GenIUS

RIVISTA DI STUDI GIURIDICI
SULL'ORIENTAMENTO SESSUALE E L'IDENTITÀ DI GENERE

DOLORES MORONDO TARAMUNDI

Le sfide della discriminazione algoritmica

PUBBLICAZIONE TELEMATICA SEMESTRALE REGISTRATA PRESSO IL TRIBUNALE DI BOLOGNA · ISSN 2384-9495

online first
24 ottobre 2022

Dolores Morondo Taramundi *

Le sfide della discriminazione algoritmica

Sommario

Abstract

Con l'espansione dei sistemi decisionali automatizzati e semiautomatici è cresciuto il dibattito sui problemi etici che l'uso di tali tecnologie può presentare per la democrazia, i diritti fondamentali e la giustizia sociale. Questo contributo si propone di affrontare il problema della discriminazione algoritmica e di analizzare alcune delle sfide che questo fenomeno (o insieme di fenomeni) rappresenta per i meccanismi di protezione esistenti, in primo luogo per la normativa antidiscriminatoria, nonché di prendere in considerazione le questioni che esso apre in relazione al quadro normativo che si sta sviluppando in Europa.

With the expansion of automated and semi-automated decision systems, the debate on the ethical problems that the use of such technologies can pose for democracy, fundamental rights and social justice has been growing. This paper aims to approach the specific problem of algorithmic discrimination and analyze some of the challenges that this phenomenon (or set of phenomena) poses to existing protection mechanisms, fundamentally anti-discrimination law, and what challenges it creates in relation to the regulatory framework that is developing in Europe.

1. Introduzione

Le decisioni automatizzate e semiautomatiche stanno diventando sempre più frequenti in tutti gli ambiti della vita. Sia le pubbliche amministrazioni che le aziende private utilizzano algoritmi progettati per aiutare o sostituire le persone incaricate di prendere decisioni. Soluzioni basate su algoritmi possono essere trovate nell'istruzione e nel mercato del lavoro, anche nella selezione degli studenti da parte delle università o nei processi di assunzione per coprire i posti vacanti[1]; nel settore dei servizi

* Investigadora Principal de Instituto de Derechos Humanos, Universidad de Deusto. Relazione al Convegno "L'algoritmo alla prova del caso concreto: stereotipi, serializzazione, discriminazione" ospitato dall'Università di Ferrara in data 6.4.2022. Contributo non sottoposto a referaggio a doppio cieco.

1 Ad esempio, A. Köchling, M.C. Wehner, *Discriminated by an algorithm: a systemic review of discrimination and fairness by algorithm decision-making in the context of HR recruitment and HR development*, in *Business Research*, 2020, 13 (3), p. 795 ss.

sociali, nel prendere decisioni in merito alla richiesta di sussidi o benefici, o nella determinazione dei bisogni di protezione sociale[2]; e anche in ambiti molto delicati del potere pubblico, come quando la polizia deve accertare la credibilità di una denuncia o la probabilità che le vittime di violenza domestica vengano nuovamente attaccate, in relazione all'accesso ad adeguate misure di prevenzione e protezione[3]. Le soluzioni basate su algoritmi promettono di limitare l'arbitrarietà e l'incoerenza dei decisori, di valutare volumi di dati in tempi che sarebbero impossibili per le menti umane, e tutto questo ad un costo molto inferiore. Il processo decisionale automatizzato si presenta così come uno strumento per aumentare la precisione e l'efficienza, sia nel contenuto e nella qualità delle decisioni (decisioni migliori, basate sui dati e coerenti nell'insieme), sia nel processo di maturazione delle decisioni (più veloce, meno costoso, meglio informato, esente da errore umano o negligenza, arbitrarietà o pregiudizio).

Gli algoritmi sono degli insiemi di istruzioni matematiche intese a risolvere una classe di problemi o a eseguire calcoli. La letteratura scientifica ha sviluppato la descrizione e la classificazione di diversi tipi di algoritmi e delle correlate tecnologie dell'informazione (TI), che potrebbero interagire con le persone chiamate a prendere decisioni[4]. Gli algoritmi basati su regole, ad esempio, sono stati studiati dalla filosofia giuridica sin dagli anni '80, nella speranza di facilitare o assistere i compiti legati all'azione del giudice. Il fallimento dei primi tentativi di raggiungere risultati pratici significativi è stato attribuito alle importanti dimensioni semantiche e pragmatiche del diritto e delle decisioni legali, diversamente dai modelli algoritmici basati sulle funzioni logiche del linguaggio[5]. Le tecniche di elaborazione del linguaggio naturale hanno tentato di colmare questa lacuna, di modo che i sistemi esperti di oggi possono operare con più funzioni per automatizzare dati e documenti e fornire assistenza intelligente[6]. Un potenziale ancora maggiore può essere trovato negli algoritmi di apprendimento automatico. I sistemi di apprendimento automatico definiscono i propri insiemi di regole in base agli output dei dati. A differenza dei sistemi deterministici basati su regole, i sistemi di apprendimento automatico basati sui dati sono probabilistici e cercano di risolvere i problemi rilevando i mo-

---

L'impatto del processo decisionale algoritmico sul mercato del lavoro è stato valutato in report commissionati dall'Organizzazione Internazionale del Lavoro (V. De Stefano, *"Negotiating the algorithm": Automation, artificial intelligence and labour protection*, Geneva, ILO, 2018) e dalla Commissione Europea (A.J. Wood, *Algorithmic management. Consequences for work organisation and working conditions*, Seville, European Commission, 2021); sull'uso del processo decisionale algoritmico nell'istruzione superiore, P. Prinsloo, *Of "black boxes" and algorithmic decision-making in (higher) education – A commentary*, in *Big Data and Society*, 2020, 7 (1).

2 In *Automating Inequality*, Virginia Eubanks indaga su tre casi di processo decisionale automatizzato e semiautomatizzato nell'accesso ai benefici sociali negli Stati Uniti. V. Eubanks, *Automating inequality. How high-tech tools profile, policee, and punish the poor*, New York, St Martin's Press, 2018.

3 Ad esempio, il sistema VioGén utilizzato dalla polizia spagnola raccoglie informazioni personali, sociali, della polizia e giudiziarie su ogni caso di violenza da partner intimo, fa previsioni sul livello di rischio delle vittime, stabilisce misure di protezione relative al livello di rischio e pertinenti avvisi informativi (ad esempio, quando l'aggressore viene rilasciato dal carcere).

4 Si vedano, tra tutti, i risultati del progetto e-SIDES (Ethical and social impacts of data science), disponibili su https://e-sides.eu/.

5 A. Martino, *Inteligencia artificial y derecho. Acerca de lo que hay*, in *Revista de ciencia de la legislación*, 2019 (6).

6 Si veda, ad esempio, PROMETEA, un sistema esperto di Intelligenza Artificiale (AI) creato dal Ministero Pubblico di Buenos Aires, Argentina (https://mpfciudad.gob.ar/institucional/2020-03-09-18-42-38-innovacion-e-inteligencia-artificial). I risultati di PROMETEA indicano che il Pubblico Ministero ha aumentato la propria velocità di elaborazione del 275% tra il 2017 e il 2018, poiché con PROMETEA un dossier può essere elaborato in 5 giorni anziché in 3 mesi. Vedi J.I. Solar Cayón, *La inteligencia artificial jurídica: nuevas herramientas y perspectivas metodológicas para el jurista*, in *Revus*, 2020 (4).

delli (*patterns*) di informazione nell'analisi dei *big data*.

Le decisioni basate su algoritmi, come qualsiasi altra decisione, possono avere un impatto di vario tipo sui diritti umani e le libertà fondamentali. Rispetto alle decisioni distorte che gli esseri umani possono prendere da soli, le decisioni automatizzate e semiautomatizzate, in quanto basate su criteri discriminatori, pongono un problema di scala, poiché esse vengono prese molto più velocemente e applicate ad un numero maggiore di individui: ciò riduce la probabilità di identificare e affrontare tempestivamente il problema, mentre il volume dei casi errati rende difficile esaminarli e correggerli. Gli algoritmi di apprendimento automatico possono anche discriminare su larga scala attraverso soluzioni che riproducono modelli di disuguaglianza incorporati nei dati da cui apprendono[7].

Con la crescita della preoccupazione politica per le implicazioni etiche dell'uso dell'intelligenza artificiale (IA), si stanno ampliando anche le conoscenze scientifiche e pratiche sull'argomento, con contributi interdisciplinari di professionisti di vari settori (IT, ingegneria e diritto). Sebbene la maggior parte dell'attenzione accademica si sia concentrata sulla violazione di diritti specifici (come il diritto alla privacy o alcuni principi generali del diritto penale) che possono derivare dall'uso, dall'uso improprio o dall'uso fraudolento delle tecnologie, il problema specifico della discriminazione algoritmica inizia a delinearsi e a richiedere attenzione specifica. Finora la ricerca su discriminazione e algoritmi si era concentrata principalmente sul contesto statunitense[8], ma ci sono oramai numerose analisi del fenomeno in Europa[9]. Il rapporto tra IA e discriminazione, e l'impatto che l'IA può avere sulla disuguaglianza, sul principio di uguaglianza e non discriminazione, sono stati oggetto di studi commissionati sia dal Consiglio d'Europa[10], sia dalla Commissione Europea[11], insieme ad altri realizzati da agenzie internazionali e nazionali[12].

Con un quadro normativo ancora in fase di sviluppo, è importante poter identificare cos'è la discriminazione algoritmica e come affrontarla. Al momento, la maggior parte dei casi che giungono all'attenzione della stampa e al dibattito pubblico sono legati a risultati discriminatori prodotti dall'utilizzo di dati personali sensibili, quali il sesso o l'origine etnica o razziale. Tuttavia, le normative

7 R. Xenidis, L. Senden, *EU non-discrimination law in the era of artificial intelligence: Mapping the challenges of algorithmic discrimination*, in U. Bernitz, X. Groussot, J. Paju, S.A. de Vries (eds.), *General Principles of EU Law and the EU Digital Order*, Alphen aan den Rijn, Kluwer Law, 2020, p. 151 ss.

8 S. Barocas, D. Selbst, *Big Data's disparate impact*, in *California Law Review*, 2016 (104), p. 671 ss.; T.B. Gillis, J.L. Spiess, *Big Data and discrimination*, in *University of Chicago Law Review*, 2019 (86), p. 459 ss., e la bibliografia lì citata.

9 P. Hacker, *Teaching fairness to artificial intelligence: Existing and novel strategies against algorithmic discrimination under EU law*, in *Common Market Law Review*, 2018 (55), p. 1143 ss.; F.J. Zuiderveen Borgesius, *Strengthening legal protection against discrimination by algorithms and artificial intelligence*, in *The International Journal of Human Rights*, 2020, 24 (10), p. 1572 ss.; R. Xenidis, L. Senden, *cit.*; S. Vantin, *Il diritto antidiscriminatorio nell'era digitale*, Milano, Wolters Kluwer, 2021; D. Morondo Taramundi, *Discrimination by machine-based decisions: Inputs and limits of anti-discrimination law*, in E. Fosch-Villaronga, B. Custers (eds.), *Law and artificial intelligence. Regulating AI and applying AI in legal practice*, The Hague, Asser Press-Springer, 2022, p. 73 ss.

10 F.J. Zuiderveen Borgesius, *Discrimination, artificial intelligence, and algorithmic decision-making*, 2018; https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73

11 J. Gerards, R. Xenidis, *Algorithmic discrimination in Europe: Challenges and opportunities for gender equality and non-discrimination law*, 2020, www.equalitylaw.eu/downloads/5361-algorithmic-discrimination-in-europe-pdf-1-975

12 C. Orwat, *Risks of Discrimination through the use of algorithms*, 2020, https://www.researchgate.net/publication/344815625_Risks_of_Discrimination_through_the_Use_of_Algorithms; Commission Nationale de l'Informatique et des Libertés, *Algorithms: preventing automated discrimination*, 2020, https://www.defenseurdesdroits.fr/sites/default/files/atoms/files/836200280_ddd_synthalgos_access.pdf;A. Balayn, S. Gürses, *Beyond debiasing: Regulating Ai and its inequalities. European digital rights*, 2021, https://edri.org/wp-content/uploads/2021/09/EDRi_Beyond-Debiasing-Report_Online.pdf.

sulla protezione dei dati personali non coprono tutti i criteri o i meccanismi di discriminazione, né stabiliscono misure che possano risolvere questi casi. Nell'elaborazione delle norme che regolano e controllano l'uso dell'IA, la discriminazione necessita di un trattamento specifico, che si collega ai meccanismi di garanzia antidiscriminatoria già esistenti, per evitare aree di esenzione rispetto al principio di uguaglianza e la proliferazione di sistemi basati sull'IA che possano ignorare o violare i diritti fondamentali e il principio di uguaglianza.

Nelle pagine che seguono, analizzeremo alcune sfide che la discriminazione algoritmica pone ai meccanismi di protezione esistenti – principalmente il diritto antidiscriminatorio – e come queste sfide possano essere affrontate nell'attuale sviluppo del quadro normativo, in particolare la proposta di Legge sull'Intelligenza Artificiale (Legge sull'IA) della Commissione Europea, attualmente in attesa di dibattito al Parlamento Europeo.

2. La discriminazione algoritmica

Insieme all'espansione dell'uso di decisioni automatizzate o semiautomatizzate, si sta facendo strada anche la convinzione che queste tecnologie possano avere – e abbiano – un impatto negativo sui diritti fondamentali e che sia necessario regolamentarne l'uso. La preoccupazione politica per le sfide sociali poste dall'uso dell'IA si manifesta nella pubblicazione di regolamenti e guide che cercano di incanalare il fenomeno: dal Libro Bianco della Commissione Europea[13] alla recente Proposta di Regolamento sull'IA (Legge sull'IA)[14], o agli annunci circa la creazione di organismi di sorveglianza, come il Comitato europeo per l'IA[15] o l'Agenzia spagnola di supervisione dell'IA (AESIA).

Anche i tribunali europei hanno iniziato a veder comparire cause su questo argomento. Ad esempio, nel febbraio 2020, un tribunale distrettuale dell'Aia[16] ha dichiarato illegale un sistema algoritmico utilizzato dal governo olandese, chiamato SyRI. L'obiettivo dell'algoritmo era di prevenire e contrastare le frodi ai danni della previdenza e della finanza pubblica e il suo funzionamento si basava sull'utilizzo dei *big data* nella predisposizione di profili che, assegnati a ciascun cittadino, determinassero il livello di rischio e la probabilità di frode contro le casse pubbliche. Il tribunale ha stabilito che SyRI non soddisfa i necessari requisiti di proporzionalità e trasparenza, e ha violato il diritto alla vita privata – protetto dall'articolo 8.2 della Convenzione europea dei diritti dell'uomo – di persone, per lo più a basso reddito e di origine immigrata, provenienti da "distretti problematici"[17]. Nel dicembre 2020, un tribunale di Bologna ha anche stabilito che il cosiddetto algoritmo Frank, utilizzato da Deliveroo per classificare (*rank*) i suoi *raiders*, possedeva un carattere discriminatorio. L'algoritmo ordinava la categoria "indice di affidabilità" in base alle assenze e alla non cancellazioni in orario dei propri turni all'interno del sistema di prenotazione di *slots*, senza valutare le cause di fondo delle assenze

13 Commissione Europea, *Libro Bianco sull'intelligenza artificiale. Un approccio europeo all'eccellenza e alla fiducia*, COM(2020) 65 final, 19 febbraio 2020; https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_it.pdf.

14 Commissione Europea, *Proposta di Regolamento del Parlamento Europeo e del Consiglio che Stabilisce Regole Armonizzate sull'Intelligenza Artificiale (Legge sull'Intelligenza Artificiale) e modifica alcuni attivi legislativi nell'Unione* COM (2021) 206 final, 21 aprile 2021; https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0006.02/DOC_1&format=PDF.

15 Legge sull'IA, Articoli 56-58.

16 Tribunale di Distritto dell'Aia, sentenza del 5 febbraio 2020, ECLI:NL:RBDHA:2020:1878.

17 J.A. Eguiluz Castañeira, *Desafíos y retos que plantean las decisiones automatizadas y los perfilados para los derechos fundamentales*, in *Estudios de Deusto*, 2020, 68 (2), p. 325 ss.

dei fattorini. Insomma, per il sistema di IA era equivalente mancare al lavoro per motivi banali, per incuria, per malattia e per l'esercizio del diritto di sciopero. Secondo il tribunale, la mancata considerazione da parte dell'algoritmo delle ragioni di una cancellazione costituisce una discriminazione e penalizza ingiustamente i lavoratori che hanno ragioni legittimi per non lavorare.[18]

Sebbene le sfide alla protezione dei diritti fondamentali poste dall'uso dei nuovi sistemi informatici e di IA siano state studiate soprattutto in relazione alla protezione dei dati personali e della privacy, esse coprono una gamma molto più ampia. L'uso incontrollato dell'IA può mettere a repentaglio la democrazia e i valori su cui si basa lo Stato di diritto, come nel caso delle "camere d'eco" prodotte dagli algoritmi dei social network, che ledono la libertà di informazione e il diritto di voto, o favoriscono l'amplificazione di *fakes* e messaggi di odio che violano la dignità umana e il diritto all'onore. Le nuove generazioni di strumenti algoritmici offrono infatti opportunità senza precedenti di sorveglianza indiscriminata o di massa – il cosiddetto capitalismo di sorveglianza[19] – di *profiling* e classificazione dei cittadini. Inoltre, i sistemi di IA possono pregiudicare l'effettiva tutela giurisdizionale e le garanzie procedurali. A causa della loro opacità, che non consente di conoscere i processi interni con cui gli algoritmi generano i loro risultati, tali sistemi creano incertezza nel determinare l'applicabilità e l'esecuzione delle disposizioni di legge che tutelano i diritti fondamentali, attribuiscono responsabilità e consentono di chiedere un risarcimento.

La discriminazione algoritmica è, quindi, una parte di questo più ampio fenomeno di ripercussione degli sviluppi tecnologici sui diritti fondamentali e su altri principi della democrazia e della giustizia sociale.

Secondo il gruppo indipendente di esperti di alto livello sull'IA[20], la discriminazione algoritmica, la distorsione (*bias*) dell'IA o la distorsione algoritmica consistono in errori sistematici e ripetibili in un sistema informatico, che portano o generano risultati (*output*) discriminatori o iniqui da un punto di vista legale, come ad esempio favorire un gruppo di utenti a scapito degli altri.

Esistono diversi tipi di distorsioni o *bias*[21] che svolgono un ruolo rilevante nel contesto dell'IA:

(i) *Bias* di automazione: un'eccessiva dipendenza dai risultati o dalle soluzioni generate dagli algoritmi, che attenua le capacità umane;

(ii) Distorsioni del passato: l'utilizzo di dati di addestramento (*training sample*) distorti come *input* in un sistema di IA riproduce la distorsione, che influenzerà i risultati futuri (ad esempio, un algoritmo di *screening* di *curricula vitae* che si nutre di dati con un *bias* di genere, apprende che le donne sono meno valide per il lavoro);

(iii) *Bias* di correlazione (*proxy*): la correlazione di diversi insiemi di dati da parte di un algoritmo può portare a pratiche discriminatorie (ad esempio, associare il genere a una minore produttività sul lavoro, non a causa di una relazione causale, ma perché storicamente le donne sono state valutate più negativamente degli uomini a parità di prestazioni lavorative);

(iv) Il *bias* di amplificazione: la capacità degli algoritmi di elaborare enormi quantità di dati ad alta velocità crea un rischio specifico consistente nello "*scaling up*", potendo interessare un numero molto elevato di persone in tempi troppo brevi. Questo rischio di amplificazione implica, quindi, l'esacerbazione e l'accelerazione nella riproduzione di stereotipi e distorsioni discriminatorie che ri-

18 Tribunale Ordinario di Bologna, ordinanza del 31 dicembre 2020, ECLI:NL:RBMNE:2015:178

19 S. Zuboff, *Il capitalismo della sorveglianza. Il futuro dell'umanità nell'era dei nuovi poteri*, Roma, Luiss University Press, 2019.

20 High-Level Expert Group on Artificial Intelligence, *The Assessment List for Trutsworthy Artificial Intelligence* (ALTAI), 2020, disponibile su https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.

21 National Institute of Standards and Technology, *A proposal for identifying and managing bias in Artificial Intelligence,* 2021, disponibile su https://www.nist.gov/artificial-intelligence/proposal-identifying-and-managing-bias-artificial-intelligence-sp-1270.

specchiano i dati di cui si alimenta il sistema di IA, oltre – come abbiamo già segnalato – alle difficoltà per intervenire tempestivamente per mitigare o correggere i risultati distorti nelle decisioni.

Discriminazione e pregiudizi algoritmici possono essere associati a vari fattori.

Il primo di questi fattori è la qualità dei dati. L'uso di dati incompleti, distorti, errati o obsoleti è considerato una fonte primaria di distorsione algoritmica. A volte si sostiene che gli algoritmi in realtà non discriminano: non possono perché sono solo matematica, raccolgono ed elaborano dati. Se il risultato è discriminatorio, guarda i dati: un argomento che è stato espresso graficamente come "spazzatura dentro, spazzatura fuori"[22]. L'impatto differenziato in base a motivi di discriminazione vietati come sesso, razza o origine etnica, disabilità ed età, mostrato da vari studi[23] è perciò attribuito alle disuguaglianze strutturali riflesse nei dati utilizzati dagli algoritmi.

La qualità dei dati può portare a distorsioni in vari punti: i campioni di addestramento (*training samples*) utilizzati per i sistemi di apprendimento automatico possono essere distorti, oppure i dati a cui accede il sistema possono riflettere gerarchie sociali radicate, rappresentazioni imprecise o insufficienti delle condizioni sociali di determinati gruppi, una distribuzione ineguale dei beni, le opportunità e gli oneri, ecc. Potrebbe anche accadere che i progettisti e gli sviluppatori di modelli di IA introducano in essi i propri *bias* e pregiudizi. Possono incorporarli volontariamente o involontariamente durante la preparazione dei campioni di addestramento. Un classico esempio è l'algoritmo dei sistemi di riconoscimento facciale o il sistema di classificazione di genere[24], che utilizzava *dataset* di addestramento distorti con una rappresentazione insufficiente di donne e persone non bianche e, quindi, produceva risultati con un tasso di errore più elevato (del 34,4%) nel caso di donne non bianche rispetto agli uomini bianchi.

Un altro fattore associato alla discriminazione algoritmica si riferisce all'imprevedibilità del risultato (*output*). Se il sistema di IA è troppo grande e complesso, potrebbe non essere compreso dagli esseri umani, ma anche dagli esperti, e portare a una mancanza di trasparenza e responsabilità nel processo decisionale; è il caso, ad esempio, delle reti neurali moderne per dati di testo su larga scala[25]. Gli algoritmi possono avere anche un comportamento parzialmente autonomo, ovvero possono funzionare interpretando determinati dati in ingresso e utilizzando un insieme di istruzioni predeterminate, senza limitarsi ad esse, nonostante il comportamento del sistema sia limitato e orientato a soddisfare l'obiettivo che gli è stato assegnato, nonché a eseguire altre istruzioni di progettazione pertinenti[26]. In questo caso si attenua – anche per i programmatori del sistema – la possibilità di comprendere, controllare o prevedere il funzionamento specifico dell'algoritmo. Ciò complica anche l'individuazione della responsabilità dell'impatto o del risultato dell'applicazione del sistema.

Infine, un fattore legato alla discriminazione algoritmica – perché rende difficile la rilevazione e il trattamento – è l'opacità o l'effetto "scatola nera", che rende complicato determinare dove si trova la radice della discriminazione: sistemi decisionali automatizzati o *software* semiautomatizzati possono contenere *bias* non intenzionali introdotti da loro programmatori, o, se essi sono intenzionali, possono essere nascosti o mascherati in un codice molto complesso. Gli algoritmi possono produrre risultati

22 R. Xenidis, L. Senden, *cit*., p. 157.

23 S. Barocas, D. Selbs, *cit*.; V. Eubanks, *cit*.

24 I. Serna, A. Peña, A. Morales, J. Fierrez, *Inside bias: Measuring bias in deep networks and application to face gender biometrics*, arXiv, 2020, https://arxiv.org/pdf/2004.06592.pdf.

25 M. Altman, A. Wood, E. Vayena. *A harm-reduction framework for algorithmic fairness*, in *IEEE Security & Privacy*, 2018, 16 (3), p. 34 ss.

26 Parlamento Europeo, *Proposta di Risoluzione sulla relazione di valutazione della Commissione concernente l'attuazione del regolamento generale sulla protezione dei dati due anni doppo la sua applicazione* (2020/2717(RSP)), 17 marzo 2021, https://www.europarl.europa.eu/doceo/document/B-9-2021-0211_IT.html.

discriminatori per alcuni gruppi, non perché utilizzino una categoria vietata nei loro codici, ma attraverso *proxies* e correlazioni stabilite nell'ambito dei *big data* e quindi molto difficili, se non impossibili, da tracciare o controllare. L'effetto opacità o "scatola nera" coinvolge sia il funzionamento dell'algoritmo, sia, molto spesso, la qualità dei dati che gestisce, il che rende ulteriormente difficile la definizione del problema della discriminazione. Inoltre, il problema dell'opacità degli algoritmi è dovuto a problemi sia tecnici, sia legali: la complessità del codice e dell'elaborazione per il ragionamento umano ostacola o impedisce la trasparenza e l'accessibilità, ma bisogna anche segnalare il fatto che i modelli e gli algoritmi sono prodotti protetti da copyright e leggi sui segreti commerciali, e che gli sviluppatori e i loro clienti generalmente non sono disposti a renderli apertamente disponibili al pubblico[27].

La caratterizzazione della discriminazione algoritmica come distorsione o errore di sistema che porta a risultati inaccettabili o iniqui pone anche alcuni problemi in relazione alla definizione giuridica e al significato sociale e politico della discriminazione. Va notato che la definizione giuridica di discriminazione non si adatta alla definizione precedentemente riportata di discriminazione algoritmica o distorsione algoritmica. Né la definizione giuridica di discriminazione copre l'intero spettro di situazioni di disuguaglianza che sono socialmente e politicamente considerate discriminatorie. La definizione algoritmica di discriminazione utilizzata nei contesti tecnici si basa sull'idea di *bias* o errore (insito nella progettazione del modello o dovuto alla scarsa qualità dei dati), mentre le nozioni giuridiche e sociali di discriminazione si concentrano sull'idea di uno svantaggio ingiusto. È opportuno tenere presente, quindi, che il problema che tecnologi e giuristi formulano sotto il nome di discriminazione algoritmica potrebbe non essere sempre lo stesso, e quindi le soluzioni che essi ricercano potrebbero essere reciprocamente irrilevanti o divergenti.

3. I limiti della protezione esistente: il diritto antidiscriminatorio

Il diritto antidiscriminatorio europeo si basa sul doppio divieto di discriminazione diretta e di discriminazione indiretta.

Per discriminazione diretta si intende la situazione in cui una persona è, è stata o potrebbe essere trattata meno favorevolmente di un'altra in una situazione comparabile, a causa di una qualsiasi delle caratteristiche vietate.

Come si vede, ciò che definisce la discriminazione diretta è che il trattamento differenziato si basa su una delle ragioni vietate. Nel diritto dell'Unione Europea le ragioni vietate sono sesso, razza o etnia, religione o credo, orientamento sessuale, età e disabilità.

In generale, la categoria della discriminazione diretta è considerata poco applicabile o utile nel caso di discriminazione algoritmica[28]. Non lo è tuttavia, come talvolta si sostiene, perché gli algoritmi

27 Questo problema è chiaramente evidenziato dal caso che mette a confronto la piattaforma civica CIVIO con l'amministrazione (Ministero della Transizione Ecologica e Consiglio della Trasparenza e del Buon Governo, CTBG) in Spagna. Visto il diniego da parte del Ministero del bonus sociale alle persone in possesso dei requisiti per riceverlo, CIVIO ha richiesto informazioni sul sistema BOSCO, che viene utilizzato dalle aziende elettriche per inserire i dati della domanda e comunicare la risposta. Il CTBG non ha fornito il codice del programma, adducendo che esso è protetto da proprietà intellettuale, nonostante sia stato sviluppato dall'amministrazione stessa, impedendo così il controllo degli errori nel programma e l'elaborazione dei reclami. https://civio.es/tu-derecho-a-saber/2019/05/16/la-aplicacion-del-bono-social-del-gobierno-niega-la-ayuda-a-personas-que-tienen-derecho-a-ella/.

28 P. Hacker, *cit.*; R. Xenidis, L. Senden, *cit*.

sarebbero solo matematica e non avrebbero alcun pregiudizio o intenzione di discriminare. In Europa, a differenza del diritto statunitense, l'intenzionalità non svolge alcun ruolo nel determinare la discriminazione, nemmeno la discriminazione diretta. Piuttosto, va detto che i processi decisionali automatizzati basati su algoritmi gestiscono grandi quantità di dati per generare i loro risultati, ed è improbabile che la decisione si basi su uno qualsiasi dei criteri vietati, o che si possa determinare che esso è stato utilizzato (o in che modo). È stato osservato che il rischio è esattamente l'opposto, e cioè che la discriminazione algoritmica diventi più "sottile", più "granulare", altamente intersezionale e superi ampiamente il numero limitato di categorie protette[29]. Ad esempio, il lavoro di Virginia Eubanks mostra come l'uso degli algoritmi dia la possibilità di un accesso differenziato ai servizi alle persone senza fissa dimora che presentano specifiche traiettorie di relazione con i servizi sociali e si trovano in determinate aree urbane[30]. Anche se avessero un effetto differenziato su uomini e donne, o per fasce d'età, o per origine razziale o etnica, queste categorie protette sarebbero disaggregate dall'interazione con gli altri assi, rendendo molto difficile l'approccio al caso attraverso il confronto con le persone in cui una delle caratteristiche protette non concorre.

Gli effetti discriminatori nell'uso degli algoritmi sono spesso dovuti a distorsioni nei dati utilizzati sia nella fase di addestramento dell'algoritmo, sia in quelle successive. I dati con cui viene addestrato l'algoritmo potrebbero non essere sufficientemente rappresentativi di alcuni gruppi (come è accaduto nei sistemi di riconoscimento facciale, che hanno fallito con le donne, le persone non bianche e soprattutto le donne non bianche, perché le immagini con cui il sistema era stato addestrato erano prevalentemente quelle di maschi bianchi). Può anche accadere che i dati a cui accede l'algoritmo riflettano gli stereotipi e le discriminazioni strutturali di cui sono vittima alcuni gruppi della società (ad esempio, i dati che riflettono la segregazione delle donne in diversi settori lavorativi o il divario salariale), e l'algoritmo non faccia altro che elaborare, riprodurre ed esacerbare quei *bias* e pregiudizi.

Per tutti questi motivi, la discriminazione algoritmica sembra più trattabile dal punto di vista della discriminazione indiretta, che si concentra sugli effetti e non sulle azioni.

La discriminazione indiretta è la situazione in cui una disposizione, un criterio o una pratica apparentemente neutri mettono una persona in una posizione di svantaggio rispetto agli altri a causa del loro sesso, razza o origine etnica, religione o convinzioni personali, disabilità, età o orientamento sessuale, a meno che detta disposizione, criterio o prassi non possano essere oggettivamente giustificati in base a uno scopo legittimo, e che i mezzi per conseguire tale scopo siano necessari e adeguati.

Per determinare un caso di discriminazione indiretta non è necessario, quindi, sapere se l'algoritmo ha utilizzato o meno criteri vietati nel processo di formazione della soluzione (*output*) o della decisione offerta dal sistema. In questo modo, possiamo evitare le difficoltà derivanti sia dall'effetto "scatola nera", sia dall'uso di *proxies* e correlazioni, o le distorsioni esistenti nei dati a cui accede l'algoritmo. È sufficiente verificare se tale soluzione/decisione produce effetti differenziati su qualcuno dei gruppi tutelati dalle categorie comprese nella normativa europea o nella dottrina costituzionale.

Sebbene da un punto di vista concettuale la discriminazione indiretta sembri adattarsi al fenomeno della discriminazione algoritmica, l'applicazione pratica della protezione antidiscriminatoria pone una serie di problemi.

Innanzitutto, può essere difficile stabilire lo svantaggio di un gruppo protetto. Sebbene la discriminazione indiretta non richieda il confronto con un gruppo "simile"[31], è comunque necessario identi-

29 R. Xenidis, L. Senden, *cit*., p. 163.

30 V. Eubanks, *cit*.

31 Corte di Giustizia dell'Unione Europea, sentenze del 10 marzo 2005, *Nikoloudi*, ECLI:EU:C:2005:141, e sentenza del 17 luglio 2014, *Leone*, ECLI:EU:C:2014:2090.

ficare un gruppo. Come abbiamo già visto, l'alto livello di intersezionalità o granularità nelle categorie utilizzate dagli algoritmi per produrre le loro soluzioni può rendere difficile l'identificazione del gruppo svantaggiato e la rappresentatività di quel gruppo in termini di categorie protette. In altre parole, può essere difficile percepire un insieme di persone come un "gruppo", se l'algoritmo offre, ad esempio, soluzioni meno vantaggiose a un sottogruppo molto particolare di donne in determinati quartieri o aree di codici postali, o in determinati settori del mercato di lavoro, o con alcune caratteristiche molto specifiche in termini di consumo di beni o uso di servizi; può anche essere difficile che quel sottogruppo specifico sia considerato rappresentativo della categoria "sesso".

Un altro aspetto problematico della discriminazione indiretta in relazione alla discriminazione algoritmica è che, anche se si potesse dimostrare lo svantaggio che crea la decisione automatizzata su un gruppo protetto, non si considererà che vi sia discriminazione se c'è una giustificazione oggettiva. Gran parte della letteratura scientifica ritiene che le opportunità offerte dalla giustificazione oggettiva minano la capacità della discriminazione indiretta di affrontare casi di discriminazione algoritmica. La giustificazione oggettiva richiede che chiunque utilizzi un sistema decisionale basato su algoritmi dimostri che esso persegue uno scopo legittimo e che l'uso è appropriato e proporzionato. Autori come Hacker[32] sostengono che la funzione predittiva degli algoritmi funzionerebbe, di per sé, come uno scopo legittimo (come un meccanismo per misurare la produttività del lavoro o la capacità di affrontare la restituzione del credito da parte di un cliente) e che sembrerebbe appropriato rispetto a tale scopo. Sarebbe difficile per i giudici stabilire che altri algoritmi o soluzioni offerte dai decisori umani presentino meno *bias* o pregiudizi, o che utilizzino correlazioni meno discriminatorie.

Infine, altre difficoltà pratiche derivano dalla mancanza di trasparenza. È difficile, per le vittime di discriminazione indiretta, rendersi conto di esserlo, tanto più se la loro situazione è dovuta all'applicazione di un algoritmo in un universo potenziale di vittime che non hanno continuità o relazione tra di loro. Se, ad esempio, l'algoritmo discrimina donne o membri di minoranze etniche o razziali nell'offerta di lavoro per mezzo di una piattaforma, questi non hanno modo di confrontare tra loro i risultati del sistema. In generale, le vittime della discriminazione indiretta, soprattutto quando si tratta di fenomeni strutturali di discriminazione, come accade con i risultati basati su dati distorti, non sono in grado di apprendere i processi che le mettono in condizione di svantaggio, né hanno la possibilità di accedere a dati e risorse (conoscenza, tempo e denaro) per portare avanti cause giudiziarie per discriminazione.

4. I meccanismi di regolazione della discriminazione algoritmica

Precisamente, i problemi legati all'opacità dei sistemi di IA e le logiche proprietarie delle aziende che li sviluppano sono stati tra i primi ad essere individuati come bisognosi di regolamentazione. Se è persino difficile sapere chi e quando utilizza quali sistemi decisionali basati su algoritmi, non si può procedere a stabilire o ad applicare altri meccanismi di regolazione dei sistemi o protezione per le potenziali vittime di discriminazione.

La Legge sull'Intelligenza Artificiale dell'Unione Europea[33], attualmente in discussione nel Par-

32 P. Hacker, *cit*.

33 La Legge sull'IA è la prima proposta di legge sull'IA da parte di un importante regolatore al mondo e potrebbe diventare uno standard globale, come nel caso del Regolamento Generale sulla Protezione dei Dati (GDPR) dell'UE nel 2018, determinando in che misura l'IA ha un effetto positivo anziché negativo. Per entrare in vigore, è necessario che il Consiglio e il Parlamento europeo concordino una versione comune del testo. Il 1 giugno 2022 è scaduto il termine per la presentazione

lamento Europeo, richiede che la progettazione di sistemi di IA ad alto rischio sia sufficientemente trasparente in modo che gli utenti possano interpretare e utilizzare correttamente i loro *output*. I sistemi di IA devono essere accompagnati da istruzioni per l'uso, con informazioni concise, complete e chiare. Ciò include le caratteristiche, le capacità e i limiti del funzionamento del sistema di intelligenza artificiale, vale a dire le specifiche relative ai dati di *input*, formazione, convalida, ecc., nonché le relative misure di sorveglianza umana.

Tuttavia, nel valutare l'importanza di stabilire requisiti di trasparenza relativi all'uso degli algoritmi nel processo decisionale, e alle loro caratteristiche, occorre tenere conto del fatto che non si tratta solo di fornire informazioni, ma che ciò si concretizza in un diritto alla "spiegabilità"[34], ovvero che queste informazioni siano comprensibili per i destinatari. Si tratta di un diritto controverso[35] e spesso si sostiene che i codici di molti algoritmi non sono comprensibili o rintracciabili nemmeno per gli esperti di informatica.

Sarebbe opportuno che le normative che impongono requisiti di trasparenza stabilissero anche i meccanismi che garantiscono la spiegabilità o che ne mitighino la mancanza, quando non è tecnicamente possibile risalire o comprendere l'algoritmo. Tuttavia, l'approccio attuale sembra più imporre un obbligo di "divulgazione" (*disclosure*) di alcuni dati concreti, in particolare nei casi limitati di sistemi ad alto rischio, per garantire l'effettivo diritto a una spiegazione comprensibile a coloro che potrebbero vedere i propri diritti compromessi dall'uso dei sistemi di IA.

La regolamentazione esplicita dei processi decisionali automatizzati o semiautomatizzati al fine di evitare discriminazioni è una sfida legale importante. La Legge sull'IA assegna, nei Motivi della proposta, un posto rilevante al principio di non discriminazione, arrivando ad affermare che "… la presente proposta integra inoltre il diritto dell'Unione in vigore in materia di non discriminazione con requisiti specifici che mirano a ridurre al minimo il rischio di discriminazione algoritmica (…)"[36].

Tuttavia, va notato che la discriminazione, in quanto tale, non è menzionata nemmeno una volta negli articoli della proposta. La proposta stabilisce un sistema di classificazione per i sistemi di IA basato su diversi livelli di rischio. Un sistema di IA è considerato ad alto rischio in base alla funzione che svolge e allo scopo o alle modalità specifiche per cui il sistema viene utilizzato. La proposta contiene un allegato III che elenca i sistemi di IA i cui rischi si sono già concretizzati e autorizza la Commissione ad ampliare tale elenco. Per i diversi livelli di rischio sono previsti specifici obblighi di prevenzione e di mitigazione. I sistemi ad alto rischio hanno obblighi relativi alla progettazione di sistemi di valutazione del rischio, alla qualità dei dati, alla documentazione tecnica, alla registrazione e tracciabilità del sistema, alle garanzie di supervisione umana, alla sicurezza informatica e alla robustezza del sistema.

Numerose organizzazioni hanno criticato questo approccio: 61 di loro hanno scritto una lettera aperta alla Commissione chiedendo che siano stabilite delle "linee rosse" per l'uso dei sistemi di IA, vietando quegli usi che riproducono discriminazioni strutturali o mettono in discussione i diritti fondamentali[37]. L'approccio basato sul rischio è stato anche criticato per la sua mancanza di rigore[38] e

degli emendamenti da parte dei gruppi politici del Parlamento europeo. Gli emendamenti, migliaia in totale, preludono a complessi negoziati in quanto gli eurodeputati nelle due commissioni principali, la commissione per il mercato interno (IMCO) e la commissione per le libertà civili (LIBE), sono quasi equamente divisi attorno all'asse di centrodestra e di centrosinistra, e sui punti più controversi.

34 D. Selbst, J. Powles, *Meaningful information and the right to explanation*, in *International Data Privacy Law*, 2017, 7 (4), p. 233 ss.

35 S. Wachter, B. Mittelstadt, L. Floridi, *Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation*, in *International Data Privacy Law*, 2017, 7 (2), p. 76 ss.

36 Legge sull'IA, p. 4.

37 *Open Letter: Civil society call for the introduction of red lines in the upcoming European Commission proposal on Artificial Intelli-*

perché fondamentalmente non è finalizzato tanto alla gestione del rischio, quanto a stabilire un equilibrio, una proporzionalità tra i rischi posti dai sistemi di IA e i diritti umani o i valori democratici[39].

L'approccio dei livelli di rischio non intende affrontare in modo specifico il rischio di discriminazione. L'elenco dei sistemi ad alto rischio nell'allegato III contiene aree e funzioni decisionali in cui la discriminazione è comune (accesso all'istruzione o al lavoro, gestione della migrazione, accesso a beni e servizi, ecc.). Tuttavia, gli obblighi contenuti nella proposta, per i sistemi di IA che trovano applicazione in tali ambiti e con funzioni di classificazione, selezione e persino di "predizione" (che, non lo possiamo dimenticare, significa calcolo di probabilità), non tengono conto delle specificità della discriminazione algoritmica, come ad esempio l'uso di *proxies*, la granularità delle categorie, i problemi legati all'individuazione dello svantaggio o il problema della giustificazione oggettiva. Quest'ultimo è particolarmente importante, vista l'importanza che la proposta attribuisce all'idea di proporzionalità.

Anche la proposta di Regolamento sui Servizi Digitali (Legge sui Servizi Digitali)[40] riconosce che determinati gruppi o individui possono trovarsi in una situazione di vulnerabilità o svantaggio nell'uso dei servizi online a causa del loro genere, della razza o dell'origine etnica, della religione o delle convinzioni, della disabilità, dell'età o dell'orientamento sessuale. Costoro potrebbero essere lesi "da distorsioni (consapevoli o inconsapevoli) potenzialmente introdotte nei sistema di notifica da utenti e da terzi e replicati negli strumenti automatizzati di moderazione dei contenuti usati dalle piattaforme"[41]. La proposta mira a mitigare questo rischio di discriminazione, compresa la presentazione discriminatoria di pubblicità che incida sulla parità di trattamento e di opportunità. La proposta comprende anche la disciplina dei sistemi di ricorso interno e di risoluzione extragiudiziale delle controversie per i destinatari dei servizi. Sono inoltre previsti obblighi di audit e un meccanismo di vigilanza sovranazionale, con particolare attenzione alle grandi piattaforme online.

Come si vede, le proposte regolatorie europee – pur sottolineando il rischio di discriminazione e di violazione dei diritti fondamentali – non prevedono nei loro articoli meccanismi di diritto antidiscriminatorio (non contengono nemmeno un divieto di discriminazione indiretta o un divieto di istruzioni discriminatorie). Piuttosto, esse stabiliscono meccanismi che aumentano la trasparenza, la tracciabilità e l'accesso alle informazioni dei sistemi di IA, nonché alcuni obblighi relativi alla valutazione del rischio e alla definizione di piani di mitigazione.

Queste proposte di intervento normativo sulla discriminazione algoritmica che troviamo nei documenti dell'UE sono soluzioni prevalentemente tecnologiche. La concezione della discriminazione in termini di "distorsioni" (*bias*) derivanti dalla scarsa qualità dei dati ha portato a questo approccio tecnologico: la discriminazione sarebbe intesa non tanto come riflesso di un problema strutturale di disuguaglianza sociale, quanto come un problema tecnico nella progettazione del modelli, nei dati di addestramento, nella raccolta e analisi dei dati, ecc.; tutti problemi che possono e devono essere risolti attraverso soluzioni tecniche "anti-bias" (*debiasing*).

---

gence, 12 gennaio 2021, https://edri.org/our-work/civil-society-call-for-ai-red-lines-in-the-european-unions-artificial-intelligence-proposal/.

38 T. Mahler, *Between risk management and proportionality: The risk-based approach in the EU's Artificial Intelligence Act Proposal*, in L. Colonna, S. Greenstein (eds.), *Nordic yearbook of law and informatics 2020-2021. Law in the era of artificial intelligence*, Stockholm, IRI, 2022, pp. 247-269.

39 Access Now, *The EU should regulate AI on the basis of rights, not risks*, 17 febbraio 2021, https://www.accessnow.org/eu-regulation-ai-risk-based-approach/.

40 Commissione Europea, *Proposta di Regolamento del Parlamento Europeo e del Consiglio relativo a un mercato único dei servizi digitali (Legge sui servizi digitali) e che modifica la Direttiva 2000/31/CE* COM(2020) 825 final, 15 dicembre 2020; https://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=CELEX:52020PC0825&from=en.

41 Si veda la legge sui servizi digitali, p. 13.

Il meccanismo anti-bias si riferisce a metodi che affrontano i bias introducendo una forma di parità statistica (chiamata "metrica di equità") nel set di dati, nell'algoritmo o nei risultati. Con questo stesso approccio funzionano anche gli audit, che fanno riferimento a processi di valutazione della parità statistica del sistema esaminato.

Alcuni lavori indicano già i limiti di questo approccio tecnocentrico alla discriminazione nel contesto dell'IA, nonché delle soluzioni anti-bias[42]. Delle soluzioni tecnologiche, si critica che le indicazioni su cosa siano le distorsioni e i loro problemi sono molto vaghe sia nel Libro bianco, sia nella Legge sull'IA, e che non possono essere utilizzati come criteri guida: i meccanismi anti-bias non sono una panacea universale per qualsiasi tipo di applicazione di IA e hanno – ad oggi – un'efficacia (comprovata) molto limitata, che la documentazione europea sembra ignorare.

Inoltre, l'approccio tecnocentrico alla discriminazione come *data bias* non può spiegare la discriminazione strutturale (che può avere una perfetta corrispondenza statistica e tuttavia essere ingiusta). I meccanismi anti-bias non sono progettati per affrontare il problema della discriminazione in senso sociale. Infine, è stato anche notato che le soluzioni tecnologiche anti-bias concentrano ancora più potere nelle aziende tecnologiche che sono, in definitiva, attori commerciali il cui interesse principale è il proprio vantaggio e non le importanti questioni politiche e sociali che stanno alla base delle decisioni in materia di uguaglianza e discriminazione.

Sebbene la riduzione della discriminazione a distorsione possa causare una valutazione errata dei problemi di discriminazione, e sebbene i meccanismi anti-bias non siano la panacea contro la discriminazione algoritmica, va notato che l'IA può anche contribuire alla supervisione e monitoraggio dell'uso dei sistemi di IA, o alla creazione di prove su discriminazione e disuguaglianza più in generale, rivelando i nostri pregiudizi[43]. Ad esempio, il sistema Claudette utilizza il *machine learning* per identificare clausole abusive nei termini di servizio o nelle informative sulla *privacy* delle piattaforme online[44]. Questo sistema, sviluppato attraverso una collaborazione tra il mondo accademico e le organizzazioni dei consumatori, mostra anche il vantaggio di espandere l'uso della tecnologia ad attori diversi dalle grandi aziende tecnologiche (Big Tech) o dai governi, consentendo di mitigare uno dei rischi già menzionati nell'approccio tecno-centrico. Per questo, come previsto dall'Agenzia per i diritti fondamentali o dal Parlamento europeo, i *big data* e l'IA potrebbero rappresentare anche opportunità e strumenti nella lotta alle discriminazioni e nella tutela dei diritti fondamentali[45].

42 A. Balayn, S. Gürses, *Beyond debiasing: Regulating Ai and its inequalities*, European Digital Rights Report, 2021, https://edri.org/wp-content/uploads/2021/09/EDRi_Beyond-Debiasing-Report_Online.pdf.

43 J.L. Kleinberg, S. Mullainathan, C. Sunstein, *Algorithms as discrimination detectors*, in *Proceedings of the National Academy of Science*, 2020, 117 (48), p. 30096 ss.

44 F. Lagioia, G. Sartor, *Artificial intelligence in the big data era: risks and opportunities*, in J. Cannataci, V. Falce, O. Pollicino (eds.), *Legal challenges of big data*, 2020, Edward Elgar Publishing, Cheltenham Glos, p. 280 ss.

45 European Union Agency for Fundamental Rights, Council of Europe, *#BigData: Discrimination in data-supported decision making*, 2018, https://fra.europa.eu/en/publication/2018/bigdata-discrimination-data-supported-decision-making; Parlamento Europeo, *Rizoluzione recante raccomandazioni alla Commissione concernenti il quadro relativo agli aspetti etici dell'intelligenza artificiale, della robotica e delle tecnologie correlate* (2020/2012(INL)), 20 ottobre 2020, https://www.europarl.europa.eu/doceo/document/TA-9-2020-0275_ES.html, punto 27 e ss.

5. Conclusioni

Tra le sfide etiche poste dall'uso dell'IA, la discriminazione algoritmica sta cominciando ad attirare l'attenzione per la sua capacità di influenzare i valori e i diritti fondamentali delle nostre democrazie, la coesione sociale e gli sforzi politici contro la disuguaglianza.

La discriminazione algoritmica presenta profili distintivi che richiedono soluzioni specifiche. L'utilizzo della normativa in materia di protezione dei dati personali appare chiaramente insufficiente per far fronte ai problemi posti dalla discriminazione algoritmica.

Un problema fondamentale, di fronte alla progettazione di soluzioni antidiscriminatorie nel contesto dell'IA, risiede nella divergenza tra le diverse concezioni di discriminazione algoritmica sostenute da tecnologi, giuristi e decisori di politica pubblica. La necessaria collaborazione interdisciplinare per generare soluzioni a un problema complesso come la discriminazione nel contesto dell'IA richiede che questa divergenza di concezioni sia riconosciuta come tale e trattata. Solo in questo modo si può evitare che le diverse concezioni di discriminazione siano usate in modo intercambiabile, generando confusione e soluzioni inadeguate a problemi mal posti.

Ciò diventa particolarmente urgente quando si cominciano a produrre normative per regolamentare l'uso e l'applicazione dell'IA, come la proposta di legge europea, che nei suoi Motivi pone il problema della discriminazione in senso giuridico-politico, ma successivamente limita i rimedi proposti ai problemi della discriminazione algoritmica intesa in senso tecnologico.

È necessario raggiungere un livello più elevato di integrazione interdisciplinare nella progettazione di soluzioni, sia tecnologiche, sia legali sia di politiche pubbliche.

Il quadro dell'antidiscriminazione in Europa presenta, in generale, evidenti debolezze per affrontare un fenomeno come la discriminazione algoritmica. Questa può essere l'occasione per rafforzare tale quadro normativo contro i fenomeni strutturali di discriminazione e disuguaglianza attraverso lo sviluppo di una vera politica pubblica antidiscriminatoria.

È necessario che le normative che regolano l'IA includano esplicitamente il divieto di discriminazione diretta, indiretta e intersezionale, e che l'uso di sistemi di IA, di sistemi decisionali automatizzati o semiautomatizzati o altre applicazioni, non diventi una forma di esenzione dal divieto di discriminazione.

Alcune caratteristiche del diritto antidiscriminatorio sono troppo rigide per affrontare la discriminazione algoritmica. È necessario rivedere sia i meccanismi per determinare lo svantaggio, sia le formule per risolvere il problema dell'operabilità dell'intersezionalità.

Al di là degli sviluppi che ciò potrebbe comportare per la normativa antidiscriminatoria in generale, applicabili ad altre forme di discriminazione strutturale come i divari (i *gender gaps*) o la segregazione, è necessario sviluppare una politica pubblica antidiscriminatoria che possa garantire un'azione di tutela a cui, al giorno d'oggi, le vittime non possono accedere individualmente, e che i tribunali con scarsa formazione e supporto non possono offrire.

Questo sviluppo si basa su una premessa di trasparenza. Sebbene non tutto negli algoritmi sia comprensibile alla mente umana, è necessario sapere quando e per fare cosa i sistemi di IA vengono utilizzati e qual è il loro impatto. È inoltre necessario che vengano assunte responsabilità per il loro utilizzo: che i modelli e i codici siano soggetti a controllo, che esistano meccanismi di certificazione, monitoraggio e valutazione d'impatto. E infine, è necessario demistificare l'IA e assumere che è uno strumento, e che tale strumento non è inevitabile: quando nessuna misura di prevenzione o mitigazione può garantire il diritto a non subire discriminazioni o a prevenire la violazione dei diritti fondamentali, anche l'uso dell'IA può essere limitato, condizionato o vietato.