"*Do Algorithms dream about Electric Sheep?*" Percorsi di studio in tema di discriminazione e processi decisori algoritmici tra le due sponde dell'Atlantico*

Giacomo Capuzzo

Abstract

Questo studio si propone di descrivere il panorama giuridico che relaziona la discriminazione al mondo dell'intelligenza artificiale, con particolare riferimento all'impiego degli algoritmi. L'autore analizza il funzionamento di queste macchine automatiche nell'ambito dei processi decisori di attori privati e pubblici sottolineando i potenziali effetti discriminatori derivanti da tale impiego. Il testo approfondisce i vari aspetti della tutela discriminatoria tra gli ordinamenti multilivello europeo e statunitense approfondendo le discipline normative e la trattazione di alcuni casi pratici per fornire una mappa introduttiva alla tematica.

This study aims at describing the legal framework that relates discrimination to the world of artificial intelligence, with particular reference to the use of algorithms. The author analyzes the operation of these automatic machines within the decision-making processes of private and public actors, emphasizing the potential discriminatory effects resulting from such use. The paper explores the various aspects of discriminatory protection between the European and US multilevel legal systems, deepening the regulatory disciplines and dealing with some practical cases to provide an introductory map to the issue.

Sommario

* Su determinazione della direzione, in conformità all'art. 15 del regolamento della Rivista, l'articolo èstato sottoposto a referaggio anonimo

sul trattamento dei dati personali – 2.5 La tutela antidiscriminatoria negli Stati Uniti d'America nel campo dei processi decisori algoritmici – 3. La discriminazione algoritmica in pratica: una serie di casi sostanziali – 4. A mo' di conclusione.

Keywords

intelligenza artificiale - algoritmi - tutela antidiscriminatoria - privacy - controllo sociale

1. Alcuni cenni introduttivi

Il crescente impiego dell'intelligenza artificiale all'interno dei settori produttivi e dell'amministrazione pubblica ha contraddistinto i primi decenni del nuovo secolo. Lo sviluppo tecnologico ha consentito l'utilizzo sempre maggiore di elaboratori elettronici in grado di eseguire compiti di supporto e talvolta in sostituzione dell'attività umana[1].

Un passaggio fondamentale ha riguardato la possibilità di raccogliere e processare masse di dati mediante i quali elaborare informazioni capaci di consentire un funzionamento intelligente da parte delle macchine stesse. Attraverso questo percorso si può concepire un processo decisorio automatico affidato a degli elaboratori che, sulla base dei dati inseriti, consente l'individuazione di una specifica soluzione ad un determinato problema. Queste nuove applicazioni hanno favorito l'introduzione dell'intelligenza artificiale in molti ambiti delle attività produttive e della pubblica amministrazione. In particolare, è aumentato l'impiego di algoritmi, sequenze di istruzioni informatiche ben definite che sono impiegate per risolvere una serie di problemi o per eseguire un determinato calcolo[2].

L'algoritmo è quindi una delle diverse applicazioni dell'intelligenza artificiale e comprende un'ampia gamma di strumenti, che a loro volta influenzano una varietà di operazioni. Si tratta di un tipo di decisione che viene presa miliardi di volte all'anno in

[1] In tema di nuove tecnologie in generale ed intelligenza artificiale, a carattere introduttivo, si veda S. Rodotà, *Elaboratori elettronici e controllo sociale*, Bologna, 1973; F. Pasquale, *The black Box Society: The Secret algorithms that Control Money And Information*, Cambridge, 2015, 5 ss.; S. Barocas - A. D. Selbst*, Big Data's Disparate Impact*, in 104 *Calif. L. Rev.* 671, 674 N.10, 2016; I. Ajunwa, *Algorithms At Work: Productivity Monitoring Platforms And Wearable Technology As The New Data-Centric Research Agenda For Employment And Labor Law*, in 63 *St. Louis U. L.J.* 2019; I. Ajunwa, *Genetic Testing Meets Big Data: Tort And Contract Law issues*, in 75 *Ohio St. L.J.* 1225, 2014; D. K. Citron - F. Pasquale, *The Scored Society: Due Process For Automated Predictions,* in 89 *Wash. L. Rev.* 1, 2014; K. Crawford - J. Schultz, *Big Data And Due Process: Toward A Framework To Redress Predictive Privacy Harms*, in 55 *B.C. L. Rev.* 93, 2014; L. Edwards-M. Veale, *Slave To The Algorithm? Why A 'Right to an explanation' Is Probably Not The Remedy You Are Looking For,* in 16 *Duke L. & Tech. Rev.* 18, 2017; G. Resta-V. Zeno Zencovich (a cura di), *La protezione transnazionale dei dati personali,* Roma, 2016; G. Resta, *Diritti esclusivi e nuovi beni immateriali*, Torino, 2011; G. Pitruzzella, *Big Data, Competition and Privacy: A Look from the Antitrust Perspective*, in *Concorrenza e Mercato*, 2016, 15 ss.; F. Pizzetti, *Privacy e diritto europeo nella protezione dei dati personali*, Torino, 2016; G. Pascuzzi*, Il diritto nell'era digitale. Tecnologie informatiche e regole privatistiche*, Bologna, 2002. A. D. Selbst-S. Barocas, *The Intuitive Appeal Of Explainable Machines*, in 87 *Fordham L. Rev.* 2018.

[2] S. Barocas-S. Hood-M. Ziewitz, *Governing Algorithms: A Provocation Piece, Governing Algorithms*, 29 Mar. 29, 2013.

svariati settori come quello creditizio, concorsuale o diagnostico in ambito medico[3]. Questa tipologia di intelligenze artificiali è sviluppata utilizzando metodi di apprendimento automatico (*machine learning*). Si tratta di tipologie di algoritmi che costruiscono funzioni di previsione sulla base di una serie di dati allo scopo di realizzare tali predizioni; una volta consolidata tale funzione, l'algoritmo riceve un particolare "input" (come le caratteristiche di un candidato ad una particolare posizione lavorativa) e predice alcuni risultati (come la sua performance in specifici ambiti)[4].

L'utilità di queste macchine è tale che sono molteplici ormai gli ambiti in cui viene sfruttata e le sue implicazioni sociali sono del tutto evidenti. Molte delle decisioni che sono prese all'interno della società si basano o sono influenzate dagli esiti di queste elaborazioni algoritmiche. Per tali ragioni, i giuristi hanno cominciato ad interessarsi alle ripercussioni nel mondo giuridico di queste nuove tecnologie[5].

Il loro impiego è rilevante per il diritto[6], sia perché le regole dell'ordinamento si estendono al funzionamento dell'algoritmo quando è impiegato in relazione a particolari attività umane, sia perché il diritto stesso ha cominciato ad impiegare algoritmi nell'ambito della sua applicazione, come ad esempio nei contratti di appalto o nel calcolo dell'assegno di mantenimento. In particolare i principali ambiti di interesse hanno riguardato finora la tutela della privacy e il trattamento dei dati personali e in ambito di protezione antidiscriminatoria. Quest'ultimo aspetto ha attirato di recente le attenzioni degli studiosi perché l'impiego degli algoritmi nel contesto di diversi processi decisionali, sia nel comparto privato, che in quello pubblico, li ha spinti a chiedersi se non vi potessero essere possibili ambiti di discriminazione connessi con questa tipologia di intelligenze artificiali[7].

Tali rilievi si fondano su due serie di considerazioni che hanno entrambe origine nei

---

[3] A. Mantelero, *I Big Data nel quadro della disciplina europea della tutela dei dati personali*, in *Il Corriere giuridico - Speciali Digitali 2018*, 2018, 46 ss.; V. Morabito, *Big Data and Analytics. Strategic and Organizational Impacts,* New York, 2015, 23 ss; P. T. Kim, *Auditing Algorithms for Discrimination*, in 166 *U. Pa. L. Rev. Online* 189, 2017; Id., *Data-Driven Discrimination At Work,* in 58 *Wm. & Mary L. Rev.* 857, 2017; P. Kim-S. Scott, *Discrimination In Online Employment Recruiting*, in 63 *St. Louis U. L.J.*, 2019; C. A. Sullivan, *Employing*, 2018 (Seton Hall Public Law Research Paper); J. A. Kroll et al., *Accountable Algorithms*, in 165 *U. Pa. L. Rev.* 633, 2017.

[4] C. Angelopoulos, et al., *Study of fundamental rights limitations for online enforcement through self regulation*, report IViR Institute for Information Law, University of Amsterdam, 2016; J. Angwin et al., *Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks*, in *ProPublica*, 23 maggio 2016 J. R. Bambauer-T. Zarsky, *The algorithm game*, in *Notre Dame Law Review,* 2018.

[5] R. Avraham, *Discrimination and Insurance*, in K. Lippert-Rasmussen (a cura di), *The Routledge handbook of the ethics of discrimination*, New York, 2017; K. Charles-J. Guryan, *Prejudice and wages: An empirical assessment of Becker's the Economics of Discrimination*, in *J. Polit. Econ.,* 116, 2008, 773 ; M. Turner et al., *All other things being equal: A paired testing study of mortgage lending institutions–final report*, Tech. Rep., US Department of Housing and Urban Development Office of Policy Development and Research, Washington, 2002. M. Bertrand-S. Mullainathan, *Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination, in Am. Econ. Rev.* 94, 2004, 991 ; V. Zeno-Zencovich, *Dati, grandi dati, dati granulari e la nuova epistemologia del giurista*, in *questa Rivista,* 2, 2018, 5-6.

[6] Su questo punto il rimando è a S. Rodotà, *Elaboratori elettronici e controllo sociale*, cit.; Id. *Tecnologie e diritti*, Bologna, 1995.

[7] S. Barocas-A. Selbst, *Big data's disparate impact*, in *Calif. Law Rev.* 104, 2016, 671 ss. S. Bornstein, *Antidiscriminatory algorithms*, in *Ala. Law Rev*. 70, 2018, 519 ss.; J. Kleinberg-J. Ludwig-S. Mullainathan-C. Sunstein, *Discrimination in the age of algorithms*, in *J. Legal Anal.* 10, 2018, 113 ss.; L. Sweeney, *Discrimination in online ad delivery*, in *Queue*, 2013, 10 ss..

discorsi e nelle elaborazioni che hanno accompagnato l'impiego di questi strumenti, secondo la prima, sebbene gli algoritmi consentano di ridurre la discrezionalità legata ai processi decisori gestiti dagli esseri umani, si ritiene altamente probabile che gli stessi riproducano le disuguaglianze già esistenti. Rispetto alla seconda linea argomentativa, l'idea che i calcoli algoritmici siano percepiti come neutrali, nella convinzione che la tecnologia sia sempre un aiuto, semplifichi le cose e che quindi il funzionamento di queste macchine ad apprendimento automatico non possa essere ricostruito o alterato dall'esterno (*black box*), rendendo di fatto il settore immune alla regolamentazione di natura giuridica[8].

2. L'antidiscriminazione alla prova dei processi decisori algoritmici: l'approccio *ex post*

Se le questioni principali relative all'utilizzo degli algoritmi risultano condivise dalla dottrina prevalente negli ordinamenti occidentali, per quanto concerne le soluzioni, le opinioni sono piuttosto diverse e non solo per la presenza di normative e tutele differenti tra le due sponde dell'Atlantico in tema di discriminazioni. In particolare, sono stati evidenziati due diversi approcci alla protezione antidiscriminatoria nell'ambito dei processi decisori algoritmici. Il primo preferisce una tutela *ex post* attraverso l'estensione dell'attuale regolamentazione in materia antidiscriminatoria anche ai casi relativi all'impiego degli algoritmi[9]. Tale approccio punta ad informare la creazione dell'algoritmo ai principi di trasparenza e di responsabilità in modo da prevenire la possibilità di produrre un esito discriminatorio[10]. In questo modo è possibile operare una verifica di un particolare algoritmo, attraverso l'accesso alle informazioni che lo riguardano e agli schemi della sua elaborazione, identificare difetti, errori intenzionali e forse scovare risultati indesiderati e possibilmente non intenzionali come la discriminazione. Questo approccio consente quindi uno stretto controllo di tutte le componenti dell'algoritmo allo scopo di escludere ogni possibile risultanza che possa condurre a discriminazioni o ad altri esiti non in linea con le normative vigenti nel

[8] Su questo punto si veda F. Pasquale, *The Black Box Society: The Secret Algorithms that control Money and Information,* cit., 34-35; I. Bogost, *The Cathedral of Computation,* cit; K. Fink , *Opening the government's black boxes: freedom of information and algorithmic accountability*, in *Information, Communication & Society*, 21(10), 2018, 1453.

[9] Sui rischi e le problematiche sollevate dall'impiego massivo dell'intelligenza artificiale, A. G. Ferguson, *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*, New York, 2017; A. Danna-O.H. Gandy Jr., *All that glitters is not gold: Digging beneath the surface of data mining*, in *J Bus Ethics*, 40(4), 2002, 373; J. Burrell, *How the machine 'thinks': understanding opacity in machine learning algorithms*, in *Big Data & Society* 3(1), 2016,1 ss.; D. M. Boyd-K. Crawford, *Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon*, in *Information, Communication & Society*, 15(5), 2012, 662.

[10] T. Khaitan, *A theory of discrimination law*, Oxford, 2015; P. Hacker, *Teaching fairness to artificial intelligence: Existing and novel strategies against algorithmic discrimination under EU law Common Market Law Review*, 4, 55, 2018, 1143 ss.; M. Hardt, *How big data is unfair. Understanding sources of unfairness in data driven decision making*, 2014; R. Gellert-K.De Vries-P. De Hert-S. Gutwirth, *A comparative analysis of anti-discrimination and data protection legislations,* in *Discrimination and privacy in the information society,* Berlin, Heidelberg, 2013, 61-89.

campo di applicazione dell'algoritmo[11]. Una simile condizione difficilmente può verificarsi con rispetto agli algoritmi, se si fa riferimento, ad esempio, a quei casi che hanno destato maggiori perplessità nel campo della tutela antidiscriminatoria.
La principale problematica dell'approccio *ex post* per regolare l'automazione risiede nel fatto che anche con tutte le informazioni relative ad uno specifico algoritmo, elaborare una mappa di tutte le possibili variabili intervenute all'interno del processo decisorio è molto spesso alquanto complicato[12]. Al contrario una protezione *ex ante* che tiene conto dell'elaborazione dell'algoritmo attraverso le procedure tecniche impiegate per realizzarlo e predispone una regolamentazione giuridica puntuale rispetto alle casistiche previste dal funzionamento dei processi decisori algoritmici[13]. proibire quei risultati che possono comportare discriminazioni, vietare gli usi inappropriati e fino a richiedere che il software sia costruito secondo determinate specifiche che possono essere testate o controllate[14].

2.1 *segue* L'approccio *ex ante*

Sulla base di quanto appena esposto, è necessario quindi soffermarsi sugli interrogativi connessi con un approccio *ex ante* alla protezione antidiscriminatoria rispetto all'impiego degli algoritmi. I contorni di una tutela *ex post* come quella fondata sull'applicazione dei principi neoliberali di trasparenza e responsabilità sono stati delineati nel precedente paragrafo, svelando una concezione che, sebbene consenta una tutela generale per tutti i casi nei quali emerga una decisione al termine di un processo decisorio algoritmico che possa essere considerata discriminatoria, rimane su di un piano meramente formale per la difficoltà di mappare il funzionamento dell'algoritmo e di reperire prove effettive che possano attribuire il risultato in questione ad uno o più responsabili[15].
L'approccio *ex ante* si basa invece su un'opera di classificazione e analisi di questi sistemi automatici per comprendere la tipologia di danno che possono comportare, le soluzioni che possono essere prodotte e gli impieghi consentiti di questi software. Per

[11] Il passaggio si può ritrovare approfondito in P. T. Kim, *Data-Driven Discrimination at Work*, cit., 857; D. J. Weitzner et al., *Information Accountability*, cit., 86.

[12] Per l'argomento in generale, si veda, M. Kaminski, *Binary governance: A two-part approach to accountable algorithms* (2018), in 92 *S. Calif. L. Rev.* 2019. Per gli esempi, A. Datta et al., *Discrimination in online advertising: A multidisciplinary inquiry* (Conference on Fairness, Accountability and Transparency 2018) 20; A. Datta-M.C. Tschantz, *Automated experiments on ad privacy settings*, in 1 *Proceedings on Privacy Enhancing Technologies*, 2015, 92.

[13] Su di un approccio all'antidiscriminazione che si fondi su alcuni principi generali si veda J.H. Gerards, *Discrimination grounds*, in M. Bell-D. Schiek (a cura di), *Ius commune case books for a common law of Europe – Non-discrimination*, Oxford, 2007, 33 ss.. Federal Trade Commission, *Big data: A tool for inclusion or exclusion? Understanding the issues* (Gennaio 2016); C. Dwork et al., *Fairness through awareness*, in *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference ACM*, 2012, 214.

[14] Su questi punti programmatici si veda European Group on Ethics in Science and New Technologies, *Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems*, Marzo 2018.

[15] Sul punto si richiama la trattazione di D. R. Desai-JA. Kroll, *Trust but Verify: A guide to Algorithms and the Law*, cit., 11 ss.

fare ciò è però necessario comprendere lo sviluppo di tali processi decisori, le componenti degli stessi, in quale modo questi sistemi algoritmici riescano a discriminare e come questo possa essere rilevato su di un piano strettamente giuridico[16].

In questo senso, si può iniziare ad approfondire questi temi specificando come, nonostante le persone spesso si riferiscano impropriamente a qualsiasi processo che rielabora dati e produce una previsione come ad un "algoritmo", è importante notare che ci sono effettivamente due processi algoritmici separati al lavoro nelle applicazioni di *screening* del tipo che stiamo considerando: 1) L'algoritmo di screening (o *screener*) prende semplicemente le caratteristiche di un individuo (come un candidato di lavoro) e restituisce una previsione del risultato di questo individuo. Questa previsione quindi informa una decisione. 2) L'algoritmo di formazione (o *trainer*) è ciò che produce l'algoritmo di screening[17].

La costruzione di questo secondo algoritmo implica (tra le altre cose) l'assemblaggio di istanze passate da utilizzare come dati di addestramento, la definizione del risultato da prevedere e la scelta di predittori candidati da considerare. L'algoritmo di screening è solo il risultato meccanico dell'applicazione dell'algoritmo di addestramento su un insieme di dati di addestramento. Quindi, mentre il primo può produrre decisioni distorte, il momento in cui si genera il trattamento discriminatorio è spesso la fase di addestramento che coinvolge il secondo[18].

2.2 Il potenziale discriminatorio degli algoritmi: una guida pratica

ll processo decisionale guidato dall'intelligenza artificiale può portare alla discriminazione in diversi modi. In un articolo fondamentale, Barocas e Selbst distinguono cinque modi in cui il processo decisionale algoritmico può portare alla discriminazione. I problemi riguardano (I) come vengono definite la *target variable* e le *class labels*; (II) l'etichettatura dei dati di addestramento; (III) la raccolta dei dati di addestramento; (IV) la selezione degli indicatori; (V) i *proxies* ed infine (VI) l'impiego degli algoritmi per fini discriminatori in modo volontario[19].

[16] S. Wachter, *Normative challenges of identification in the Internet of Things: Privacy, profiling, discrimination, and the GDPR*, in 34 *Computer Law & Security Review*, 3, 2018, 436 ss.; A.D.Selbst-S. Barocas, *The intuitive appeal of explainable machines*, in *Fordham Law Review*, 86, 2018; R. Swedloff , *Risk classification's Big Data (r) evolution*, in 21 *Connecticut Insurance Law Journal*, 2014, 339; A. Moretti, *Algoritmi e diritti fondamentali della persona. Il contributo del regolamento (UE) 2016/679* in *Dir. Inf.,* 4-5, 2018, 799 ss..

[17] Il funzionamento dell'algoritmo come procedura in due fasi è ben spiegata in: J. Kleinberg-J. Ludwig-S. Mullainathan-C. Sunstein, *Algorithms as discrimination detectors*, in *Proceedings of the National Academy of Sciences of the United States of America*, 28 luglio 2020.

[18] *Ibid.*

[19] La trattazione delle diverse ipotesi di distorsione discriminatoria del percorso decisorio algoritmico sono delineate in S. Barocas-A.D. Selbst, *Big Data's disparate impact*' cit., 671. Le stesse ipotesi sono riprese in F. Zuiderveen Borgesius, *Discrimination, artificial intelligence, and algorithmic decision-making*, (Consiglio d'Europa) Strasburgo, 2018, 10 ss.

I) la definizione di target variable e class labels.

Come si è visto in precedenza l'algoritmo di screening lavora sulla base dell'algoritmo di addestramento, entrambi cercano correlazioni tra gruppi di dati, il secondo per fornire dati al primo, il primo per elaborare la soluzione richiesta. Ad esempio, quando un'azienda sviluppa un filtro antispam, l'algoritmo alla base viene allenato attraverso l'inserimento di una serie di messaggi di posta elettronica che sono etichettati dai programmatori come "spam" e "non spam". I messaggi etichettati sono i dati di addestramento[20].

L'algoritmo rileva quali caratteristiche dei messaggi sono correlate all'essere etichettati come spam, l'insieme di queste correlazioni individuate è spesso chiamato "modello predittivo". L'algoritmo viene addestrato a studiare i dati inseriti per capire quali caratteristiche possono essere prese in considerazione per ottenere i risultati richiesti, che vengono definiti come *target variable*[21]. Se la *target variable* definisce ciò che gli operatori stanno cercando, le *class labels* dividono in categorie mutualmente escludibili i risultati richiesti, nell'esempio riportato in precedenza del filtro spam, le persone concordano grosso modo sulle etichette delle classi: quali messaggi sono spam o meno. In altre situazioni, è meno ovvio quali dovrebbero essere le variabili di destinazione[22].

Questo punto fa riferimento ad uno degli aspetti più controversi relativi all'impiego degli algoritmi nonché una delle maggiori cause di trattamenti discriminatori attraverso processi decisionali algoritmici. Chiunque desideri servirsi di un algoritmo allo scopo di predire una soluzione ad una particolare questione, necessita di semplificare il quadro degli attributi ai quali vuole fare attenzione per consentire alla macchina di elaborarli. In altre parole, gli operatori, anche di fronte a dati non discriminatori e correttamente campionati, dovranno procedere ad un'attività interpretativa degli stessi per consentire il loro utilizzo da parte dell'algoritmo[23].

Si prenda, ad esempio, il caso di un'azienda che sia alla ricerca di una/o candidata/o per ricoprire una posizione al suo interno. L'idea è quindi quella di assumere un soggetto modello e per farlo sarà necessario indicarne le qualità: L'azienda potrebbe scegliere "essere raramente in ritardo" come etichetta di classe per valutare se un dipendente è considerato positivamente per l'assunzione. In questo caso i soggetti con redditi più bassi, che solitamente vivono più lontano dal luogo di lavoro, si troverebbero in una posizione di svantaggio, anche se superano gli altri dipendenti sotto altri aspetti[24].

II) e III) etichettatura (labelling) dei dati di addestramento e (III) raccolta dei dati di addestramento

Queste due operazioni si riferiscono a due tipologie di casi piuttosto simili, la discriminazione deriverebbe in queste ipotesi dai passaggi relativi al trattamento dei dati di

[20] F. Zuiderveen Borgesius, *Discrimination, artificial intelligence and algorithmic decision-making*, cit.

[21] P. Dourish, *Algorithms and their others: Algorithmic culture in context* in *Big Data & Society*, 3(2), 2016; F. Zuiderveen Borgesius, *Discrimination, artificial intelligence and algorithmic decision-making*, cit. C. O'Neil, *Weapons of math destruction: How big data increases inequality and threatens democracy*, cit.

[22] S. Barocas-A.D. Selbst, *Big Data's disparate impact*, cit., 678.

[23] Ivi, 679.

[24] F. Zuiderveen Borgesius, Ivi, 11.

addestramento dell'algoritmo[25].

Secondo il punto (II), i dati risultano discriminatori non perché introducano nuovi elementi o correlazioni che portino ad un simile risultato, ma perché riproducano un comportamento discriminatorio già presente nella società. L'esempio classico in questo caso è quello di una azienda sanitaria inglese che di fronte ad un numero rilevante di candidati per alcune posizioni nella sua struttura decide di utilizzare un programma per velocizzare le procedure di selezione. Il programma viene addestrato utilizzando i dati provenienti dalle precedenti assunzioni[26].

Il risultato è stato che il programma discriminava nei confronti di alcune categorie sociali come le donne e la popolazione non autoctona. In questo caso non è stato introdotto alcun elemento discriminante, semplicemente nelle precedenti procedure qualcuno aveva sistematicamente discriminato quelle particolari categorie sociali, di conseguenza i dati utilizzati per l'addestramento dell'algoritmo erano già di partenza alterati[27].

Al punto (III) invece si fa riferimento al momento precedente a quello appena trattato, al passaggio relativo al campionamento dei dati da utilizzare nell'addestramento dell'algoritmo[28]. Un esempio utile per questo caso riguarda l'app *Street Bump*, un'app che utilizza le informazioni GPS degli utenti per segnalare alle amministrazioni pubbliche quali sono le strade che devono ricevere manutenzione. In questo caso il campionamento risultava distorto dal fatto che le segnalazioni avvenivano per le strade dove il numero di smartphone era maggiore, perciò le strade dei quartieri più abbienti ricevevano maggiore assistenza rispetto a quelle di quelli meno abbienti, dove la percentuale di telefoni di nuova generazione era inferiore[29].

IV) la selezione degli indicatori

Questa operazione riprende alcuni aspetti del punto (II) sulla creazione delle *class labels*, classificazioni attraverso le quali i dati vengono trattati per consentire alla macchina di processarli, talvolta questo processo può essere troppo costoso o risultare troppo lungo, per questo motivo si scelgono delle particolari caratteristiche che fungono da indicatori per l'algoritmo, ad esempio un particolare tipo di educazione o qualche corso specialistico. Tali scelte però possono tradursi in comportamenti discriminanti nei confronti di alcune categorie sociali, se si addestra l'algoritmo a preferire candidati da università private particolarmente costose, automaticamente verranno esclusi gli individui meno abbienti e appartenenti alle minoranze che statisticamente sono meno presenti all'interno di quel tipo di istituti educativi[30].

---

25 S. Barocas - A.D. Selbst, Ivi, 680-1; F. Zuiderveen Borgesius, *Ibid.*

26 L'esempio è ampiamente illustrato in S. Lowry - G. Macpherson, *A blot on the profession*, in *Br Med J*, 296, 1988, 657.

27 F. Zuiderveen Borgesius, *Discrimination, artificial intelligence and algorithmic decision-making,* cit., 12.

28 S. Barocas-A.D. Selbst, *Big Data's disparate impact*, cit., 685; D. Robinson-L. Koepke. *Stuck in a pattern,* 2016.

29 Federal Trade Commission, *Big data: A tool for inclusion or exclusion? Understanding the issues*, cit, 27. Per un approfondimento su temi delle politiche di sorveglianza al tempo dell'AI, si veda A.G. Ferguson, *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*, New York, 2017.

30 S. Barocas-A.D. Selbst, Ivi, 689.

V) Proxies

Questo punto riguarda i cosiddetti *proxies*, quei dati che sono stati campionati per essere utilizzati come addestramento per l'algoritmo, ma che sono indirettamente collegati a particolari categorie sociali e possono quindi comportare un comportamento discriminatorio da parte della macchina. I dati sono, in questo caso, perfettamente neutrali e non comportano come al punto precedente una scelta a monte da parte degli operatori, si tratta di dati che possono ricollegarsi a determinate situazioni e quindi sono denominati *proxies*, perché il risultato discriminatorio non si determina per effetto dei dati immessi nella macchina, ma le caratteristiche ad essi ricollegate[31].

VI) discriminazione algoritmica volontaria

Infine l'ultimo caso contempla l'ipotesi più diretta nella quale gli operatori abbiano un positivo intento discriminatorio alla base dell'impiego dell'algoritmo, in questi casi di solito si servono di *proxies* ovvero di campionature di dati non corrette per consentire gli esiti discriminatori che si vogliono ottenere[32].

Il quadro che si presenta è piuttosto complicato, la tipologia di funzionamento dell'algoritmo e la complessità del processo decisorio che lo coinvolge complica non poco la regolamentazione giuridica dell'utilizzo di questo tipo di software. In questo senso, appare difficile seguire un approccio *ex post* sul tema perché l'utilizzo di principi generali come quelli esposti in precedenza non consentirebbe una tutela efficace nei diversi tipi di casi appena trattati[33].

2.3 La tutela antidiscriminatoria multilivello

La normativa in materia antidiscriminatoria, nei diversi ordinamenti occidentali, si serve di strumenti diversi in particolare nei sistemi giuridici che si prenderanno in considerazione: quello statunitense e quello europeo. In entrambi questi contesti si parla di tutele multilivello delle situazioni giuridiche soggettive che si integrano nella protezione di taluni interessi come quelli relativi all'eguaglianza sostanziale degli individui. Nello specifico a livello europeo si possono distinguere due diverse fonti del diritto, da una parte c'è la Convezione Europea dei Diritti dell'Uomo e dall'altra la Carta dei diritti fondamentali, la direttiva in tema di antidiscriminazione e il regolamento GDPR[34]. La convenzione disciplina la tutela antidiscriminatoria all'art. 14 della Convezione, che tutela in via generale il godimento dei diritti senza discriminazioni di sesso, razza,

[31] Ivi, 692.

[32] Sulla discrimazione intenzionale si veda P. T. Kim, *Data-driven discrimination at work*, cit., 857 J. Bryson, *Three very different sources of bias in AI, and how to fix them,* in *Adventures in NI*, 13 July 2017; B. Friedman-H. Nissenbaum, *Bias in computer systems,* in *ACM Transactions on Information Systems* (TOIS), 14(3), 1996, 330.

[33] J. A. Kroll et al., *Accountable algorithms*, in 165 *University of Pennsylvania Law Review*, 2016, 633 ss.

[34] P. Hacker, *Teaching fairness to artificial intelligence: Existing and novel strategies against algorithmic discrimination under EU law*, 55 *Common Market Law Review*, 4, 2018 1143 ss.; N. Helberger-F. Zuiderveen Borgesius-A. Reyna, *The perfect match? A closer look at the relationship between EU consumer law and data protection law*, in *Common Market Law Review*, 54, 2017, 1427 ss.

colore, la lingua, la religione, le opinioni politiche o quelle di altro genere, l'origine nazionale o sociale, l'appartenenza a una minoranza nazionale, la ricchezza, la nascita od ogni altra condizione. Tale disposizione è ulteriormente rafforzata dalla previsione del protocollo 12 alla Convezione che riafferma una tutela omnicomprensiva in ambito antidiscriminatorio, oltre l'elenco contenuto all'art. 14[35].

La giurisprudenza della Corte Edu si è successivamente assunta il compito di definire le due tipologie principali di discriminazione, quella diretta che fonda la differenza di trattamento in analoghe o similari situazioni basata su caratteristiche identificabili. Al contrario la discriminazione indiretta si pone in essere attraverso un'azione che è all'apparenza neutrale, ma si risolve in un atto discriminatorio nei confronti di una particolare categoria sociale. Uno studio delle sentenze della Corte sul tema ha evidenziato come la regolamentazione a tutela della discriminazione indiretta non sia lineare come nel caso di quella diretta[36].

Una disciplina molto simile è quella prevista dall'ordinamento UE, la Carta fondamentale di Nizza prevede una disciplina che riprende quella della CEDU, all'art. 21, c. 1 si legge «é vietata qualsiasi forma di discriminazione fondata, in particolare, sul sesso, la razza, il colore della pelle o l'origine etnica o sociale, le caratteristiche genetiche, la lingua, la religione o le convinzioni personali, le opinioni politiche o di qualsiasi altra natura, l'appartenenza ad una minoranza nazionale, il patrimonio, la nascita, la disabilità, l'età o l'orientamento sessuale»[37].

Tale prescrizione è stata successivamente positivizzata anche in una serie di direttive a partire dalla n. 43 del 2000. Il regime disposto è sostanzialmente speculare a quello instaurato dalla CEDU; anche in questo caso viene delineata una discriminazione diretta di semplice interpretazione e una indiretta, che si distingue per una disciplina maggiormente discrezionale[38].

La normativa predispone, come si è visto in precedenza rispetto alla giurisprudenza della Corte Edu, la possibilità di liberarsi da parte del convenuto dimostrando che la presunta condotta discriminatoria è avvenuta allo scopo di ottenere un risultato legittimo e che gli strumenti utilizzati per raggiungere tale obiettivo sono necessari e proporzionati; il principio di proporzionalità è impiegato anche nella legislazione EU come strumento per contemperare la tutela antidiscriminatoria con altri interessi[39].

---

[35] Sono diversi i trattati e le convezioni che predispongono una tutela antidiscrimatoria, la Dichiarazione universale dei Diritti d'uomo dell'ONU all'art. 7, la Convezione Europea dei diritti dell'uomo all'art. 14 e al protocollo 12 che non è ancora stato ratificato da tutti i membri, la Carta fondamentale dei diritti dell'Unione Europea all'art. 21 che verrà trattato in questo paragrafo e l'art. 26 della Convenzione Internazionale dei diritti civili e politici all'art. 26.

[36] Su questo punto nella giurisprudenza della Corte Edu si vedano le sentenze, *Biao v. Denmark* (Grand Chamber), ric. 38590/10 (2016), § 89; *D.H. and Others v. Czech Republic*, ric. 57325/00 (2007), §§ 187-188.

[37] Carta fondamentale dei diritti dell'Unione Europea, art. 21, 2002.

[38] La distinzione è rinvenibile nella normativa europea in tema di antidisciminazione, la 2000/43/CE che attua la parità di trattamento tra gli individui indipendemente dalla razza e dall'origine etnica, la 2000/78/CE che stabilisce una disciplina in materia di parità di trattamento in materia di occupazione, la 2004/113/CE che disciplina la parità di trattamento nell'ambito dell'accesso a beni e servizi e la 2006/54/CE che regola le pari opportunità e la parità di trattamento tra uomini e donne in materia di occupazione e impiego.

[39] Su questa eccezione si veda l'art. 2, par. 2, lett. *b)*, della direttiva 2000/43/EC: «…a meno che

2.4 La discriminazione algoritmica attraverso la normativa sul trattamento dei dati personali

Nel contesto dell'Intelligenza artificiale, una regolamentazione che può rivelarsi particolarmente utile nell'ambito dell'ordinamento UE al trattamento dei dati personali. In questo campo, la tutela approntata è meno legata a principi generali e più imperniata su singole regole che disciplinano in modo chiaro la protezione accordata dal diritto alla privacy. La normativa promossa in ambito europeo sul tema enuclea alcuni principi come la trasparenza, l'integrità, la responsabilità e la confidenzialità, ma fonda la propria disciplina su prescrizioni che regolano ogni passaggio della gestione dei dati personali: dalla raccolta all'archiviazione, dallo scopo al controllo del trattamento[40].
Il regolamento adottato dall'UE in questo campo, il *General Data Protection Regulation* offre una serie di disposizioni che vanno proprio in questa direzione, stabilendo che le autorità garanti dei Paesi membri possono richiedere informazioni e l'accesso al sistema di elaborazione dei dati ai responsabili del trattamento, possono accedere ai luoghi fisici dove avviene la gestione dei dati e condurre audit su un particolare utilizzo di intelligenze artificiali[41]. In tema di processi decisori algoritmici, il GDPR ha predisposto una serie di regole a cominciare dall'art. 22, che si occupa di quelle decisioni prese con il solo ausilio degli algoritmi e per il quale: *«The data subject shall have the right not to be subject to a decision based solely on automated processing including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her»* (c.1)[42].
Il diniego di prestazioni assistenziali o pensionistiche, ovvero una pronuncia di un organo giurisprudenziale o di una pubblica amministrazione che fondino la propria decisione esclusivamente su di un processo algoritmico fanno nascere in capo ai soggetti coinvolti un diritto a ricorrere contro la stessa. Per la trattazione in oggetto è interessante notare come questa tutela si applica anche nei casi di *profiling*[43].

---

tale disposizione, criterio o prassi siano oggettivamente giustificati da una finalità legittima e i mezzi impiegati per il suo conseguimento siano appropriati e necessari.» che è ripreso specularmente nella sentenza della Corte Edu *Biao v. Denmark* cit., § 91-2.

[40] Sulla normativa dell'UE in tema di privacy con particolare riferimento all'intelligenza artificiale e ai Big Data si veda European Data Protection Supervisor, *Privacy and competitiveness in the age of big data: The interplay between data protection, competition law and consumer protection in the Digital Economy*, Marzo 2014; G. Buttarelli, *Towards a New Digital Ethics: Data, Dignity and Technology, Speech before the Institute of International and European Affairs*, Dublino, 2015, 1-4; F. Pizzetti (a cura di), *Intelligenza artificiale, protezione dei dati personali e regolazione,* cit.

[41] Sulla relazione tra GDPR e tutela antidiscriminatoria, si veda W. Schreurs-M. Hildebrandt-E Kindt-M. Vanfleteren, *Cogitas, ergo sum. The role of data protection law and non-discrimination law in group profiling in the private sector*' in M. Hildebrandt-S. Gutwirth, (a cura di) *Profiling the European citizen*, Heidelberg, Berlino, 2008; P. Hacker, *Teaching fairness to artificial intelligence: Existing and novel strategies against algorithmic discrimination under EU law*, cit., 1143 ss.; F. Zuiderveen Borgesius, *Discrimination, artificial intelligence and algorithmic decision-making*, cit., 21.

[42] Art. 21 GDPR. In questo ambito si rimanda a I.. Mendoza-L. A. Bygrave *The right not to be subject to automated decisions based on profiling*, University of Oslo, in Research paper no. 20 (2017). Sia sul tema in generale, che con specifico riferimento alla discriminazione in tema di online *pricing*, si veda F. Zuiderveen Borgesius - J. Poort, *Online price discrimination and EU data privacy law*, in *Journal of Consumer Policy*, 2017,1 ss.

[43] L'art. 4, par. 4, del GDPR definisce il profiling come «qualsiasi forma di trattamento automatizzato di dati personali consistente nell'utilizzo di tali dati personali per valutare determinati aspetti personali

L'art. 22 stabilisce delle eccezioni alla regola enunciata al c.1, il divieto non si applica se la decisione automatizzata (i) è basata sul consenso esplicito dell'interessato; (ii) è necessaria per un contratto tra la persona fisica e il titolare del trattamento; o (iii) questo è autorizzato dalla legge. I primi due casi però fanno scattare l'applicazione di un'altra disciplina per la quale il responsabile del trattamento è tenuto ad attuare misure adeguate per salvaguardare i diritti, le libertà e i legittimi interessi dell'interessato e almeno il diritto di ottenere l'intervento umano da parte del responsabile del trattamento e per contestare la decisione (art. 22, par. 3, GDPR)[44].

Questa previsione diviene estremamente interessante perché consente la possibilità di ricorrere nei confronti di decisioni espresse esclusivamente attraverso procedure algoritmiche che possono aver compresso i diritti dei soggetti coinvolti dalle stesse. In alcuni casi si è parlato di un *algorithmic due process* che tutela alcune libertà e diritti fondamentali degli individui nei confronti di tali processi decisionali[45].

Rimane ancora un argomento di discussione se queste obbligazioni diano vita ad un diritto di spiegazione in capo ai soggetti coinvolti dalla decisione in virtù del quale sia possibile richiedere un chiarimento rispetto a qualsiasi risultato individuato con l'ausilio dell'intelligenza artificiale. Questo obbligherebbe gli operatori a porre in essere quei passaggi richiesti dalla trasparenza informatica e consentirebbe ai soggetti coinvolti di avere informazioni tecniche attraverso le quali cercare di comprendere i passaggi del processo decisorio, come previsto dall' art. 9 *Data Protection Convention* 108 del Consiglio d'Europa del 2018. In generale però è spesso difficile spiegare la logica alla base di una decisione, come un algoritmo, analizzando grandi quantità di dati, arriva a quella decisione. In alcuni casi, non è chiaro quanto una spiegazione potrebbe aiutare i soggetti interessati, soprattutto nella misura in cui pone l'onere di comprendere la stessa decisione e la sua adeguatezza in capo a loro[46].

La normativa in tema di protezione dei dati personali offre una tutela più articolata nei confronti delle decisioni che coinvolgono l'intelligenza artificiale, la trattazione tende a reggersi non solo su principi generali, che sono più difficili da applicare alle diverse casistiche tecniche che possono verificarsi all'interno di un processo decisorio algoritmico, ma anche sul lavoro svolto dalle autorità garanti che possono sviluppare

relativi ad una persona fisica..» Su questo tema si veda, O. De Schutter-J. Ringelheim, *Ethnic profiling: A rising challenge for European human rights law*, cit., 358 ss.; B. E. Harcourt, *Against prediction: Profiling, policing, and punishing in an actuarial age*, Chicago, 2008.

44 Su questo aspetto si veda, F. Zuiderveen Borgesius, *Discrimination, artificial intelligence and algorithmic decision-making*, cit., 22. Sulla possibilità di regolamentare questo ambito, P. De Hert-S. Gutwirth, *Regulating profiling in a democratic constitutional state* in M. Hildebrandt - M.S. Gutwirth (a cura di), *Profiling the European Citizen*, cit.

45 Sull'individuazione di un *algorithmic due process* si rimanda a M. Kaminski, *Binary governance: A two-part approach to accountable algorithms*, in *S. Calif. L. Rev.* (2018) 92. Sullo stesso punto, si parla più in generale di *technological due process*, K. Citron, *Technological due process*, in 85 *Wash.UL Rev.*, 2007, 1249.

46 Sul tema del *right of explanation* si veda, L. Edwards-M. Veale, *Slave to the algorithm: Why a right to an explanation is probably not the remedy you are looking for*, in 16 *Duke L.& Tech.Rev.*, 2017, 18; Id., *Enslaving the algorithm: from a "right to an explanation" to a "right to better decisions"?*, in *IEEE Security & Privacy*, 16(3), 2018, 46 ss.; M. Kaminski, *The right to explanation, explained,* 2018; G. Malgieri, *Right to explanation and algorithm legibility in the EU Member States legislations*, 17 Agosto 2018; cfr. S. Wachter-B. Mittelstadt-L. Floridi, *Why a right to explanation of automated decision-making does not exist in the general data protection regulation*, in *International Data Privacy Law*, 2017, 2.

regolamentazioni di settore o codici di condotta in specifici ambiti che consentono una maggiore versatilità delle regole approntate, che sono così in grado di rapportarsi in modo sostanziale con le caratteristiche informatiche proprie di ogni algoritmo[47].
Questi aspetti positivi sono però attenuati da una serie di obiezioni che la dottrina in materia ha posto in essere rispetto alla tutela antidiscriminatoria attraverso la protezione dei dati personali. In primo luogo, la disciplina in esame può essere utilizzata solo nel caso in cui il processo decisorio algoritmico gestisca dati personali, nel caso in cui non siano coinvolti dati classificati come tali, non è possibile ricorrere all'autorità garante o fare affidamento sulla normativa. Ad esempio molti dei dati a cui abbiamo fatto riferimento nei casi esposti nel precedente paragrafo, il CAP, il percorso di studi, ma soprattutto perché l'algoritmo di addestramento che crea il modello predittivo non si serve di dati personali, elabora un percorso decisorio sulla base di dati generali[48].

2.5 La tutela antidiscriminatoria negli Stati Uniti d'America nel campo dei processi decisori algoritmici

L'antidiscriminazione nel mondo giuridico americano è uno strumento particolarmente sviluppato capace di adattarsi nel corso degli anni ai vari ambiti di applicazione. A livello costituzionale federale, il fulcro della tutela ruota attorno alla *equal protection clause* contenuta all'interno del XIV emendamento, che garantisce uguale protezione a tutti gli individui da parte del diritto. Questa prescrizione mirava a tutelare i diritti dei soggetti nei confronti degli stati federati, a livello federale, è garantita dalla giurisprudenza della Corte Suprema un'uguale tutela dal quinto emendamento alla costituzione. Tale prescrizione a livello costituzionale a consentito l'introduzione di una serie di strumenti normativi sia a livello nazionale che statale a tutela delle categorie sociali maggiormente esposte a trattamenti discriminatori. In particolare è bene ricordare il *Civil Right Act* del 1964 che riconosce una tutela generale nei confronti di ogni tipo di discriminazione, con particolare riferimento al tema di questo scritto, il titolo VII prevede una garanzia risarcitoria nei confronti di tutte le forme di discriminazione indiretta (*disparate treatment*)[49].
L'ordinamento statunitense, nonostante la posizione di avanguardia dell'industria nazionale nell'ambito dell'intelligenza artificiale, non possiede una legislazione generale sul tema. La produzione normativa ha riguardato principalmente un paio di ordini esecutivi delle amministrazioni Obama e Trump che hanno delineato i piani industriali nel campo dell'intelligenza artificiale, alcune regolamentazioni di determinati settori come quella relativa alle automobili senza conducente e una serie di proposte di legge

[47] Su questo punto si veda S. Bornstein, *Antidiscriminatory algorithms*, cit., 522 ss.; J. Kleinberg-J. Ludwig-S. Mullainathan-C. Sunstein, *Discrimination in the age of algorithms*, cit., 5.

[48] Sui limiti della normativa sul trattamento dei dati: S. Wachter-B. Mittelstadt, *A right to reasonable inferences: Re-thinking data protection law in the age of Big Data and AI*, cit., 81 ss; M. Ananny-M K. Crawford, *Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability*, in *New Media & Society* 20(3), 2018, 973.

[49] Su questo punto in relazione con i processi decisori algoritmici si veda, S. Bornstein, *Antidiscriminatory algorithms*, cit., 525 ss.

presentate davanti al Congresso e in attesa di essere votate[50].
Di queste ultime, gli esempi più importanti sono stati i progetti di legge presentati al Senato e alla Camera relativi all'*Algorithmic Accountability Act* (S. 1108, H.R.2231) che sono stati introdotti al Congresso il 10 aprile 2019, probabilmente in risposta ai rapporti recentemente pubblicizzati sui rischi dei risultati distorti prodotti dall'utilizzo dell'intelligenza artificiale[51]. Il disegno di legge mira a richiedere agli attori privati e alle entità pubbliche di condurre delle valutazioni d'impatto sui loro sistemi decisionali automatizzati considerati "ad alto rischio" al fine di verificare in quale modo il processo decisorio algoritmico si relaziona con i principi generali di accuratezza, correttezza, privacy e sicurezza[52]. La proposta prevede inoltre che tali valutazioni dovrebbero essere condotte da terze parti esterne, compresi revisori ed esperti tecnologici indipendenti[53]. Questa normativa richiede agli operatori privati e pubblici che impiegano intelligenze artificiali di condurre valutazioni di impatto del tutto simili a quelle proposte dal disegno di legge federali sui processi decisionali algoritmici e sui sistemi informativi. Tale verifica implica una valutazione del processo di sviluppo del sistema, compresa la componente relativa al suo *design and training data*, deve contenere una descrizione dettagliata delle migliori pratiche utilizzate per minimizzare i rischi e un'analisi costo-benefici[54].

3. La discriminazione algoritmica in pratica: una serie di casi sostanziali

Uno studio esaustivo di questo tema non può realizzarsi senza un approfondimento di alcuni casi sostanziali che hanno riguardato la tutela discriminatoria rispetto all'impiego degli algoritmi all'interno dei processi decisori elaborati dalla pubblica amministrazione o da attori privati. Come si è illustrato nella sezione precedente, la casistica è alquanto ampia e si è scelto quindi di concentrare l'attenzione su due categorie che riguardano la selezione di dipendenti e studenti, la pubblicità online. È ovvio che questa disamina esclude alcuni ambiti molto rilevanti quali la pubblica sicurezza, le traduzioni automatiche, la ricerca di immagini nella rete e la discriminazione di prezzo.

Selezione e valutazione dipendenti pubblici
Si è già visto come l'intelligenza artificiale possa essere utilizzata per selezionare po-

[50] Executive Office of the President National Science and Technology Council, Committee on Technology, *Preparing for the Future of Artificial Intelligence*, Ottobre 2016. Questo primo documento dell'Amministazione Obama è stato seguito da un ordine esecutivo dell'amministrazione Trump, Exec. Order No. 13,859, 3 C.F.R. 396, 2019.

[51] Algorithmic Accountability Act of 2019, S. 1108, H.R. 2231, 116th Cong. (2019). Sulla proposta di legge si veda S. Revanur, *In a Historic Step Toward Safer AI, Democratic Lawmakers Propose Algorithmic Accountability Act*, in *Medium* (20 aprile 2019).

[52] Algorithmic Accountability Act (2019) par. 2 c. 2 e 3 lett. (b).

[53] Algorithmic Accountability Act (2019) par. 3 lett. (b)(1)(C).

[54] E. J. Tail et al., *Proposed Algorithmic Accountability Act Targets Bias in Artificial Intelligence*, in *JD Supra*, 27 giugno 2019.

tenziali dipendenti o studenti. Un caso di questo genere ha riguardato Amazon[55], che impiegava un sistema di intelligenza artificiale per la selezione dei candidati alle posizioni di lavoro nell'azienda, i programmatori l'avevano addestrato a trovare modelli nei curriculum inviati per tecnici informatici nei dieci anni precedenti, la maggior parte dei quali, a causa della demografia di chi detiene quei posti di lavoro, proveniva da candidati uomini. Di conseguenza l'algoritmo aveva sviluppato un modello che lo portava a preferire i candidati maschi, perciò qualsiasi frase che includesse la parola "donna" o sue derivazioni, come in "capitano del club di scacchi femminile" comportava l'automatica esclusione del curriculum della candidata[56].

Nel caso di Amazon, l'azienda ha posto rimedio accantonando il software nella impossibilità di escludere ogni possibile esito discriminatorio del processo decisorio algoritmico, in altri casi distinguere la portata di alcuni dati è molto più difficile, di conseguenza provare eventuali effetti discriminatori diventa complicato. In questo senso è emblematico il caso *McKinzy v. Union Pacific*[57], nel quale l'attore, un candidato ad una posizione lavorativa ha citato in giudizio il potenziale datore di lavoro, sulla base di una supposta discriminazione razziale da parte dell'algoritmo[58]. Il procedimento ha stabilito che l'azienda aveva utilizzato un algoritmo per valutare il curriculum dell'attore, il quale ha affermato di essere stato escluso in ragione della sua origine etnica, la difesa della Union Pacific ha fornito i dati relativi all'esperienza lavorativa dell'attore, sottolineando come McKinzy non fosse qualificato per la posizione aperta secondo dati che non tenevano conto dell'appartenenza etnica. La corte si è pronunciata a favore della Union Pacific proprio per la sua capacità di motivare sulla base di criteri neutri il rigetto del candidato[59]. Si è visto in precedenza come simili criteri possano comunque comportare un risultato discriminatorio indiretto se sono collegati (*proxies*) con caratteristiche relative a determinate categorie sociali. Rispetto alla discriminazione indiretta si è già avuto modo di parlare di come sia la Corte Edu, che le direttive UE[60] consentano l'eccezione del perseguimento di un obiettivo legittimo e necessario. In questo senso la Corte Suprema degli Stati Uniti[61] aveva già tracciato la strada con la teoria della "*business necessity*", per la quale, se un datore di lavoro può dimostrare che una particolare misura o politica aziendale è necessaria per la conduzione dell'impresa anche se questa comporta il verificarsi di una discriminazione indiretta. Con riferimento ai casi di discriminazione algoritmica, questa difesa potrebbe applicarsi se il datore di lavoro può

---

55 Il caso è raccontato in J. Dastin, *Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women*, in *Reuters*, 9 ottobre, 2018.

56 La rilevanza giuridica del caso è esposta in S. Bornstein, *Antidiscriminatory algorithms,* cit., 521.

57 *McKinzy v. Union Pac. R.R.,*2010WL3700546(2010).

58 *Ibid.*

59 *Ibid.*

60 Si rimanda alla sentenza Corte Edu, *D.H. and Others v. Czech Republic*, ric. 57325/00, (2007), §§ 187-188; art. 2 , par. 2, direttiva 2000/43/EC.

61 La teoria della *business necessity* è stata elaborata nel precedente *Griggs v. Duke Power Co.*, 401 U.S. 424 (1971).Si veda sulla questione I. Ajunwa, *Automated Hiring*, cit., 41-2. L'utilizzo del dato statistico per provare la *business necessity* è richiamato sia da Ajunwa per quanto riguarda la Corte Suprema, sia da F. Zuiderveen Borgesius, *Discrimination, artificial intelligence and algorithmic decision-making*, cit., 19 per quanto concerne la Corte Edu e la normativa UE.

dimostrare che i dati su cui fa affidamento un algoritmo sono una "necessità aziendale" o, in altre parole, sono statisticamente correlati ad una corretta gestione dell'azienda.

Pubblicità online

L'intelligenza artificiale è impiegata per la pubblicità online mirata, in questo campo, l'algoritmo rielabora i dati forniti dagli utenti attraverso le loro interazioni per proporre segnalazioni pubblicitarie su misura. Questo impiego dell'intelligenza artificiale è stato segnalato perché particolarmente esposto a pratiche discriminatorie, già nel 2013 si è dimostrato come le persone che cercassero nomi di provenienza afroamericana, venivano esposti dal motore di ricerca di Google ad annunci pubblicitari per soggetti destinatari di condanne penali o precedenti di polizia. Al contrario, se si digitavano nomi di provenienza caucasica, lo stesso motore di ricerca individuava un numero significativamente inferiore di annunci connessi ai precedenti giudiziari dei destinari[62].
Nel 2015 uno studio condotto da ricercatori ha elaborato una simulazione nella quale utenti del motore di ricerca di Google, che si sono auto-dichiarati maschi o femmine nelle impostazioni, effettuavano identiche ricerche online sulla piattaforma. I ricercatori hanno quindi analizzato gli annunci presentati dall'algoritmo. Google ha mostrato annunci pubblicitari agli utenti simulati maschili da una certa agenzia di consulenza che prometteva salari elevati con frequenza significativamente maggiore rispetto a quelli proposti alle utenti simulate donne con effetti discriminatori. Lo studio ha inoltre notato come non sia possibile individuare il motivo per il quale alle utenti simulate donne è stato mostrato un numero inferiore di annunci pubblicitari per impieghi a salari elevati, a causa dell'opacità del sistema automatico che gestisce la piattaforma e elabora i diversi dati immessi dagli utenti[63].
Un altro caso interessante ha riguardato le modalità di inserzione degli annunci pubblicitari proposte dalla piattaforma Facebook. Il social network ha consentito agli inserzionisti di indirizzare gli annunci pubblicitari agli utenti sulla base di una serie di dati sensibili processati dal suo algoritmo come ad esempio, i dati relativi alle preferenze sessuali[64]. Il Garante per la protezione dei dati olandese ha posto in essere un'indagine in questo ambito e ha provveduto a registrare una serie di profili falsi sulla piattaforma indicanti, tra le varie informazioni richieste, la categoria "*men that are interested in other men*"[65]. La totalità dei profili registrati non ha posto in essere ulteriori interazioni sulla piattaforma ed è stata esposta a campagne pubblicitarie mirate per quella specifica categoria[66].

---

62 L. Sweeney, *Discrimination in online ad delivery*, in *Queue*, 2013, 11(3).

63 Lo studio empirico è descritto dagli autori in A. Datta - M. C. Tschantz - A. Datta, *Automated experiments on ad privacy settings*, in *Proceedings on Privacy Enhancing Technologies*, 2015, 92 A. Datta et al., *Discrimination in online advertising: A multidisciplinary inquiry,* (Conference on Fairness, Accountability and Transparency 2018) 20.

64 Autorità garante del trattamento dei dati olandese, *Dutch data protection authority: Facebook violates privacy law*, 16 maggio 2017.

65 Autorità garante del trattamento dei dati olandese, *Informal English translation of the conclusions of the Dutch Data Protection Authority in its final report of findings about its investigation into the processing of personal data by the Facebook group*, 23 febbraio 2017, 3.

66 *Ibid.*

Un ulteriore aspetto relativo alle inserzioni pubblicitarie sul social network Facebook è stato oggetto di un procedimento davanti ad una corte federale statunitense. La piattaforma consentiva agli inserzionisti, attraverso una serie di menù a scelta denominato "affinità etniche", di escludere una serie di categorie sociali come la popolazione di origine africana o ispanica dalla visualizzazione degli annunci[67]. La piattaforma consentiva anche l'esclusione di precisi gruppi di individui quali *"women in the workforce,", "moms of grade school kids,", "foreigners", "Puerto Rico Islanders"*; ovvero soggetti interessati a *"parenting," "accessibility," "service animal," "Hijab Fashion", "Hispanic Culture"* ovvero la pubblicizzazione di annunci di lavoro solo a persone di una determinata fascia di età[68]. Sulla base di queste risultanze una serie di organizzazioni no profit che operano nell'ambito del diritto all'abitazione e della tutela antidiscriminatoria nel mercato immobiliare hanno citato in giudizio[69] Facebook per la violazione del *Fair Housing Act* che prescrive come *«[t]o make, print, or publish, or cause to be made, printed, or published any notice, statement, or advertisement, with respect to the sale or rental of a dwelling that indicates any preference, limitation, or discrimination based on race, color, religion, sex, handicap, familial status, or national origin, or an intention to make any such preference, limitation, or discrimination»*[70]. Il ricorso lamentava essenzialmente tre condotte discriminatorie da parte della piattaforma: la possibilità per gli inserzionisti di includere o escludere gli utenti in base al loro sesso, età, interessi, comportamenti o dati demografici che si presume siano correlati o associati a razza, origine nazionale, disabilità o stato di famiglia; La possibilità per gli inserzionisti di definire un'area geografica ristretta per il pubblico degli annunci che potrebbe presumibilmente avere un impatto negativo in base alla razza o all'origine nazionale; 3) il possibile impiego dello strumento informatico *Lookalike Audience* da parte degli inserzionisti che avrebbe permesso di creare segmenti di pubblico tra gli utenti secondo dinamiche in grado di avere un impatto negativo su vari gruppi, anche in base a sesso, razza ed età[71]. Le parti hanno deciso di addivenire ad un accordo stragiudiziale[72] sulla base di una serie di comportamenti che il convenuto ha acconsentito di adottare per emendare le pratiche discriminatorie citate nel ricorso, tra le quali: a) la piattaforma creerà un portale pubblicitario separato per la creazione di annunci di alloggi, occupazione e credito ("HEC") che avrà opzioni di *targeting* limitate, per prevenire la discriminazione; b) Facebook elaborerà una pagina in cui gli utenti possano cercare e visualizzare tutti gli annunci che sono stati inseriti dagli inserzionisti per l'affitto e la vendita di alloggi indipendentemente dal fatto che gli utenti abbiano ricevuto tali annunci immobiliari sulla loro bacheca; c) la piattaforma richiederà agli inserzionisti di certificare il rispetto delle politiche di Facebook che vietano la discriminazione e tutte le leggi anti-discrimi-

---

67 Su questo punto si veda J. Angwin et al., *Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks*, in *ProPublica*, Maggio 2016; J. Angwin-A. Tobin-M. Varner, *Facebook (still) letting housing advertisers exclude users by race*, in *ProPublica*, Novembre 2017.

68 J. Angwin-N. Scheiber-A. Tobin, *Dozens of companies are using Facebook to exclude older workers from job ads*, in *ProPublica*, Dicembre 2017.

69 NFHA v. Facebook complaint, Caso 1:18-cv-02689, 25 giugno 2019.

70 42 U.S.C. § 3604(c).

71 NFHA v. Facebook complaint, Caso 1:18-cv-02689, cit.

72 NFHA v. Facebook settlement agreement, 18 marzo 2019.

nazione applicabili; d) Facebook consentirà ai querelanti di testare la sperimentazione della piattaforma pubblicitaria per garantire che le modifiche stabilite dall'accordo siano attuate in modo efficace[73].
Un ulteriore ricorso è stato presentato da parte del *Housing and Urban Development Department* degli Stati Uniti D'America sulla base delle stesse risultanze della causa appena riportata[74] Con un atto di citazione nell'estate del 2019 il Dipartimento federale conviene in giudizio la piattaforma Facebook per la violazione del *Fair Housing Act*. Non si conoscono ad oggi le sorti di questo contenzioso, le pratiche discriminatorie a cui si fa riferimento si estendono fino al primo trimestre dello scorso anno[75].

4. A mo' di conclusione

È difficile poter tracciare un quadro definitivo di una disciplina ancora in piena evoluzione. Lo sviluppo tecnologico nel campo dell'intelligenza artificiale è in rapida espansione e il diritto, per sua stessa natura, fatica a tenere il passo. I due approcci che sono stati utilizzati per presentare la disciplina antidiscriminatoria rispetto all'impiego degli algoritmi rappresentano le basi con cui si sta affrontando questo sforzo. Come si è detto l'approccio *ex ante* è spesso preferibile specie in quei contesti dove è possibile intervenire con autorità indipendenti come il Garante del trattamento dei dati che sono in grado di specificare il materiale regolamentare per adattarlo alle diverse realtà. In quegli ambiti nei quali tale cooperazione è più difficile, l'approccio *ex post* può sopperire ad eventuali lacune nella regolamentazione. Una sinergia tra questi due schemi di tutela sembra essere il modo migliore per estendere una tutela antidiscriminatoria effettiva ai processi decisori algoritmici.

[73] *Ibid.*

[74] La violazione del 42 U.S.C. § 3604(c).

[75] HUD v. Facebook complaint, FHEO No. 01-18-0323-8, 27 marzo 2019.