012, 53-77.

[14] Cfr. M. AINIS, *Il regno dell'uroboro. Benvenuti nell'era della solitudine di massa*, Milano, 2018, 24.

dall'altra parte, però, fa sorgere anche grandi preoccupazioni, forse non ancora ben comprese. Anzitutto il compito di assicurare la protezione dei diritti individuali si fa più arduo: la privacy, nozione che già di per sé non può dirsi unificante e che ha assunto valenze assai diverse a seconda dei contesti storici e sociali[15], è sotto attacco, tanto da essere considerata da alcuni antiquata se non addirittura dannosa[16]. Da abbandonare, tuttavia, è semmai la versione eremitica di privacy, l'idea antica di uno *ius solitudinis* in cui rifugiarsi e togliersi la propria maschera in tranquillità e sicurezza. Con la società dell'informazione il brandeisiano «*right to be left alone*»[17] ha subito dei profondi cambiamenti, trasformandosi quasi in una ossimorica pretesa di partecipazione, di conoscenza, di controllo su tutte le informazioni che possono identificare una persona. La privacy, in altre parole, oggi assumerebbe un significato ulteriore che si diramerebbe dallo stesso e intimo bisogno di libertà. Da un lato, le esigenze di difesa e di sicurezza nazionale, messe in luce dai recenti episodi di terrorismo, dall'altro, le nuove strategie di mercato hanno trasformato la società in una rete di sorveglianza quotidiana che investe la generalità dei consociati. Nel primo caso, il monitoraggio non più limitato per periodi eccezionali espone l'universalità delle persone e non solo quelle pericolose a una visibilità prima inimmaginabili; nel secondo, la logica mercantilistica del prodotto personalizzato aumenta il rischio che il consumatore possa essere discriminato per le sue opinioni, credenze religiose, condizioni di salute e indirizzato intenzionalmente a determinati contenuti o merci. La privacy come difesa di una sorta di cittadella privata, «*my house, my castle*», non esiste più. Oggi la tutela della privacy rimane fondamentale solo se intesa come «lo strumento necessario per difendere la società delle libertà e per opporsi alle spinte verso la costruzione di una società della sorveglianza, della classificazione, della selezione sociale»[18].

Lungi dal dover essere abbandonata, la privacy deve allora essere presa sul serio: è indubbio che l'innovazione tecnologica abbia fatto espandere insieme agli algoritmi predittivi il senso della tutela di una sfera della persona che si sottrae all'esercizio arbitrario del potere. E allora la riservatezza elasticizzando il suo significato si deve trasformare in «tutela delle scelte esistenziali contro il controllo pubblico e la stigmatizzazione sociale»[19] o «come la richiesta di strumenti sociali che ci mettano al riparo dal rischio d'essere semplificati, oggettivati e giudicati fuori contesto»[20]. È chiaramente una pretesa attiva finalizzata al controllo dei propri dati, la quale può ben tradursi nella richiesta di non esse-

[15] Per una più ampia ricognizione sul tema si suggerisce F. FABRIS, *Il diritto alla* privacy *tra presente, passato e futuro*, in *Tigor - A.I*, 1, 2009, 94-98 e S. NIGER, *Le nuove dimensioni della privacy: dal diritto alla riservatezza alla protezione dei dati personali*, Padova, 2006.

[16] Cfr., in particolare, J. E. COHEN, *What privacy is for*, in *Harv. L. Rev.*, 126, 2012, 1904-1933 e R. SHIH RAY KU, *Privacy is the problem*, in *Widener L.J.*, 19, 873, 2009, 873-891.

[17] S. D. WARREN, L. D. BRANDEIS, *The right to privacy*, in *Harv. L. Rev*, 5, 1890, 193-220.

[18] S. RODOTÀ, *Privacy, libertà, dignità. Discorso conclusivo della Conferenza internazionale sulla protezione dei dati*, 26th International Conference on Privacy and Personal Data Protection, Poland, Wroclaw, 14-16 Settembre 2004, in http://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/export/1049293 (ultima consultazione 23/04/2019).

[19] *Ibidem*.

[20] *Ibidem*. La sua tutela rappresenterebbe non più la privazione da una vita sociale bensì condizione di inclusione nella società della partecipazione, nel senso che solo la protezione dai condizionamenti che possono derivare dall'uso-abuso delle nuove tecnologie renderà libera la partecipazione dell'individuo alla società. Cfr., anche, G. ZICCARDI, *Sorveglianza elettronica, data mining e trattamento indiscriminato delle informazioni dei cittadini tra esigenze di sicurezza e diritti di libertà*, in *Ragion pratica*, 1, 2018, 29-50.

re visto da terzi e incontrarsi così con la matrice solitaria della privacy, ma attraverso strumenti di tutela nuovi che consentano a ciascuno di controllare i propri dati. Del resto, nel suo trasformarsi in uno strumento di difesa contro la società della sorveglianza[21], la privacy si configura sempre di più come un diritto alla *data protection* che, andando oltre la tutela del singolo nella sua solitudine, si estende oltre la sfera della vita privata e in particolare nelle relazioni sociali, così garantendo l'autodeterminazione decisionale e il controllo sulla circolazione dei propri dati[22]. Si tratta, a ben vedere, di garantire la libertà personale come diritto fondamentale, non solo come libertà fisica ma anche contro ogni controllo illegittimo e ogni ingerenza altrui.

Si registra così uno stretto collegamento tra privacy e dignità, nella misura in cui la dignità assume la forma della libertà di autodeterminazione; tale collegamento si erge a fondamentale fattore di contrasto contro forme di controllo di massa, e richiede di rispondere e arginare il fenomeno crescente e convulso del *data mining*[23]. Il fine è quello di proteggere la soggettività dinamica ed emergente degli individui dagli abusi di attori commerciali e governativi i quali hanno interesse a rendere gli individui e le comunità fissi e prevedibili.

## 3. Dietro il velo della neutralità: algoritmi e profilazione

L'uso sempre più diffuso dei sistemi algoritmici per la gestione e l'elaborazione dei giganteschi flussi di dati che ci lasciamo ogni giorno alle spalle pone problemi non soltanto con riguardo alla tutela della privacy, da intendersi oggi come il diritto di ciascuno di mantenere il controllo dei propri dati personali[24], ma anche con riguardo al principio di eguaglianza e di non discriminazione. Il fatto che a essere oggetto di discriminazione possa essere il corpo elettronico (c.d. *body as information*) non rende la violazione meno grave, perché molto spesso ne derivano conseguenze negative anche sulla vita reale dell'individuo, il quale rischia di vedersi rifiutato a un colloquio di lavoro, escluso da un prestito o privato dei suoi diritti.

In primo luogo, l'intelligenza artificiale è sempre più un sistema decisionale che apprende dai dati che ha a disposizione. Amazon, ad esempio, già utilizza l'intelligenza artificiale per determinare le preferenze dei consumatori e suggerire prodotti agli acquirenti, ma anche per ottimizzare l'organizzazione dei magazzini. Le decisioni sono prese sulla base di ampie raccolte di informazioni personali. Tuttavia, se i dati a disposizione, fotografando la realtà sociale esistente, sono di per sé discriminatori è molto

[21] Cfr., in particolare, D. LYON (a cura di), *Surveillance as social sorting: Privacy, risk, and digital discrimination*, New York, 2003.

[22] Cfr., tra gli altri, S. NIGER, *Le nuove dimensioni della privacy: dal diritto alla riservatezza alla protezione dei dati personali*, Padova, 2006; J. KOKOTT, C. SOBOTTA, *The distinction between privacy and data protection in the jurisprudence of the CJEU and the ECtHR*, in *International Data Privacy Law*, 3-4, 2013, 222-228; F. PIZZETTI, *Privacy e il diritto europeo alla protezione dei dati personali: dalla Direttiva 95/46 al nuovo Regolamento europeo*, Torino, 2016; A. PISAPIA, *La tutela per il trattamento e la protezione dei dati personali*, Torino, 2018.

[23] Con tale espressione si intende l'individuazione di nuove informazioni di varia natura tramite estrapolazione mirata da grandi banche dati, singole o multiple (nel secondo caso, informazioni più accurate si ottengono incrociando i dati delle singole banche). Le tecniche e le strategie applicate alle operazioni di *data mining* sono per larga parte automatizzate, consistendo in specifici *software* e algoritmi adatti al singolo scopo. Ad oggi, in particolare, si utilizzano reti neurali, alberi decisionali, *clustering* e analisi delle associazioni.

[24] Cfr. J. E. COHEN, *op. cit*.

probabile che ne deriveranno decisioni altrettanto discriminatorie. Dal momento che l'apprendimento automatico e l'intelligenza artificiale operano attraverso la raccolta, il filtraggio e l'analisi dei dati esistenti, l'algoritmo finirà per replicare i pregiudizi strutturali esistenti a meno che non sia progettato esplicitamente per rendersene conto e contrastarli. La profilazione algoritmica, pertanto, tenderà non soltanto a differenziare persone e gruppi, senza poter avere contezza né del funzionamento alla base degli algoritmi (la c.d. *black box*), né dei criteri di *input* del *software* (ovvero la qualità dei dati), ma anche ad essere staticamente conservatrice nelle sue predizioni, perché fondamentalmente non farà altro che usare dati del passato per prevedere un comportamento futuro[25].

In secondo luogo, anche l'individualizzazione dei calcoli, nei grandi *database*, produce categorizzazioni senza aver l'aria di farlo. Per esempio, negli Stati Uniti, il «*fico score*» misura, per ogni individuo, i rischi che presenta rispetto al credito al consumo[26]. Se è vero che questo programma è pubblico, molti altri registrano e confrontano, nella più grande opacità, i dati concernenti il profilo delle famiglie, l'indebitamento, il consumo, la situazione bancaria o giudiziale. Su tale questione, le legislazioni nazionali sono più o meno tolleranti, ma dappertutto si sviluppa il mercato dei dati tra imprese che rivendono o si scambiano le informazioni. Tuttavia, incrociando informazioni sugli individui, è possibile fare predizioni su certi loro attributi, come la sessualità, la religione o le opinioni politiche, che non sarebbe consentito schedare. Il confronto delle informazioni personali permette poi di indovinare nuove informazioni. Parlare di *big data*, infatti, significa riferirsi non tanto a un grande insieme di dati quanto piuttosto all'operazione di trattamento automatizzato al fine di conoscere nuove informazioni. L'applicazione classica dell'elaborazione algoritmica dei dati è la profilazione degli utenti al fine di inviare loro pubblicità personalizzata (*targeted advertising*).

La vera forza di queste intelligenze artificiali è il processo di scoperta della conoscenza a partire dai flussi di dati, il c.d. *data mining*, per il mezzo del quale è possibile estrapolare correlazioni dai dati, al fine di ricavare profili e modelli predittivi. Il prodotto finale del *data mining* è dunque la classificazione degli utenti, quella che normalmente viene definita come profilazione. Un profilo, infatti, non è altro che una lista di categorie determinate ricavando informazioni dai dati raccolti con riferimento ad un individuo. In base al comportamento di tale individuo, a ciò che fa *online*, ma anche nella vita reale, si definiscono le categorie nelle quali egli può essere inscritto. Il probabile prende così il posto del possibile e le identità digitali si ritrovano così ulteriormente classificate in segmenti sempre più fini senza che il proprietario di quei dati ne sia informato: «cliente non affidabile», «elevate spese mediche», «reddito in declino», «guida pericolosa». Tale classificazione è del tutto indipendente dalla vo-

[25] Sul punto si rinvia a F. PASQUALE, *The Black Box Society: The Secret Algorithms That Control Money and Information*, Londra, 2016.

[26] FICO (Fair Isaac Corporation) è una società di analisi dei dati, con sede a San Jose in California, che eroga servizi di *credit scoring*. Il cosiddetto punteggio FICO o *score FICO*, è una misura del rischio di credito al consumo ed è attualmente diventato uno dei principali punti riferimento nel mondo del credito al consumo negli Stati Uniti. Nel 2013, i creditori – banche per lo più – hanno acquistato più di 10 miliardi di punteggi FICO e circa 30 milioni di consumatori americani hanno avuto accesso ai propri punteggi. Cfr., per esempio, S. LUDWIG, *Credit scores in America perpetuate racial injustice*, 2015, disponibile su https://www.theguardian.com/commentisfree/2015/oct/13/your-credit-score-is-racist-heres-why (ultima consultazione 23/04/2019); L. RICE, D. SWESNIK, *Discriminatory Effects of Credit Scoring on Communities of Color* , in *Suffolk Univ. L. Rev.*, 2013, 935-966; G. ANDREEVA, J. ANSELL, J. CROOK, F. SERV MARK, *Impact of anti-discrimination laws on credit scoring*, in *Journal of Financial Services Marketing*, 9, 1, 2004, 22-33.

lontà dell'individuo e sottratta altresì a qualsiasi forma di controllo collettivo. Una persona può essere classificata come «maschio» o «femmina» (o entrambi in percentuale) in maniera del tutto indipendente da cosa egli è realmente o decide di essere. E, come in ogni forma di classificazione, ci possono essere errori, esclusione e discriminazione, con un disallineamento tra l'identità reale e l'identità digitale. Ad essere sorvegliato, controllato e discriminato non è più la persona in carne e ossa (almeno non direttamente) ma il suo *alter ego* digitale decomposto e ricomposto in un continuo processo di integrazione e disintegrazione dei suoi dati[27]. E, mentre i partecipanti al dibattito sulla privacy e le stesse leggi in materia si focalizzano sulla tutela delle persone, e quindi dei dati identificativi, si tralasciano i rischi del trattamento delle informazioni anonimizzate e aggregate che, invece, appaiono essere estremamente importanti, e foriere di discriminazioni sui gruppi e sulle minoranze, anche perché queste informazioni aggregate, anonimizzate e trattate per gruppi sono esattamente le stesse informazioni utilizzate dalle aziende per stabilire chi è «uomo» e chi è «donna», chi è «gay» e chi è «etero», chi è «cattolico» e chi è «musulmano», chi è «progressista» o «conservatore» e via dicendo.

## 4. Il diritto antidiscriminatorio alla prova dell'intelligenza artificiale: uno sguardo comparato

È noto che nelle società occidentali il dibattito sul principio di eguaglianza si è soprattutto incentrato sulle modalità e sugli strumenti più efficaci a combattere le discriminazioni, tanto attraverso l'utilizzo di clausole di divieto sempre più "sofisticate", quanto attraverso il ricorso a disposizioni normative identificate di volta in volta come azioni positive, misure temporanee speciali o *reservation policy*. Lo dimostra la grande fioritura che, negli ultimi decenni, il diritto antidiscriminatorio ha avuto nella cultura giuridica occidentale, divenendo l'estrinsecazione prevalente e pressoché unica del principio di eguaglianza e configurandosi come una sorta di laboratorio del nuovo linguaggio globalizzato ma anche di positiva regolazione, per cui modelli e strumenti normativi sono migrati da un ordinamento all'altro, da quello nordamericano a quelli europei, dove mediante la tecnica di armonizzazione legislativa sono stati poi adattati alle esigenze del diverso contesto sociale e culturale[28].

La circolazione dei modelli giuridici e degli strumenti di tutela a livello globale è certamente facilitata in campo antidiscriminatorio. I meccanismi di distorsione nella rappresentazione della realtà, che sono il fondamento epistemologico dello stereotipo, del pregiudizio e della discriminazione, possono essere rinvenuti in ogni tipo di raggruppamento umano. Si tratta di fenomeni che vanno di pari passo con l'organizzazione sociale e che, seppur con diversa intensità e con configurazioni peculiari a seconda del circuito culturale di riferimento, pongono problemi comuni. Anche se la discriminazione nel mondo contemporaneo cambia da contesto a contesto, è possibile individuare alcuni ambiti ri-

[27] Così D. LYON, *Surveillance as social sorting, op. cit.*; e, più recentemente, G. ZICCARDI, *Il ricatto digitale. Geopolitica, sorveglianza e controllo*, in *Rivista bimestrale di cultura e di politica*, 4, 2017, 671-678; ID., *Sorveglianza elettronica, data mining e trattamento indiscriminato delle informazioni dei cittadini tra esigenze di sicurezza e diritti di libertà*, in *Ragion pratica*, 1, 2018, 29-50.

[28] Sia consentito il rinvio, per un approfondimento sulla genesi e sullo sviluppo della tutela antidiscriminatoria nell'esperienza americana ed europea, a L. GIACOMELLI, *Ripensare l'eguaglianza. Gli effetti collaterali della tutela antidiscriminatoria*, Torino, 2018.

correnti in cui si manifesta e selezionare alcuni criteri generali in base ai quali gli individui sono sottoposti a trattamenti stigmatizzanti. Dalle restrizioni in ambito lavorativo, che determinano limitazioni della mobilità sociale, rifiuti di opportunità professionali, arbitrarie penalizzazioni sul luogo di lavoro, ostacoli all'acquisizione dell'autostima, alle discriminazioni in ambito sociale e familiare, con forme di segregazione e ghettizzazione, di disconoscimento di alcuni diritti civili e politici, di svilimento istituzionalizzato di certe categorie di individui. Allo stesso tempo, esistono evidenti ricorrenze anche nei criteri ritenuti rilevanti nella discriminazione: il sesso, la razza, la nazionalità, la religione, l'orientamento sessuale, l'età, la disabilità. E nonostante i grandi passi in avanti compiuti per eliminare certe forme di discriminazione, il trattamento svantaggioso riservato alle donne, o alla popolazione nera, o agli omosessuali dalla società non si è dissipato con il passaggio alla modernità ed è tuttora presente nell'attribuzione dei posti occupazionali privilegiati, nella negazione di taluni diritti civili e politici, nell'imposizione di ortodossie liturgiche, di canoni linguistici, di procedure amministrative differenziate e via dicendo. In primo luogo, gli stereotipi e i pregiudizi che danno origine alla discriminazione non sono scomparsi e continuano ad agire non solo in angoli occultati o marginali, ma addirittura nei luoghi istituzionali. In secondo luogo, l'attuale sistema politico e giuridico non appare in grado di garantire un'effettiva eguaglianza (quella sostanziale) e anzi tende a riprodurre, più o meno consapevolmente e mediante dinamiche ritenute legittime, disparità di trattamento, posizioni di privilegio e subordinazione, diseguaglianze.

Se tutto ciò è vero per l'individuo in carne ed ossa perché allora non dovrebbe esserlo anche per il suo corpo elettronico risultante dal riassemblaggio della miriade di dati sparsi nel *web* e rielaborati dalle potenti menti artificiali? Allo stesso modo, quell'identità digitale rischierà di subire le medesime sorti del suo proprietario nella vita reale e, dunque, incorrere nelle medesime discriminazioni derivanti dagli stereotipi e pregiudizi che impregnano la società e la cultura di riferimento. Diversamente da quanto si potrebbe pensare, l'astrazione del mondo virtuale non neutralizza le varie caratteristiche (sesso, genere, razza, etnia, età, orientamento sessuale etc.) che, socialmente e culturalmente, connotano gli individui, differenziandoli. In realtà, proprio l'assenza del *background* sociale e culturale renderà ancora più marcate quelle differenziazioni e le trasformerà in strumenti di discriminazione ancora più pericolosi proprio perché decontestualizzati.

Di fronte all'avanzata dello spazio digitale e alla fede sempre più assoluta riposta sull'elaborazione algoritmica, la tutela antidiscriminatoria rivela i propri limiti, sia perché pensata per persone fisiche, sia perché anch'essa fondata su una logica classificatoria, individualistica e non intersezionale.

### 4.1 I paradossi del *disparate impact* nell'esperienza americana

Come è noto, il cuore del diritto antidiscriminatorio americano è tutt'oggi rappresentato dal *Civil Rights Act* del 1964, definito da alcuni[29] come la più importante legge sui diritti civili del Secolo scor-

[29] Cfr. C. McCrudden, *Anti-Discrimination Law*, New York, 2004; K. Yoshino, *The New Equal Protection*, in *Harv. L. Rev.*, 124, 2011, 747-803; D. Oppenheimer, S. Foster, S. Han, R. Ford, *Comparative Equality and Anti-Discrimination Law (2nd edition)*, Berkeley, 2017.

so, e a cui si sono aggiunti l'*Age Discrimination in Employment Act* del 1967, il *Fair Housing Act* del 1968 (che costituisce il Titolo VIII del *Civil Rights Act*) e l'*Americans with Disabilities Act* del 1990. La vera innovazione di questa legislazione è stato il richiamo esplicito a una duplice base costituzionale, la *Commerce Clause* e la sezione quinta del XIV Emendamento: infatti, ciò ha reso possibile, per la prima volta nella storia americana, l'applicazione del divieto di discriminazione anche ai rapporti fra privati. È la Corte Suprema stessa nella sentenza *Heart of Atlanta Motel v. United States* [30] a confermare la legittimazione del Congresso a interferire nell'autonomia privata in forza della *Commerce Clause*.

Anche il diritto antidiscriminatorio americano conosce due diverse nozioni di discriminazione che comportano una lesione del principio di parità di trattamento: il *disparate treatment*, ovvero la discriminazione diretta che riguarda la situazione in cui un soggetto viene intenzionalmente trattato in maniera diversa da un altro soltanto in ragione del possesso di certi fattori "protetti" quali la razza, il sesso, la nazionalità, la religione, l'età, la disabilità, l'orientamento sessuale e via dicendo; e il *disparate impact*, ovvero la discriminazione indiretta che ricorre ogni volta che una regola o una prassi, che apparentemente non distingue in base ad alcun fattore "protetto", produce comunque un effetto ingiustificatamente discriminatorio su taluni gruppi di individui rispetto ad altri. Si deve alla giurisprudenza la teorizzazione della discriminazione indiretta che, sebbene non prevista esplicitamente dalla normativa, ha attenuato l'impostazione fortemente soggettiva, legata alla dimostrazione dell'intento discriminatorio dell'agente, il perno che definisce il modello americano di tutela antidiscriminatoria. Infatti, è atteggiamento costante della giurisprudenza statunitense quello di richiedere in ogni caso la prova dell'esistenza dell'elemento intenzionale da parte del soggetto che discrimina, ammettendo soltanto in via subordinata che l'attore possa ricorrere a elementi presuntivi per assolvere al suo onere probatorio. Si tratta di un modello procedurale completamento opposto a quello europeo, dove, in forza delle direttive di seconda generazione fondate sull'articolo 13 del Trattato di Amsterdam, l'allegazione da parte del ricorrente di elementi atti a far presumere l'avvenuta discriminazione valgono a spostare l'onere della prova in capo al convenuto (che dovrà, dunque, dimostrare di aver agito legittimamente[31]). Nel caso statunitense, invece, l'onere di provare il comportamento intenzionalmente discriminatorio del datore di lavoro rimane unicamente in capo al ricorrente.

Le resistenze all'accoglimento della nozione di discriminazione indiretta sono, quindi, riconducibili in gran parte a tale impostazione procedurale. Ad ogni modo, a partire dal famoso caso *Griggs v. Duke Power*[32] del 1971, anche il sistema americano perviene all'elaborazione della nozione di *disparate impact*: «The objective of the Congress in the enactment of Title VII is […] to achieve equality of employment opportunities and remove barriers that have operated in the past in favor an identifiable group of white employees over other employees. Under the act, practices, procedures or test neutral on their face, and even neutral in terms of intent, cannot be maintained if they operate to 'freeze'

[30] *Heart of Atlanta Motel, Inc. v. United States*, 379 U.S. 241 (1964).

[31] Per una panoramica generale sul diritto antidiscriminatorio europeo si consigliano, fra i molti, M. BELL, *Anti-Discrimination Law and the European Union*, Oxford, 2002, S. FREDMAN, *Discrimination Law*, Oxford, 2011 e E. ELLIS, P. WATSON, EU *Anti-Discrimination Law*, Oxford, 2012.

[32] *Griggs v. Duke Power Co*., 401 U.S. 424 (1971).

the status quo of prior discriminatory practices. […] The act proscribes not only overt discrimination but also practices that are fair in form, but discriminatory in operation»[33].

Dopo una prima fase di apertura, tuttavia, la Corte suprema fa un passo indietro e, confermando l'originaria impostazione individualistica dell'*antidiscrimination principle*, limita la portata della discriminazione indiretta, affermando che occorre fornire la prova dell'intento discriminatorio del provvedimento, e che questo onere non può essere adempiuto limitandosi a documentare il solo impatto discriminatorio di quel provvedimento nei confronti di un certo gruppo. Occorre dimostrare, anche e soprattutto, che quel provvedimento è stato adottato con l'intenzione di discriminare quella certa categoria di persone[34]. Emblematiche, in questo senso, sono le sentenze sulla pena di morte e sui reati di droga, nelle quali la Corte Suprema ha sempre rifiutato di far valere come sintomo del carattere discriminatorio delle relative previsioni il fatto che i condannati a morte fossero in numero disproporzionato persone nere[35]. Lo stesso criterio hanno seguito le corti di merito per escludere rilievo, ai fini discriminatori, al fatto che il tipo di droghe il cui consumo è punito più severamente riguarda proprio quelle sostanze generalmente più utilizzate dai neri[36]. Il legislatore, per riequilibrare la distribuzione dell'onere della prova e attenuare, in parte, il rigore delle corti nel riconoscimento dell'impatto discriminatorio, è intervenuto a codificare la nozione di discriminazione indiretta attraverso il *Civil Rights Act* del 1991, ma nemmeno questo è bastato a fare chiarezza, tanto che l'orientamento della giurisprudenza è tuttora altalenante.

Nell'era dei *big data*, l'attuale modello di tutela antidiscriminatoria non appare adeguatamente equipaggiato per tutelare gli individui contro gli esiti discriminatori delle decisioni derivanti dall'utilizzo dei sistemi di intelligenza artificiale. Paradossalmente vengono meno proprio i due pilastri principali sui quali quel modello si regge: da un lato, l'uso delle categorie "protette" (genere, razza, età) quale *input* nell'elaborazione algoritmica è generalmente irrilevante per un potenziale risultato discriminatorio poiché la miriade di variabili aggregate e l'oscurità del processo computazionale rendono di fatto impossibile stabilire la correlazione tra il fattore "protetto" e il risultato discriminatorio; dall'altro, è una sfida quasi impossibile ricondurre alla volontà di un "soggetto" la *data discrimination*, derivando quest'ultima da un calcolo matematico/statistico.

Nel caso di una disparità di trattamento derivante dall'uso di un algoritmo, difficilmente potrà essere riconosciuta la discriminazione diretta proprio perché non sarà possibile provare in giudizio che la decisione algoritmica è stata presa con l'intenzione di discriminare proprio quella determinata categoria di persone. L'unica analogia possibile per l'applicazione del *disparate treatment* si potrebbe avere di fronte ai meccanismi di *data mining* che riflettono il pregiudizio di una persona in carne e ossa; si pensi, per esempio, ai motori di raccomandazione come il *Talent Match* di LinkedIn, che si basa su valutazioni umane potenzialmente pregiudizievoli dei lavoratori[37]. Come regola generale, un dato-

[33] *Idem*, 430.

[34] Si veda sul punto il clamoroso *revirement* della Corte Suprema nel caso *Ricci v. DeStefano*, 557 U.S. 557 (2009).

[35] Cfr. *McCleskey v. Kemp*, 481 U.S. 279 (1987).

[36] Cfr. *United States v. Clary*, 846 F. Su 768, 1994.

[37] Cfr. D. WOODS, *LinkedIn's Monica Rogati on "What is a Data Scientist?"*, in *Forbes*, 2011, disponibile su https://www.forbes.com/sites/danwoods/2011/11/27/linkedins-monica-rogati-on-what-is-a-data-scientist/#50a228673e15 (ultima consultazione 23/04/2019).

re di lavoro non può eludere la responsabilità da discriminazione diretta giustificando il suo comportamento dietro le preferenze espresse da altri soggetti. Però, anche in questa ipotesi, appare alquanto difficile riuscire a dimostrare che il datore di lavoro abbia voluto trattare diversamente il lavoratore sulla base di uno dei fattori "protetti", sapendo dell'effetto discriminatorio di quel meccanismo.
Dove non vi è *discriminatory intent*, potrà allora trovare applicazione la dottrina della discriminazione indiretta. In un giudizio per *disparate impact*, infatti, il ricorrente è tenuto a provare che una decisione o una prassi apparentemente neutrale ha prodotto un risultato ingiustificatamente svantaggioso per un certo gruppo "protetto". Di fronte alla «*business necessity*» opposta dal convenuto a giustificazione del proprio comportamento, il ricorrente potrà controbattere dimostrando l'esistenza di una «*alternative employment practice*» che avrebbe determinato un effetto meno svantaggioso per il gruppo in questione. Senza addentrarsi nella tortuosa giurisprudenza in materia di discriminazione indiretta, che tuttora fatica a delinearne l'effettiva portata, di fronte agli effetti discriminatori derivanti da decisioni algoritmiche essa può rappresentare l'unico rimedio giuridico disponibile poiché colpisce la discriminazione non intenzionale[38].
Particolarmente interessante per le sue possibili ripercussioni è la decisione della Corte Suprema nel caso *Texas Department of Housing & Community Affairs v. The Inclusive Communities Project Inc.*[39] del 2015, in cui è stato riconosciuto il *disparate impact* di un programma governativo di assegnazione degli alloggi che, fondandosi su un sistema algoritmico, discriminava sulla base della razza e dell'etnia, segregando di fatto le persone nere nei quartieri più poveri e periferici della città. Il progressivo spostamento dei cittadini a basso reddito, per la maggior parte neri e ispanici, dai quartieri residenziali verso le aree più periferiche delle città è qualcosa di più di un cambiamento demografico: è una segregazione di ritorno, un problema di diseguaglianza economica e di discriminazione razziale. L'aumento degli affitti e delle tasse di proprietà, sfratti e pignoramenti, aumento dei prezzi, l'assenza di agevolazioni per l'edilizia nei quartieri urbani residenziali costituiscono delle barriere quasi insormontabili che spingono le famiglie di colore a segregarsi ai margini delle regioni metropolitane. Questo è ciò che accade anche a Dallas dove il 92,29% degli alloggi costruiti usufruendo delle agevolazioni per l'edilizia popolare si trova in aree urbane abitate quasi esclusivamente da persone di colore. Alla luce di questi dati, l'*Inclusive Communities Project*, organizzazione *no-profit* con sede in Texas, decide di ricorrere contro l'agenzia governativa che si occupa di concedere crediti d'imposta per la costruzione di case popolari. L'accusa era che il programma governativo – che utilizza un *software* algoritmico per stabilire i beneficiari e distribuire le agevolazioni fiscali in «*an objective, transparent, predictable and race-neutral manner*» – assegnando troppe poche agevolazioni all'edilizia popolare nei quartieri residenziali della città, ad alta concentrazione di bianchi, determinava una segregazione delle fasce più povere, per la maggior parte appartenenti a minoranze razziali ed etniche, nelle zone più periferiche e marginali.
La Corte Suprema ha riconosciuto per la prima volta la discriminazione indiretta nell'accesso alla casa, ritenendo sufficiente la prova statistica dell'effetto discriminatorio del programma governativo sulle persone di colore. Dunque, al pari del *Civil Rights Act* e dell'*Age discrimination Act*, anche nel

[38] Su punto cfr. J. KOBICK, *Discriminatory Intent Reconsidered: Folk Concepts of Intentionality and Equal Protection Jurisprudence*, in *Harv. L. Rev.*, 45, 2010, 517-551.
[39] *Texas Dept. of Housing and Community Affairs v. Inclusive Communities Project, Inc.*, 576 U.S. ___ (2015).


ISSN 2284-4503

*Fair Housing Act* del 1968 è possibile riscontrare un divieto di discriminazione indiretta quando alla *Section* 804(a) si vieta «to refuse to sell or rent after the making of a bona fide offer, or to refuse to negotiate for the sale or rental of, or otherwise make unavailable or deny, a dwelling to any person because of race, color, religion, sex, familial status, or national origin». Proprio l'espressione «otherwise make unavailable» viene interpretata dai giudici come riferita alle conseguenze di una azione piuttosto che all'intenzione del soggetto che agisce. Questo consente di discostarsi dalla rigorosa prova della volontà di discriminare – che nel caso di un algoritmo diventa quasi impossibile – e guardare agli effetti prodotti concretamente da quella pratica o decisione automatizzata. Non riconoscere questo, sostiene il giudice Kennedy nell'opinione di maggioranza, significherebbe depotenziare irrimediabilmente una legge che ha l'obiettivo di eradicare la discriminazione in un settore nevralgico della società e dell'economia. La discriminazione indiretta, infatti, contribuisce alla sostanzializzazione del principio di eguaglianza, consentendo un approccio sistemico in grado di smascherare politiche e pratiche che perpetuano segregazione ed esclusione. È il caso delle c.d. *zoning laws*[40] e delle restrizioni all'accesso alla casa che molto spesso finiscono per marginalizzare ingiustamente certe minoranze in determinati distretti elettorali o in quartieri periferici e degradati.

La Corte, pur accogliendo il ricorso, non allarga però le maglie della tutela antidiscriminatoria, imponendo significative limitazioni all'applicazione della *disparate impact doctrine*: di fronte alle difese del convenuto, che può opporre una ampia varietà di ragioni legittime per giustificare la propria politica – elevato costo degli alloggi, traffico, sovraffollamento, vincoli storici –, il ricorrente non può limitarsi ad allegare meri dati statistici per provare la disparità di trattamento, ma dovrà piuttosto dimostrare il nesso causale tra la politica e l'impatto discriminatorio e l'esistenza di un'alternativa meno svantaggiosa per il gruppo "protetto". Inoltre, la Corte aggiunge che, anche laddove la responsabilità per discriminazione indiretta sia accertata dal giudice, i rimedi concessi dovranno limitarsi all'eliminazione della politica o della pratica discriminatoria ed essere rigorosamente *race-gender-neutral* onde evitare di incorrere in una discriminazione al rovescio.

Quest'ultime precisazioni – o salvaguardie – manifestano i limiti e i paradossi dell'attuale modello antidiscriminatorio di fronte alle sfide dei *big data* e delle intelligenze artificiali. La dottrina del *disparate impact*, proprio perché *result-oriented*, si presta molto bene ad agire contro gli effetti discriminatori delle decisioni algoritmiche, ma solo in linea teorica. Come mostra il caso *Texas Department of Housing & Community Affairs v. The Inclusive Communities Project Inc.*, l'applicazione del *disparate impact* incontra molti ostacoli tecnici e giuridici: in primo luogo, l'impossibilità di controbattere a una decisione – quella del *software* algoritmico – che si fonda su dati statistici con altrettanti dati statistici che comprovino l'ingiustizia dei suoi risultati; in secondo luogo, la difficoltà di provare il nesso causale tra la pratica discriminatoria e lo svantaggio per un determinato gruppo, essendoci un moltepli-

[40] La c.d. zonizzazione è uno strumento utilizzato in urbanistica consistente nel suddividere il territorio di ciascun comune in aree omogenee secondo determinate caratteristiche. L'attività di zonizzazione è quella mediante la quale la pubblica amministrazione suddivide il proprio territorio comunale in zone alle quali viene riconosciuta o attribuita una determinata funzione con conseguente attribuzione di vincoli ed altri limiti da osservare per ciascuna zona. La prima *zoning law* negli Stati Uniti fu adottata nel 1916 dalla città di New York. Per un'introduzione sul tema della discriminazione derivante dalla zonizzazione si suggerisce C. BERRY, *Land Use Regulation and Residential Segregation: Does Zoning Matter?*, in *American Law and Economics Review*, 2, 2001, 251-274.

cità di dati che concorrono alla decisione ed essendo sempre più complessa la logica della scatola nera che conduce a quel risultato; la difficoltà, infine, di provare l'esistenza di soluzioni alternative che perseguano il medesimo fine ma con un minor svantaggio per il gruppo "protetto" allorché lo sviluppo tecnologico dichiara di offrire la soluzione matematicamente migliore, più efficiente e oggettiva.

### 4.2 La "recidività" del pregiudizio: la lezione canadese

L'impiego di algoritmi predittivi in campo giudiziario è ormai pratica diffusa. Ma, mentre in Europa sistemi di questo tipo sono ancora in corso di sperimentazione e si discute della loro compatibilità con i diritti del giusto processo[41], negli Stati Uniti e in Canada non solo esistono, ma vengono concretamente applicati: dalle *online dispute resolution* (ODR) alle cause commerciali, dagli arbitrati al diritto penale[42].

I rischi legati all'utilizzo di tecniche di profilazione sono evidenti. Secondo il Parlamento europeo, innanzitutto, esiste «the risk of data being used for discriminatory or fraudulent purposes and the marginalisation of the role of humans in these processes, leading to flawed decision making procedures that have a detrimental impact on the lives and opportunities of citizens, in particular marginalised groups, as well as bringing about a negative impact on societies and businesses»[43]. Ciò ha l'effetto di rendere l'utilizzo dei *big data* capace di causare non solo violazioni dei diritti individuali fondamentali (la presunzione di innocenza, il diritto di essere informato, il diritto di difesa, il diritto al contradditorio etc.) ma anche trattamenti discriminatori o discriminazioni indirette nei confronti di gruppi di persone con caratteristiche simili. In secondo luogo, vi è il pericolo che i giudici utilizzino questi strumenti in maniera inappropriata, cadendo nella tentazione di delegare loro la decisione finale, senza che questo sia l'obiettivo per cui sono stati programmati e confidando in una loro neutralità che, come visto, potrebbe non esserci affatto.

Proprio per questo è particolarmente interessante il caso *Ewert v. Canada*[44], deciso dalla Corte Suprema canadese nel giugno 2018, che, sebbene non direttamente connesso alla legittimità costituzionale del *decision-making* algoritmico, tocca uno dei nodi essenziali della questione, ovverosia la qualità e la trasparenza delle informazioni che ne sono alla base. Jeffrey Ewert, condannato a due ergastoli per omicidio e tentato omicidio per strangolamento e aggressione sessuale, si trova da oltre trent'anni sotto la custodia federale di media e massima sicurezza. In più di un'occasione chiede di essere sottoposto a una valutazione psicologica per la riduzione del livello di sicurezza e per essere autorizzato a fare domanda per la libertà condizionale. Tuttavia, il *Correctional Service of Canada*

[41] Sul punto si consiglia la lettura del volume A. GARAPON, J. LASSÈGUE, *Justice digitale. Révolution graphique et rupture anthropologique*, Parigi, 2018. Si veda anche R. SICURELLA, V. SCALIA, *Data Mining and Profiling in the Area of Freedom, Security and Justice: State of Play and New Challenges in the Balance between Security and Fundamental Rights Protection*, in *New Journal of European Criminal Law*, 4, 2013, 409-460.

[42] Per una panoramica sull'applicazione negli Stati Uniti, cfr. R. SIMMONS, *Quantifying Criminal Procedure: How to Unlock the Potential of Big Data in Our Criminal Justice System*, in *Mich. St. L. Rev.*, 947, 2016, 969-975; D.B. MARLOWE, *Adaptive Programming Improves Outcomes in Drug Court: An Experimental Trial*, in *Crim. Justice Behav.*, 39, 2012, 514–532.

[43] Risoluzione del Parlamento europeo 14 marzo 2017 sulle «*implicazioni dei Big Data per i diritti fondamentali: privacy, protezione dei dati, non discriminazione, sicurezza e attività di contrasto*» (2016/2225(INI)), disponibile su http://www.europarl.europa.eu/ (ultima consultazione 23/04/2019).

[44] *Ewert v. Canada*, 2018 SCC 30.

(CSC), l'agenzia federale che si occupa di svolgere la valutazione dei detenuti e redigere la relazione per il giudice, puntualmente respinge la richiesta del ricorrente. Già nel 2005 il signor Ewert chiedeva che fosse effettuata una verifica sull'accuratezza dei test psicologici quando si trattava di valutare soggetti di etnia *Metis* perché a suo avviso i risultati erano discriminatori e violavano le garanzie del *due process*. Il Servizio ammetteva la possibilità di questo rischio, essendo il sistema settato su un campione di dati non rappresentativo, e annunciava che avrebbe affidato a un ente esterno il compito di accertarlo. Per questa ragione in primo e secondo grado veniva rigettata la domanda del ricorrente. Nel 2018 la Corte Suprema canadese, però, dà ragione al signor Ewert affermando che le autorità hanno l'obbligo di «to take all reasonable steps to ensure that any information about an offender that it uses is as accurate, up to date and complete as possible»[45].

Nello specifico, la Corte ritiene che il concetto di "*information*", che a norma dell'art. 24 del *Correction and Conditional Release Act* (CCRA) deve per l'appunto essere «as accurate, up to date and complete as possible», includa anche i profili elaborati e le previsioni fatte su un determinato individuo tramite l'utilizzo delle tecnologie informatiche. Diversamente da quanto sostenuto dal Governo, secondo cui l'accuratezza non era riferibile ai dati predittivi generati dai test, i quali possono avere diversi livelli di validità predittiva, nel senso che predicono male, moderatamente o enormemente bene, i giudici di maggioranza sono convinti che il concetto di accuratezza possa essere adattato anche ai risultati derivanti dall'uso di algoritmi e pertanto sia compito dell'autorità verificare la loro completezza, precisione e oggettività. Se gli strumenti di intelligenza artificiale incorporano pregiudizi, o non sono appropriatamente sensibili alle differenze culturali, allora il risultato è "oggettivo" solo in un senso molto ristretto della parola. Se questi strumenti, fa notare la Corte, sono considerati utili proprio in ragione del fatto che le informazioni che ne derivano possono essere convalidate scientificamente, questa è una ragione in più per concludere che la legge imponga al CSC l'obbligo di adottare tutte le misure ragionevoli per garantire che le informazioni siano accurate.

In effetti, nonostante il CSC fosse da tempo a conoscenza delle preoccupazioni riguardo alla possibilità che tali test psicologici perpetuassero certi pregiudizi etnico-culturali, non era stata intrapresa alcuna iniziativa concreta per valutare l'affidabilità dei test che hanno continuato ad essere applicati indistintamente a tutti i detenuti, compresi quelli di origine indigena. Secondo l'opinione di maggioranza, questo rappresenta una violazione dei principi generali indicati nell'art. 4 del CCRA, sulla base dei quali le politiche e le pratiche dell'autorità federale devono rispettare le differenze culturali, linguistiche e di altro tipo e tener conto dei bisogni speciali delle donne, delle popolazioni aborigene, delle persone che richiedono cure mentali e di altri gruppi. Un principio, questo, che è volto a riconoscere «la discriminazione sistemica affrontata dalle persone indigene nel sistema giudiziario e correzionale canadese e non solo»[46]. Da ciò discende, secondo i giudici, l'onere della prova a carico del CSC, il quale deve dimostrare di aver adottato tutte le misure adeguate a garantire l'affidabilità degli strumenti di valutazione. La Corte non specifica chi è destinato a svolgere questa verifica, quali standard dovrebbero essere soddisfatti, o quanto dovrebbe essere estesa tale valutazione. Questi sono problemi importanti, specialmente quando si ha a che fare con discriminazione algoritmiche che implicano una valutazione indipendente da parte di soggetti terzi.

[45] *Idem*, par. 35.
[46] *Idem*, par. 54.

Alla sentenza in commento è mancato inoltre di fare il passo successivo, quello più importante: testare la tenuta dei principi costituzionali del *due process* e dell'eguaglianza di fronte all'uso – sempre più determinante – dell'intelligenza artificiale nei processi decisionali. La Corte, infatti, liquida questi due argomenti, pure sollevati dal ricorrente, per mancanza di prove sufficienti, non essendo stato possibile dimostrare né l'arbitrarietà dell'affidamento sui test, né l'effettivo impatto discriminatorio sui detenuti indigeni rispetto ai detenuti non indigeni. Con riferimento al primo argomento, il ricorrente aveva evidenziato che il mancato previo accertamento dell'attendibilità dei test costituisse una violazione dei principi fondamentali di giustizia in base all'art. 7 della *Canadian Charter* che stabilisce che «everyone has the right to life, liberty and security of the person and the right not to be deprived thereof except in accordance with the principles of fundamental justice». In proposito, la Corte richiama la propria giurisprudenza ribadendo che «the principles of fundamental justice are to be found in the basic tenets of our legal system. They do not lie in the realm of public policy but in the inheret domain of the judiciary as guardian of the justice system»[47]. Con tale clausola, infatti, si impone all'interprete di trovare un punto d'equilibrio tra l'interesse dello Stato e quello della persona. Nel caso di specie, i giudici non ritengono adeguatamente dimostrata la natura di principio fondamentale di giustizia posto che la constatazione che vi sia incertezza sull'accuratezza dei test quando applicati ai trasgressori indigeni non è sufficiente di per sé a stabilire l'assenza di una connessione razionale tra l'affidamento a detti test e l'interesse legittimo del governo[48]. Invece, con riferimento all'argomento egualitario, il ricorrente aveva sostenuto che il CSC stesse usando "informazioni affidabili o veritiere" per prendere decisioni sui detenuti non indigeni e "informazioni inaffidabili o false" per prendere decisioni sui detenuti indigeni. Questa pratica aveva pertanto determinato trattamenti più severi e detenzioni più lunghe nei confronti dei soli detenuti indigeni. La Corte, pur riconoscendo il difetto di accuratezza dei test utilizzati dal CSC, esclude comunque la violazione dell'*equal protection* di cui all'art. 15 della *Canadian Charter*[49], affermando che «the evidence before the trial judge established a risk that the impugned tools are less accurate when applied to Indigenous inmates than when they are applied to non-Indigenous inmates. However, the trial judge did not find, and indeed could not have done so on the evidence before him, that the impugned tools do in fact overestimate the risk posed by Indigenous inmates or lead to harsher conditions of incarceration or to the denial of rehabilitative opportunities because of such an overestimation»[50].

A tal proposito, lo sguardo comparato ci invita al confronto con un caso analogo deciso però dalla Corte Suprema del Wisconsin (Stati Uniti) nel 2016. Anche nel sistema giudiziario americano si utilizzano dei *software* che assistono i giudici nel calcolare la percentuale di rischio che il soggetto che è sotto giudizio possa ricommettere in futuro altri crimini. Questi programmi informatici – chiamati di *risk assessments* – sono ampiamente utilizzati in tutto il sistema penale. Gli algoritmi di calcolo del rischio di recidiva vengono usati per assistere nella decisione di chi può rimanere in libertà in ogni

[47] *Re B.C. Motor Vehicle Act*, 2 S.C.R. 486, 1985.
[48] *Ewert v. Canada*, par. 76.
[49] «*Every individual is equal before and under the law and has the right to the equal protection and equal benefit of the law without discrimination and, in particular, without discrimination based on race, national or ethnic origin, colour, religion, sex, age or mental or physical disability*», Art. 15, comma primo, Canadian Charter of Rights and Freedoms.
[50] *Ewert v. Canada*, par. 79.

momento del procedimento giudiziario: dalla determinazione della cauzione fino alla decisione che incide definitivamente sulla libertà dell'imputato. In particolare poi nelle corti dell'Arizona, del Colorado, del Delaware, del Kentucky, della Louisiana, dell'Oklahoma, della Virginia, di Washington e del Wisconsin, i risultati di questi calcoli di rischio sono consegnati al giudice come elemento determinante per la sentenza penale.

È il caso di Eric Loomis che veniva arrestato nel 2013 per ricettazione e resistenza a pubblico ufficiale e successivamente condannato alla pena di sei anni di reclusione, una pena particolarmente severa determinata dalla Corte distrettuale sulla base dell'alto punteggio (*score*) risultante a carico dell'imputato da COMPAS – Correctional Offender Management Profiling for Alternative Sanctions[51] –, uno degli algoritmi predittivi di valutazione del rischio di recidiva più diffusi. Loomis impugnava la sentenza sostenendo che l'utilizzo da parte del giudice di un algoritmo predittivo per addivenire alla pronuncia di condanna e alla comminazione della pena avesse violato le garanzie del *due process*, in quanto COMPAS è un algoritmo proprietario, il cui meccanismo di funzionamento – che si basa sulla raccolta e sull'elaborazione dei dati emersi dal fascicolo processuale e dall'esito di un test a 137 domande, a cui viene sottoposto l'imputato, riguardanti età, attività lavorativa, vita sociale, grado di istruzione, legami, uso di droga, opinioni personali e percorso criminale – non è pubblicamente noto e dunque la sua validità e oggettività non è accertabile.

Nel luglio 2016 la Corte Suprema statale, pronunciandosi sul caso *State v. Loomis*[52], ha tuttavia dichiarato, all'unanimità, la legittimità dell'uso giudiziario di algoritmi che misurano il rischio di recidiva specificando che lo strumento non può essere l'unico elemento su cui si fonda una pronuncia di condanna. Il ricorrente aveva sostenuto la violazione della *due process clause* del XIV Emendamento, ritenendo che l'algoritmo, anche se viene considerato "imparziale" perché agisce solo sui dati provenienti dal questionario, lo aveva discriminato, rivelandosi altamente punitivo soprattutto per il genere maschile e per le persone di razza non bianca. Inoltre, anche il fatto della società Northpointe di

[51] *Compas* (*Correctional offender managment profiling for alternative sanctions*) è tra gli algoritmi processuali più applicati in via sperimentale negli Stati Uniti. Brevettato da una società americana e inizialmente adottato da un ristretto numero di Corti di merito americane per la previsione su base statistica della *probability of the offender's recidivism* ai fini della quantificazione della pena e della successiva *extended supervision*, oggi il suo utilizzo è stato legittimato ufficialmente in diversi Stati federati. Attraverso l'inserimento di una serie di dati oggettivi concernenti il trascorso criminale, le condizioni socioeconomiche e personali dell'imputato, nonché 137 domande a risposta vincolata poste allo stesso, *Compas* misura il rischio di recidiva come *low*, *medium* e *high*. Si tratta di un *software* già largamente in uso ad alcuni corpi di polizia nazionali per calibrare l'intensità della supervisione dei soggetti in libertà vigilata. È bene precisare che, trattandosi di algoritmo brevettato, «la metodologia usata per formulare tale valutazione non è stata comunicata né al giudice né all'imputato». Pertanto, non è possibile vagliare in concreto l'idoneità del meccanismo di calcolo approntato, ma solo la sua astratta idoneità a fornire informazioni rilevanti per il giudizio sul pericolo di recidiva. Così coniugati dati oggettivi relativi alla storia criminale dell'imputato e risposte fornite dallo stesso imputato, *Compas* processa tutte le informazioni raccolte suddividendole in 12 sezioni: «*Current Charges*, *Criminal History*, *Non-Compliance*, *Family Criminality*, *Peers*, *Substance Abuse*, *Residence/Stability*, *Social Environment*, *Education*, *Vocation*, *Leisure/Recreation*, *Social Isolation*, *Criminal Personality*, *Anger* e *Criminal Attitudes*. Informazioni disponibili sul sito della società Northpointe Inc. proprietaria dell'algoritmo: (www.northpointeinc.com/files/downloads/FAQ_Document.pdf). Per un approfondimento si veda T. BRENNAN et Al., *Evaluating the predictive validity of the COMPAS risk and needs assessment system*, in *Crim. Just. Behav.*, 21, 2009, 21-40.

[52] *State v. Loomis*, 881 N.W.2d 749 (Wis. 2016).

non rivelare i meccanismi attraverso i quali COMPAS giunge alle sue "classifiche" di rischio da sottoporre ai giudici rappresenterebbe una violazione dei basilari diritti della difesa perché tale metodo, che si basa sulle informazioni raccolte, potrebbe venir alterato di volta in volta, modificando il sistema di calcolo dei rischi e delle pene.

La Corte, tuttavia, ha respinto tutti gli argomenti del ricorrente affermando che l'utilizzo di fattori quali il genere e la razza per la valutazione del rischio ha, all'opposto, uno scopo non discriminatorio, finalizzato ad assicurare la completezza e l'accuratezza delle valutazioni e che in mancanza di prove evidenti che la decisione sia stata presa soltanto sulla base di quei fattori si debba escluderne la natura discriminatoria. La Corte ha anche aggiunto, con riferimento al rischio di de-individualizzazione della decisione, che se è vero che COMPAS procede per aggregazioni di dati sul rischio di recidiva che inevitabilmente collegano il soggetto al gruppo etnico-razziale di appartenenza, è altresì vero che questo non è il solo elemento su cui la decisione si basa e che in ogni caso il giudice conserva la propria discrezionalità.

La Corte sembra essere conscia dei pericoli che si nascondono dietro la fiducia nella scientificità di questi sistemi di calcolo: i punteggi sul rischio e sulla recidività non possono essere usati «to determine whether an offender is incarcerated» o «to determine the severity of the sentence». Il giudice, infatti, nell'utilizzare i punteggi sul rischio deve sempre motivare la sua sentenza, dando conto di tutti gli altri fattori considerati[53]. L'obbligo di motivazione da parte del giudice e il divieto di fondare la sua decisione soltanto sui punteggi provenienti dal sistema informatico sono sufficienti, per la Corte del Wisconsin (ma anche per quella dell'Indiana[54]), a salvaguardare i diritti fondamentali dell'imputato[55].

La differenza tra il caso *Loomis* e l'applicazione ormai consolidata degli algoritmi predittivi del rischio di recidiva sta nel fatto che, per la prima volta, il programma COMPAS è stato utilizzato in fase di cognizione quale elemento determinante per un giudizio di condanna. L'incidenza dell'algoritmo è dunque di gran lunga superiore rispetto all'uso che ne veniva fatto in precedenza e, conseguentemente, i

[53] Nelle motivazioni della sentenza, la Corte suprema del Wisconsin ha evidenziato che algoritmi a base statistica, come *Compas*, non prevedono la specifica verosimiglianza che un determinato imputato commetterà un nuovo reato. Al contrario, forniscono una previsione basata su una comparazione di informazioni del singolo imputato su quelle di un gruppo di soggetti simili. Sulla scorta di tale constatazione, la Corte suprema ha concluso che *Compas* è solo uno dei fattori che possono essere considerati e ponderati nella decisione. È, in sostanza, la conferma dell'impossibilità di eliminare l'*intime conviction*, specie in tema di valutazione prognostica del rischio di recidivanza, inevitabilmente affidato all'«*intuition, instinct and sense of Justice*» del giudicante. Peraltro, nonostante i limiti imposti all'uso processuale di *Compas*, la Corte si mostra conscia dei rischi di un abuso, posto che la maggior parte dei giudici non è in grado di comprendere gli esiti degli algoritmi di calcolo del rischio e che l'esercizio del diritto di difesa appare estremamente compromesso dalla segretezza del loro meccanismo di funzionamento. Cfr., *State v. Loomis*, par. 15 e par. 99.

[54] *Malenchick v. State*, 928 Ind. S.C., 2010.

[55] È proprio quest'ultimo aspetto il più criticabile: la Corte Suprema del Wisconsin ha escluso infatti la contrazione dei diritti di difesa ridimensionando il ruolo effettivamente svolto dagli algoritmi nel processo. Contro tale conclusione si scaglia, tra gli altri, K. FREEMAN, *Algorithmic injustice: how the Wisconsin Supreme Court Failed to protect due process rights in State v. Loomis*, in *North Carolina Journal of Law & Technology*, XVIII, 2016, 75 ss., ove si propone di trasformare *Compas* da *property algorithm* a *open-source algorithm* che i difensori possano analizzare e utilizzare autonomamente e di introdurre un processo di verifica in cui un supervisore esterno alla società effettui costanti controlli sul sistema per garantirne precisione e funzionamento appropriato.

suoi effetti sono potenzialmente molto più pericolosi. La decisione, salvo un intervento in senso contrario della Corte Suprema federale, apre all'utilizzo di COMPAS anche in giudizio. Le cautele suggerite appaiono però insufficienti a rassicurare sul fatto che tali algoritmi non enfatizzino e non perpetuino i problemi di discriminazione sociale, sessuale e razziale. Sotto quest'ultimo profilo, per esempio, da alcuni studi sul funzionamento degli algoritmi predittivi è emerso che gli imputati neri sono risultati avere uno *score* più alto rispetto a quello reale e che i non recidivi neri hanno quasi il doppio delle probabilità di essere erroneamente classificati come a rischio più elevato rispetto ai loro omologhi bianchi[56]. È stato inoltre osservato che il più grosso limite degli algoritmi predittivi è rappresentato dal fatto che si basano su di un metodo statistico, per cui i punteggi di rischio sono correlati ad una probabilità di recidiva generica (calcolata su casi simili) e non alla probabilità specifica che il soggetto a cui l'algoritmo viene applicato commetta in futuro un altro reato[57]. Quest'ultima considerazione coglie l'essenza della questione che può essere sintetizzata in una domanda: è socialmente accettabile depersonalizzare (con riferimento all'imputato) il procedimento logico che conduce a una sentenza penale di condanna?

### 4.3 GDPR e algoritmi discriminatori: un'arma spuntata?

È opportuno fare un cenno, infine, anche al contesto europeo che ha recentemente visto l'entrata in vigore del nuovo regolamento in materia di protezione dei dati personali 2016/679/UE[58], che modifica la disciplina già esistente introducendo nozioni aggiornate al nuovo orizzonte tecnologico, quali quella di profilazione, *data mining*, dati genetici e dati biometrici[59]. Con riferimento al tema trattato,

[56] Cfr., in particolare, J. ANGWIN, J. LARSON, S. MATTU, L. KIRCHNER, *Machine Bias. There's software used across the country to predict future criminals. And it's biased against blacks*, in ProPublica, 2016, disponibile su https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing; L. KIRCHNER, *When Big Data Becomes Bad Data*, in ProPublica, 2015, disponibile su https://www.propublica.org/article/when-big-data-becomes-bad-data (ultima consultazione 23/04/2019). *ProPublica* è un'agenzia stampa indipendente e no-profit, che si occupa di giornalismo investigativo e che ha nel suo statuto la missione di portare alla luce gli abusi di potere e il tradimento della fiducia pubblica da parte del governo, del mondo degli affari e di altre istituzioni.

[57] Cfr. anche C. COSTANZI, *La matematica del processo: oltre le colonne d'Ercole della giustizia penale*, in *Quest. Giust.*, 4, 2018.

[58] 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).

[59] In Europa il diritto alla protezione dei dati personali è sancito come diritto fondamentale della persona dall'art. 8 della Carta dei diritti fondamentali dell'Unione europea che lo riconosce come diritto autonomo; ad esse deve aggiungersi l'art. 16, par. 1, TFUE che attribuisce a ciascuno il diritto alla protezione dei propri dati; la principale fonte normativa primaria è stata fino all'entrata in vigore del nuovo Regolamento (GDPR), la direttiva 95/46/CE che sanciva il principio del consenso informato come presupposto indefettibile dell'esercizio del diritto alla protezione dei propri dati personali. Il nuovo Regolamento conferma e rafforza questo presupposto,

di particolare interesse è la disciplina sulla profilazione che viene definita come «qualsiasi forma di trattamento automatizzato di dati personali consistente nell'utilizzo di tali dati personali per valutare determinati aspetti personali relativi a una persona fisica, in particolare per analizzare o prevedere aspetti riguardanti il rendimento professionale, la situazione economica, la salute, le preferenze personali, gli interessi, l'affidabilità, il comportamento, l'ubicazione o gli spostamenti di detta persona fisica»[60].

Come abbiamo visto, le conseguenze giuridiche della profilazione potrebbero essere molteplici e ledere anche i diritti fondamentali della persona: il diniego di attraversamento di una frontiera, l'adozione di misure di sicurezza, l'esclusione da forme di assistenza sociale, il diniego di un impiego, il rifiuto della concessione di un prestito. Per questo sono imposte una lunga serie di cautele e di obblighi al titolare del trattamento, primo fra tutti l'obbligo di informare gli interessati dell'esistenza di una decisione basata sul trattamento di dati automatizzato comprendente la profilazione[61]. Nell'informativa devono infatti essere esplicitate le modalità e le finalità della profilazione. Inoltre, deve essere chiarita la logica inerente al trattamento e le conseguenze previste per l'interessato a seguito di tale tipo di trattamento, intendendo in tal senso i criteri utilizzati per giungere alla decisione, senza necessariamente dover fornire una spiegazione complessa degli algoritmi utilizzati o l'apertura della scatola nera dell'algoritmo. L'articolo 22 del GDPR, paragrafo 1, chiarisce l'ambito di applicazione delle norme in materia di profilazione che riguarda le ipotesi in cui tale attività produca effetti giuridici o incida in modo significativo sulla persona dell'utente, e la decisione sia basata interamente sul trattamento automatizzato dei dati. Questo, almeno per il momento, dovrebbe evitare casi come quelli americani e canadesi poiché il regolamento è piuttosto netto nell'affermare il diritto di non essere sottoposti a una decisione giuridica che riguardi o che incida significativamente sulla persona basata unicamente sul trattamento automatizzato, inclusa la profilazione. Vi è dunque la dif-

---

stabilendo che i dati personali debbano essere trattati in modo lecito, corretto e trasparente nei confronti dell'interessato (art. 5, GDPR) e introduce nuove e più chiare definizioni (art. 4), sancisce il diritto all'oblio (art. 17), affronta il tema dei *big data* e dei trattamenti dati automatizzati (art. 22), impone ulteriori cautele e responsabilità in capo ai titolari del trattamento e inasprisce le sanzioni. Per un approfondimento sulla strategia europea in materia di intelligenza artificiale e *big data* si suggerisce l'interessante rapporto dello *European Political Startegy Centre* della Commissione Europea, pubblicato il 27 marzo 2018 e consultabile all'indirizzo: https://www.labparlamento.it/wp-content/uploads/2018/04/epsc_strategicnote_ai.pdf (ultima consultazione 23/04/2019).

[60] Art. 4, comma quarto, Regolamento (UE) 2016/679.

[61] «1. 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. 2. 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 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. 3. Nei casi di cui al paragrafo 2, lettere a) e c), il titolare del trattamento attua misure appropriate per tutelare i diritti, le libertà e i legittimi interessi dell'interessato, almeno il diritto di ottenere l'intervento umano da parte del titolare del trattamento, di esprimere la propria opinione e di contestare la decisione. 4. Le decisioni di cui al paragrafo 2 non si basano sulle categorie particolari di dati personali di cui all'articolo 9, paragrafo 1, a meno che non sia d'applicazione l'articolo 9, paragrafo 2, lettere a) o g), e non siano in vigore misure adeguate a tutela dei diritti, delle libertà e dei legittimi interessi dell'interessato». Art. 22, e Considerando 71, GDPR.

fidenza nei confronti dell'algoritmo e negli strumenti di *data mining* e di *processing* dei dati che in maniera automatizzata arrivino a prospettare conseguenze giuridiche e il rimedio individuato è un diritto all'opposizione alla procedura di tal tipo, con la possibilità di domandare l'intervento di un essere umano.

Ma lo stesso divieto di decisione esclusivamente automatizzata soffre di eccezioni. La decisione del robot è legittima qualora «sia autorizzata dal 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». L'individuo può essere inoltre sottoposto a un processo decisionale automatizzato, compreso la profilazione, anche quando: a) il trattamento sia necessario per la conclusione o l'esecuzione di un contratto tra l'interessato e il titolare (la necessità deve essere interpretata in modo restrittivo, anche se i Garanti europei precisano che i motivi di efficienza sono ritenuti sufficienti per giustificare l'utilizzo di sistemi decisionali basati su profilazione, a condizione che non vi siano metodi meno intrusivi che raggiungano lo stesso risultato), ma tale eccezione non si applica in caso di trattamento di dati sanitari; b) vi sia esplicito consenso al trattamento (un consenso distinto rispetto al consenso relativo ad altri trattamenti)[62].

Infine, allo scopo di prevenire violazioni dei diritti e delle libertà degli individui, il trattamento basato su sistemi automatizzati (anche se non esclusivamente basato sulla valutazione della macchina, ma anche quello con intervento umano) deve essere preceduto dalla valutazione di impatto[63], proprio perché dalle elaborazioni possono derivare dettagli informativi ritenuti di natura particolarmente invasiva ma anche perché possono essere impiegati una quantità significativa di dati ai quali devono essere assicurati gli opportuni livelli di protezione e garanzia. A tal fine, il regolamento impone altresì che il titolare adotti una politica di periodica revisione dei sistemi di decisione automatizzata, per verificare se producono errori, classificazioni non corrette e discriminazioni.

Indubbiamente il nuovo Regolamento europeo è indice di un mutato contesto giuridico e politico in cui si è assistito ad una trasformazione o meglio ad una riformulazione ed estensione del diritto alla riservatezza, che, da guscio protettivo della persona, va evolvendo a garanzia del patrimonio informativo digitalizzato e circolante, come denominatore comune di tutte le realtà industriali, commerciali, culturali, pubbliche e private e come riferimento costante in tutti i campi di attività. Gli strumenti di tutela, quindi, devono essere adeguati all'evoluzione delle innovazioni tecnologiche, risultando, pertanto, strutturalmente dinamici. Nel nuovo mondo digitale, dunque, la protezione dei dati, intesa come controllo effettivo delle proprie informazioni personali, è un diritto fondamentale in Europa e, in quanto tale, deve essere tutelato. Non sorprende se parte della dottrina ha visto proprio nell'avvento della Carta dei diritti fondamentali la spinta verso «un'ondata costituzionalizzatrice»[64] che ha determinato una reazione da parte delle Istituzioni e, in particolar modo, della Corte di Giusti-

[62] Art. 22, par. 2, GDPR.

[63] «Quando un tipo di trattamento, allorché prevede in particolare l'uso di nuove tecnologie, considerati la natura, l'oggetto, il contesto e le finalità del trattamento, può presentare un rischio elevato per i diritti e le libertà delle persone fisiche, il titolare del trattamento effettua, prima di procedere al trattamento, una valutazione dell'impatto dei trattamenti previsti sulla protezione dei dati personali. Una singola valutazione può esaminare un insieme di trattamenti simili che presentano rischi elevati analoghi». Art. 35, par. 1, GDPR.

[64] Così in O. POLLICINO, *La «transizione» dagli atomi ai bit nel reasoning delle Corti europee*, in *Ragion pratica*, 1, 2015, 53-82.

zia nel dar vita a una "nuova era" del diritto alla privacy: a partire dai casi *Digital Rights Ireland*[65] e *Google Spain*[66], infatti, i giudici di Lussemburgo hanno ampliato il più possibile il margine di protezione dei dati personali e della privacy degli individui allo scopo di realizzare un quadro globale, coerente, solido e moderno per la protezione dei dati nell'Unione. Un processo di rinvigorimento e sostanzializzazione dei diritti alla privacy e alla *data protection* confermato anche dal caso *Schrems*[67] che ha costituito la risposta, in parte anche emotiva, dell'Europa allo scandalo "*Datagate*" e alle operazioni di sorveglianza globale messe in atto dal governo degli Stati Uniti, dietro il consenso o quantomeno la complicità (forse inconsapevole) di alcuni Stati membri[68]. Tutto questo ci porta, a maggior ragione, a ritenere che il cammino del diritto alla privacy sia tutt'altro che giunto al termine ma che anzi sia solo all'inizio.

[65] Corte di Giustizia, 8 aprile 2014, cause n. C-293/12 e C-594/12, *Digital Rights Ireland, Seitlinger e altri*. Per un approfondimento si rinvia a G. FINOCCHIARO, *La giurisprudenza della Corte di giustizia in materia di dati personali da Google Spain a Schrems*, in G. RESTA, V. ZENO-ZENCOVICH (a cura di), *La protezione transnazionale dei dati personali. Dai "Safe Harbour Principles" al "Privacy Shield"*, Roma, 2016, 113-136 e L. TRUCCO, *Data retention: la Corte di giustizia si appella alla Carta UE dei diritti fondamentali*, in *Giur. it*., 8-9, 2014, 1850 ss.

[66] Corte di Giustizia, 13 maggio 2014, causa n. C-131/12, *Google Spain SL, Google Inc. v. Agencia Española de Protección de Datos e Mario Costeja Gonzàlez*. Per un approfondimento si rinvia a G.E. VIGEVANI, *Identità, oblio, informazione e memoria in viaggio da Strasburgo a Lussemburgo, passando per Milano*, in *Danno e resp*., 7, 2014, 731 ss. e T.E. FROSINI, *Diritto all'oblio e Internet*, in www.federalismi.it, 10 giugno 2014 (ultima consultazione 23/04/2019).

[67] Corte di Giustizia, 6 ottobre 2015, causa n. C-362/14, *Maximillian Schrems v. Data Protection Commissioner*.

[68] La decisione ha riguardato l'art. 25 della Direttiva 95/46/CE che stabilisce che il trasferimento di dati personali verso paesi terzi può aver luogo soltanto se il paese di destinazione garantisca un livello di protezione adeguato, fatte salve le misure nazionali di attuazione delle altre disposizioni della direttiva. Il compito di valutare l'adeguatezza del livello di protezione offerto da un paese terzo è affidato alla Commissione dal par. 6 dell'art. 25, secondo la procedura regolata all'art. 31 della direttiva. Nel caso la Commissione constati l'inadeguatezza del livello di tutela assicurato da un paese terzo, gli Stati membri sono tenuti ad adottare le misure necessarie a evitare ogni trasferimento di dati personali verso tale stato. La Corte ha proceduto a un'interpretazione costituzionalmente orientata della Direttiva ha ritenuto, contrariamente a quanto valutato dalla Commissione, che gli Stati Uniti non garantirebbero i presupposti per una tutela efficace, esponendo a pregiudizio i dati personali trasferiti dall'Unione europea soprattutto nell'ambito dei trattamenti effettuati da autorità pubbliche, trasformando il principio di adeguatezza, cui la direttiva all'art. 25, àncora la valutazione della legittimità del trasferimento di dati personali verso paesi terzi, in principio di protezione sostanzialmente equivalente della tutela dei diritti fondamentali in gioco. Cfr., tra gli altri, M. NINO, *Il caso Datagate: i problemi di compatibilità del programma di sorveglianza PRISM con la normativa europea sulla protezione dei dati personali e della privacy*, in *Diritti umani e diritto internazionale*, 3, 2013, 727-746; G. RESTA, *La sorveglianza elettronica di massa e il conflitto regolatorio USA/UE*, 23-48; V. ZENO-ZENCOVICH, *Intorno alla decisione nel caso Schrems: la sovranità digitale e il governo internazionale delle reti di telecomunicazione*, 7-22; P. PIRODDI, *I trasferimenti di dati personali verso Paesi terzi dopo la sentenza Schrems e nel nuovo regolamento generale sulla protezione dei dati*, 169-214, tutti in G. RESTA, V. ZENO-ZENCOVICH (a cura di), *La protezione transnazionale dei dati personali. Dai "Safe Harbour Principles" al "Privacy Shield"*, Roma, 2016; F. COUDERT, *Schrems vs. Data Protection Commissioner: A Slap on the Wrist for the Commission and New Powers for Data Protection Authorities*, in www.europeanlawblog.eu, 15 ottobre 2015; P. FALLETTA, *La Corte di Giustizia, ancora una volta, contro le multinazionali del web (riflessioni su Corte di Giustizia UE (Grande sezione), 6 ottobre 2015, Schrems c. Data Protection Commissioner, C-362/14)*, in *federalismi.it*, 23 dicembre 2015; M. BASSINI, O. POLLICINO, *La Corte di giustizia demolisce il safe harbor e ridisegna i confini del diritto alla privacy in ambito transnazionale*, in www.diritto24.ilsole24ore.com, 7 ottobre 2015 (ultima consultazione 23/04/2019).

## 5. Osservazioni conclusive

Il quadro delineato apre a numerose riflessioni per il futuro e, al contempo, evidenzia dubbi non risolti: nonostante un impianto normativo volto a valorizzare l'autodeterminazione dell'individuo e a proteggerlo contro attività, come quella della profilazione, considerate estremamente invasive e potenzialmente lesive dell'eguaglianza e dei diritti fondamentali della persona[69], il GDPR resta ancorato a una visione "vecchia" del mondo digitale e tecnologico e non apporta alcuna reale innovazione nella tutela del corpo digitale. Può bastare la manifestazione, benché esplicita, del consenso al trattamento dei propri dati[70] e la trasparenza sull'uso dei dati – intesa come intellegibilità e comprensibilità del trattamento e dei suoi fini – per assicurare il rispetto della privacy e dei diritti fondamentali dell'utente? Può bastare – e, se sì, come si concilia con la tutela del segreto industriale e della proprietà intellettuale – l'apertura della c.d. *black box* dell'algoritmo per escludere errori e discriminazioni nel *decision-making* automatizzato? Può bastare la disciplina della "raccolta" e del "trattamento dei dati" e la tutela contro le pratiche di "profilazione" e di decisione "automatizzata" a proteggere adeguatamente il corpo digitale? L'approccio del GDPR appare deficitario laddove si scontra con gli ultimi sviluppi in materia di *big data*, tracciabilità, profilazione, algoritmi di *machine learning* e di *deep learning*, aggregazione dei dati da parte di più soggetti nel tempo, *digital surveillance*, *Internet of Things*. Da un lato, appare un po' ingenua la pretesa di assicurare la trasparenza e il consenso informato a ogni tipo di trattamento in uno scenario caratterizzato dalla complessità di questi fenomeni[71]. Per esempio, la profilazione può avvenire utilizzando dati individuali o identificativi (come quelli anagrafici), oppure dati aggregati derivanti da dati personali individuali. Il livello di aggregazione è variabile, e quindi potrebbe accadere che i dati utilizzati, anche se in forma aggregata, consentano comunque, a seguito dell'incrocio con altri dati, l'identificazione dei soggetti interessati e l'emersione anche di informazioni sensibili[72]. Dall'altro, vista l'assoluta centralità dei dati nello sviluppo di algoritmi predittivi e di apprendimento automatico, la *data protection* si sovrapporrà con la tutela di altri interessi pubblici e privati, quali la proprietà intellettuale.

Inoltre, gli algoritmi non sono certamente perfetti e per questo motivo possono portare anche a errori che sono il riflesso dei dati che hanno a disposizione. Anche l'attenzione correttamente posta sul

[69] Per una disamina più approfondita del nuovo assetto europeo in materia di protezione dei dati personali si rimanda, tra gli altri, a F. PIZZETTI, *Privacy e il diritto europeo alla protezione dei dati personali: dalla Direttiva 95/46 al nuovo Regolamento europeo*, Torino, 2016; A. PISAPIA, *La tutela per il trattamento e la protezione dei dati personali*, Torino, 2018.

[70] I «dati personali» nel nuovo Regolamento sono definiti in senso molto ampio come «qualsiasi informazione riguardante una persona fisica identificata o identificabile [...] direttamente o indirettamente, con particolare riferimento a un identificativo come il nome, un numero di identificazione, dati relativi all'ubicazione, un identificativo online o uno o più elementi caratteristici della sua identità fisica, fisiologica, genetica, psichica, economica, culturale o sociale». Art. 4, par. 1, GDPR.

[71] Così anche C. CASONATO, *Intelligenza artificiale e diritto costituzionale: prime considerazioni*, in *Dir. pub. comp. eur.*, numero speciale, 2019, 107, il quale parla di un «consenso consapevolmente disinformato» per riferirsi a una categoria – quella appunto del consenso informato – che è divenuta una «mera finzione che, con il nostro consapevolmente disinformato accordo, ci espone quotidianamente ad essere profilati in ogni nostra dimensione e attività».

[72] Per un approfondimento si consiglia M.R. CALO, *Robots and Privacy*, in P. LIN, K. ABNEY, G.A. BEKEY (a cura di), *Robot Ethics: The Ethical and Social Implications of Robotics*, Londra, 2012.

concetto della "correttezza e trasparenza" dei dati, vale a dire la disponibilità di strumenti in grado di verificarne l'affidabilità, quali per esempio accurati sistemi di *audit* per gli algoritmi e di certificazione per i loro sviluppatori, come raccomandato, per esempio, dal Comitato «Scienza e tecnologia» della Camera dei Comuni del Regno Unito[73], rischia di rivelarsi insufficiente (specialmente dopo la comparazione con i casi americani e canadesi).

A questo, si aggiunge un ulteriore problema legato all'impossibilità di aprire la scatola nera degli algoritmi, ovvero di accedere al *software* alla base del loro funzionamento poiché generalmente protetto dal segreto commerciale[74]. In tal senso, quindi, anche i soggetti che utilizzano tali algoritmi, se non ne sono i programmatori, potrebbero non conoscere affatto le logiche alla base degli stessi, e dunque non comprendere fino in fondo gli effetti della loro applicazione. Il regolamento europeo prevede un apposito obbligo di informazione sulla logica alla base della profilazione ma, ancora una volta, rischia di non rivelarsi efficace, spesso perché per gli stessi programmatori è difficile darne conto, specialmente nei casi in cui l'intelligenza artificiale è in grado di crescere e apprendere autonomamente, appropriandosi dei dati stessi (che diventano parte integrante dell'algoritmo[75]). «La moderna *data protection*, infatti, sembra non prendere direttamente in considerazione questo rapporto nella sua totalità, fermandosi alla disciplina della raccolta e del trattamento dei dati [...]. Ma *quid iuris* nel caso in cui i dati fuoriusciti dalla sfera di dominio dell'individuo, continuino a essere utilizzati al di fuori del contesto nel quale erano stati forniti o si trasformino addirittura in tanti ulteriori *layer* di una rete neurale usata da un algoritmo di intelligenza artificiale?»[76].

Se il ricorso alla privacy e alla *data protection* mostra delle criticità, nonostante i recenti sviluppi soprattutto in ambito europeo, limiti ancora maggiori evidenzia la tutela offerta dal diritto antidiscrimi-

[73] Nel rapporto su «*Algorithms in decision-making*» pubblicato il 23 maggio 2018, il Comitato «Scienza e tecnologia» della Camera dei Comuni inglese ha opportunamente ricordato che la tecnologia deve essere utilizzata per migliorare la qualità dei servizi pubblici e guidare l'innovazione, in particolare in settori come i trasporti e la sanità. Alla base di tutti vi sono i dati, soprattutto quelli in mano al settore pubblico, sui quali operano algoritmi che non sono affatto la formula magica che produce automaticamente benefici in assenza di un quadro accurato di regole, anche di carattere etico. In sostanza, il rapporto evidenzia che l'applicazione degli algoritmi, al pari di ogni decisione umana, può essere condizionata da errori che comportano esiti imprevedibili e talora discriminatori, soprattutto nei confronti di determinate categorie sociali, se il loro funzionamento non è corretto oppure viene alterato. Per evitare tale rischio, il Parlamento invita il governo ad affidare al «*Centre for Data Ethics and Innovation*», organismo consultivo che sta per essere creato, previsto dalla legge di bilancio dello scorso anno, il compito di verificare il funzionamento degli algoritmi. Il che significa controllare e garantire la qualità dei dati sui quali gli stessi si basano assicurandosi che i loro sviluppatori siano in grado di spiegare come funzionano. Tali meccanismi dovrebbero infatti essere pubblicati e conoscibili a tutti, nel momento in cui incidono sui diritti e la libertà dei cittadini. Il Rapporto è consultabile all'indirizzo: https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/351/351.pdf (ultima consultazione 23/04/2019).

[74] Cfr., F. PASQUALE, *The Black Box Society*, op. cit.; Cfr. anche, sempre in tema di *big data* e concorrenza, F. DI PORTO (a cura di), *Big data e concorrenza*, in *Mercato e concorrenza*, 23, 2016, 5-14.

[75] Si veda, fra gli altri, R. PISELLI, *La protezione dei dati personali al tempo degli algoritmi intelligenti e dei robot umanoidi*, in *LLR*, 2, 2017, 197. L'Autore parla di un vero e proprio processo di "appropriazione" e "rielaborazione" dell'informazione da parte della macchina, divenendo quell'informazione (o un frammento di essa) parte integrante dell'algoritmo stesso. «La relazione che si instaura tra dati e algoritmo è bidirezionale. Da un lato, l'algoritmo processa e interpreta i dati che compongono l'informazione, dall'altro, l'informazione fa evolvere l'algoritmo, diventando essa stessa componente dell'algoritmo».

[76] *Idem*, cit. 197-198.

natorio. La profilazione e le decisioni automatizzate basate sui *big data* pongono problemi ancor più difficilmente risolvibili facendo ricorso agli schemi classici della tutela antidiscriminatoria: sebbene, rispetto al modello americano, il quadro europeo offra un più ampio elenco di fattori "protetti"[77] e di misure proattive per il contrasto alla discriminazione, permangono le medesime criticità che derivano soprattutto dalla mancanza di un soggetto umano (c.d. *bad actor*) che, esplicitamente o implicitamente, discrimina un certo individuo o gruppo e dall'attenuazione del nesso causale tra il risultato discriminatorio e la caratteristica "protetta" che determina la disparità di trattamento. Mentre è piuttosto evidente come la fiducia nella neutralità e nella razionalità matematica dell'algoritmo offuschi la possibilità che esso possa cadere in errore e che, elaborando dati intrisi di pregiudizio e stereotipi, riproduca e perpetui la discriminazione e la segregazione nei confronti di certi gruppi sociali, meno immediato è invece il fatto che la "normalità" – che rappresenta l'implicito fondamento della categorizzazione delle differenze rilevanti ai fini della tutela antidiscriminatoria – non sia più definita dalle norme sociali, morali o giuridiche bensì dalle correlazioni statistiche dei dati. Pertanto anche intervenendo sulle istruzioni di funzionamento dell'algoritmo, ordinandogli di non tenere conto della «razza» o del «sesso» o dell'«orientamento sessuale» delle persone, non è possibile escludere che aggregando e processando dati e frammenti dati, le correlazioni statistiche alla base del suo funzionamento non lo conducano a risultati comunque discriminatori sulla base di quei fattori. Tuttavia, nel momento in cui si esclude formalmente che l'algoritmo abbia tenuto conto di quei dati sensibili per giungere alla sua decisione, di fatto si depotenzia il rimedio antidiscriminatorio perché si cancella o comunque si attenua il nesso causale tra l'ingiusto trattamento e la caratteristica "protetta". Come si può dopo sostenere di essere stati discriminati sulla base della razza se si dimostra che l'algoritmo è stato programmato per non tenere conto proprio della razza? Per esempio, negli Stati Uniti l'*Home Mortgage Disclosure Act* (HMDA) impone alle banche di divulgare i dati sulla «razza» nelle decisioni di concessione o rifiuto dei mutui per l'acquisto della casa per scongiurare una discriminazione algoritmica a danno della popolazione di colore. Uno studio ha tuttavia dimostrato che, nonostante l'approccio *color blind* imposto dalla pubblica amministrazione, le persone nere e ispaniche risultano essere escluse dai prestiti tre volte di più delle persone bianche; similmente, anche coloro che chiedono un mutuo per l'acquisto di proprietà situate in aree ad alta concentrazione di neri e ispanici hanno una maggiore probabilità di vedersi rifiutato il prestito[78].

[77] Bisogna tuttavia segnalare come anche nel quadro antidiscriminatorio europeo vi siano lacune significative per esempio con riferimento alla tutela dell'identità di genere, dell'espressione di genere e delle caratteristiche sessuali (c.d. SOGIESC – *Sexual orientation, gender identity and expression, sexual characteristics*). La disciplina del GDPR non fa eccezione: mentre tra i fattori "protetti" sono inseriti l'orientamento sessuale e la vita sessuale (art. 9, comma primo), non vi è alcun accenno all'identità di genere. Questo rappresenta un grosso limite considerando quanto incidono nella profilazione dell'individuo questo genere di informazioni. Un'interpretazione estensiva delle nozioni di «dato genetico» e «dato biometrico» potrebbe essere quindi la sola soluzione per tamponare tale lacuna, anche alla luce del fatto che l'interpretazione estensiva della categoria «sesso», tradizionalmente usata dalle Corti per tutelare anche le persone transessuali e intersessuali, sembrerebbe limitata dal fatto che tale caratteristica non è stata inserita tra i «*sensitive grounds*» ai sensi del GDPR.

[78] Cfr., per esempio, A.H. Munnel, G.M.B. Toonell, L.E. Browne, J. McEneaaney, *Mortgage lending in Boston: interpreting HDMA data*, in *The American Economic Review*, 1, 1996, 25-53.

Sebbene le soluzioni proposte da autori come Frank Pasquale, il quale invoca l'apertura della scatola nera degli algoritmi in nome della piena trasparenza[79], abbiano un certo grado di persuasività, se non altro per il fatto che impongono sin dalla progettazione dell'algoritmo di tenere in considerazione gli effetti discriminatori che possono derivarne, non convincono del tutto. Infatti, se il problema della discriminazione algoritmica risiedesse soltanto nel modo in cui un algoritmo viene progettato e può essere manipolato, allora aprire la scatola nera, garantendo la totale trasparenza dei codici e delle istruzioni che lo fanno funzionare, sarebbe indubbiamente la risposta al problema. Tuttavia, se, come è, il problema della discriminazione algoritmica risiede non tanto nel modo in cui è stato progettato ma nel modo in cui esso replica la discriminazione del mondo reale attraverso i dati che utilizza, allora la c.d. *transparency by design* non è sufficiente, anzi può diventare un ostacolo in più per l'applicazione del diritto antidiscriminatorio poiché attenua (se non addirittura elimina) il nesso causale tra risultato discriminatorio e caratteristiche "protette". Cosicché «revealing the facially neutral algorithm may help defend that algorithm from accusations of discrimination»[80]. È necessario, allora, concentrarsi sulla trasparenza degli *input* ed *output* dell'algoritmo, ovvero sulla qualità dei dati in ingresso e dei risultati in uscita. Focalizzarsi su questi elementi, piuttosto che sul modo in cui l'algoritmo opera, diventa particolarmente utile soprattutto quando i meccanismi computazionali e il ragionamento artificiale diventano sempre più complessi, talvolta incomprensibili anche per i progettatori stessi.

Una proposta di riabilitazione della tutela antidiscriminatoria di fronte alle sfide dell'intelligenza artificiale potrebbe allora essere quella di sviluppare un «algorithmic affirmative action»[81], ovverosia un piano di pratiche e azioni proattive «not to focus on identifying the *how* of discrimination, but on working on to correct it, *regardless of its source*»[82]. Lo scopo deve essere quello di correggere gli effetti discriminatori che derivano da una realtà sociale e culturale intrinsecamente discriminatoria, anziché concentrarsi su come evitare che gli algoritmi discriminino, perché tanto lo faranno e più si cercherà di renderli "neutrali", più diventerà difficile coglierne gli effetti perpetuativi delle diverse forme di razzismo. «The counterintuitive result of affirmative action is that the decisionmaker must take race and gender into account in order to ensure the fairness of the result. This is what struck Chief Justice John Roberts as implausible: "The way to stop discrimination on the basis of race is to stop discriminating on the basis of race". The obvious remedy to the problem of manipulations of algorithms that produce racist or sexist outcomes would seem to be to mandate race or gender neutrality. In reality, however, even while neutrality is certainly better than hard-coded racism or sexism, racial or sex neutrality would in fact perpetuate the problem of algorithmic replication of existing racism. Justice Sonia Sotomayor responded sharply to Chief Justice Roberts' claim in a recent opinion: "The way to stop discrimination on the basis of race is to speak openly and candidly on the subject of race, and to apply the Constitution with eyes open to the unfortunate effects of centuries of racial discrimination"»[83].

[79] F. PASQUALE, *The Black Box Society: The Secret Algorithms That Control Money and Information*, Cambridge - Londra, 2015.
[80] A. CHANDER, *The Racist Algorithm?*, in *Michigan L. Rev.*, 115, 2017, 1040.
[81] *Idem*, 1041.
[82] *Ibidem*.
[83] *Ibidem*.

È pertanto indispensabile, per riabilitare i rimedi antidiscriminatori, prevedere azioni positive volte a riconoscere la "discriminazione virale" che risiede nei dati, quegli stessi dati utilizzati dalle intelligenze artificiali per prendere le loro decisioni. Allora, piuttosto che richiedere l'accesso alla scatola nera degli algoritmi, bisogna richiedere l'accesso ai dati in entrata da un lato, per avere la possibilità di verificarli e di correggerli qualora siano imparziali, non rappresentativi o inaccurati e ai dati in uscita dall'altro, per identificare e bloccare l'eventuale discriminazione indiretta. In questo modo, come suggerisce Frank Pasquale, «[e]ven if algorithms at the heart of these processes "transcend all understanding", we can inspect the inputs (data) that go into them, restrict the contexts in which they are used, and demand outputs that avoid disparate impacts»[84].

Allo stato attuale, pertanto, il diritto antidiscriminatorio rischia di non superare le sfide prospettate dall'avvento dei *big data* e del *decision-making* algoritmico e su questo fronte il GDPR non aggiunge nulla agli strumenti di tutela tradizionali. I corpi elettronici – i nostri *alter ego* virtuali ricostruiti dall'infinita quantità di dati che ci lasciamo dietro quotidianamente – sono, nello spazio digitale, nudi e sforniti di tutela contro le discriminazioni, nonostante siano sempre più protagonisti delle scelte e delle decisioni che poi si ripercuotono sui diritti fondamentali dei loro proprietari in carne ed ossa, i quali finiscono per subire sommessamente la "verità della macchina" senza sapere nemmeno il perché.

[84] F. PASQUALE, *Bittersweet Mysteries of Machine Learning (A Provocation)*, in *London Sch. Econ. & Pol. Sci.: Media Pol'y Project Blog*, 5 febbraio 2016, consultabile all'indirizzo web: http://blogs.lse.ac.uk/mediapolicyproject/2016/02/05/bittersweet-mysteries-ofmachine-learning-a-provocation/ [https://perma.cc/XSS9-2D58] ((ultima consultazione 23/04/2019).