Algoritmi e discriminazione

Fundamental rights
Rivista di studi giuridici, storici e antropologici

Viviana Molaschi

Algoritmi e discriminazione*

Algorithms and discrimination

SOMMARIO: 1. L'ambivalenza degli algoritmi: opportunità di uguaglianza e causa di discriminazione. Cenni su *digital divide* ed esclusione − 2. Alcuni esempi: quando a discriminare è lo Stato… − 3. Anatomia della discriminazione algoritmica – 4. La risposta del diritto. Alcune considerazioni sui principi in materia di decisioni algoritmiche del GDPR e sulla proposta di regolamento UE in materia di intelligenza artificiale − 5. La rilevanza di una tutela *by education*.

Algorithms bring both economic and social benefits and can help fight social inequalities. However, they do not escape the ambivalence that characterizes technological progress: they can cause discrimination and marginalization. After disproving the myth of algorithmic infallibility and neutrality, the paper brings to attention some examples of algorithmic discrimination that have occurred in public activity. Once the issue has been exemplified, it investigates how discriminatory algorithms are generated (however, these mechanisms are not entirely perspicuous and this lack of clarity is itself part of the problem). Finally, the paper focuses on the solutions given by law, with particular reference to the algorithmic decision-making principles of the GDPR and the proposed EU regulation on artificial intelligence. Since the current regulatory system, as well as those envisaged in the future, have inherent limitations due to the speed and unpredictability of technological developments, the analysis concludes by underlining the need for a responsible education both of developers, who design algorithms, and of users of digital instruments.

KEYWORDS: Algorithms, artificial intelligence, principle of equality, discrimination.

1. L'ambivalenza degli algoritmi: opportunità di uguaglianza e causa di discriminazione. Cenni su *digital divide* ed esclusione.

L'articolo si propone di formulare alcune riflessioni sui rischi di forme di discriminazione che possono derivare dal massivo impiego degli algoritmi nelle società contemporanee[1].

Gli algoritmi sono strumenti di calcolo eccezionali, con capacità predittive un tempo impensabili. Essi sono sempre più utilizzati tanto nel settore privato, nel contesto di attività commerciali, quanto in quello pubblico e sono indubitabilmente forieri di benefici sia economici che sociali.

Come sottolineato anche dalla recente proposta di regolamento in materia di intelligenza artificiale, rappresentano «uno strumento per le persone e un fattore positivo per la società, con il fine ultimo di migliorare il benessere degli esseri umani»[2].

Proprio nell'ottica del principio di uguaglianza, che è il fulcro dell'analisi condotta in questo scritto, non si può non ricordare come gli algoritmi possano contribuire a ridurre le disparità sociali. Essi, infatti, creano occasioni di sviluppo e di crescita sia per i singoli che per la collettività. Analogamente

*Questo contributo sarà destinato agli Scritti in onore di Carlo Emanuele Gallo di prossima pubblicazione. Ringrazio i Direttori di *Fundamental Rights* per la possibilità di anticiparlo in questa sede.

[1] Sintetizzando, gli algoritmi sono «procedure codificate per trasformare i dati di 'input' in un 'output' desiderato, in base a calcoli specifici»: è questa la definizione di M. DURANTE, *Potere computazionale. L'impatto delle ICT su diritto, società, sapere*, Milano, 2019, p. 237, che si rifà, tra gli altri, a T. GILLESPIE, *The relevance of alghorithms*, in T. Gillespie, P.J. Boczkowski e K.A. Foot (curr.), *Media Technologies: Essays on Communication, Materiality and Society*, Cambridge Mass., 2014, p. 167. Gli algoritmi sono insiemi di regole/istruzioni per svolgere funzioni o risolvere problemi attraverso una serie di passi definiti. Invero, essi presentano diversi livelli di complessità, in base ai quali possono essere classificati secondo una scala qualitativa che ne mette in evidenza prevedibilità, comprensibilità e 'intelligenza'. Quello più semplice è il c.d. *white box*, interamente predeterminato nei suoi passaggi. *Grey box* è invece un algoritmo non completamente predeterminato, ma i cui aspetti non predefiniti possono essere facilmente predetti e compresi. Si parla di *black box* quando è difficile o impossibile spiegare le caratteristiche dell'algoritmo ossia capirne il funzionamento. Un algoritmo *sentient* è un algoritmo che è in grado di superare il Test di Turing, ossia ha raggiunto o anche superato l'intelligenza umana. Sono dotati della c.d. *singularity* gli algoritmi capaci di *recursive self improvement*, vale a dire di apprendere e migliorarsi senza alcuna azione esterna. Per tale tassonomia v. A. TUTT, *An FDA for Algorithms?*, in *Administrative Law Review*, LXIX (2017), p. 107, http://www.administrativelawreview.org/wp-content/uploads/2019/09/69-1-Andrew-Tutt.pdf. Il fenomeno degli algoritmi, peraltro, non può essere definito unicamente sulla base di canoni tecnologici. Rileva al riguardo R. KITCHIN, *Thinking critically about and researching algorithms*, in *Information, Communication & Society*, XX (2016), p. 1, http://dx.doi.org/10.1080/1369118X.2016.1154087, che gli algoritmi possono essere concepiti in vari modi: «*technically, computationally, mathematically, politically, culturally, economically, contextually, materially, philosophically, ethically*».

[2] Si veda il punto 1.1. della proposta della Commissione di Regolamento europeo che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale) e modifica alcuni atti legislativi dell'Unione {SEC(2021) 167 final} - {SWD(2021) 84 final} - {SWD(2021) 85 final}. Per un commento v. C. CASONATO, B. MARCHETTI, *Prime osservazioni sulla proposta di regolamento dell'Unione Europea in materia di intelligenza artificiale*, in *Biolaw Journal*, 3/2021; G. MARCHIANÒ, *Proposta di regolamento della commissione europea del 21 aprile 2021 sull'intelligenza artificiale con particolare riferimento alle IA ad alto rischio*, in *AmbienteDiritto*, 2021, https://www.ambientediritto.it/.

alle innovazioni tecnologiche che hanno contrassegnato le precedenti rivoluzioni industriali, gli algoritmi operano una redistribuzione del potere sociale ed hanno altresì un profondo valore emancipatorio[3].

Il Libro Bianco sull'Intelligenza Artificiale al servizio del cittadino, Versione 1.0 (marzo 2018), curato dalla *Task force* dell'Intelligenza Artificiale dell'Agenzia per l'Italia Digitale, al riguardo reca alcuni esempi interessanti di utilizzo di sistemi intelligenti in materia di istruzione e formazione: si pensi al sostegno dell'apprendimento di studenti aventi problematiche cognitive[4] o al contributo alla riduzione dei *gap* linguistici di chi proviene da altri Paesi. Per quanto concerne il mondo della disabilità, grazie alla capacità di completare i processi comunicativi sia scritti che orali, rendendoli pertanto più agevoli, gli algoritmi possono migliorare la qualità di vita di individui ipovedenti o affetti da malattie degenerative come la SLA.

Anche il loro impiego in processi decisionali e di controllo può avere risvolti nella direzione di una maggiore uguaglianza. Gli algoritmi, infatti, consentono una migliore e più mirata distribuzione delle prestazioni sociali[5]. Da questo punto di vista rileva anche il loro utilizzo per il contrasto alle frodi, che privano i servizi di risorse che invece andrebbero destinate ai soggetti più svantaggiati.

Tuttavia, anche in materia di algoritmi si disvela l'ambivalenza del progresso tecnologico[6].

Essi possono essere all'origine tanto di concentrazioni di potere e privilegi quanto di discriminazioni ed emarginazione, divenendo quindi causa e talora strumento di nuove schiavitù[7].

Agli algoritmi vengono delegati compiti e decisioni sempre più rilevanti e con sempre maggiore frequenza, il che determina una crescente

[3] La «forte valenza emancipatoria» degli algoritmi è sottolineata da G. RESTA, *Governare l'innovazione tecnologica: decisioni algoritmiche, diritti digitali e principio di uguaglianza*, *in Politica del diritto*, 2/2019, p. 200.

[4] Trattasi dei sistemi di *Computer Assisted Instruction* (CAI) e degli *Intelligent Tutoring Systems* (ITS).

[5] Per tale esempio v. G. RESTA, *Governare l'innovazione tecnologica: decisioni algoritmiche, diritti digitali e principio di uguaglianza*, cit., p. 218.

[6] In tema v. A. SIMONCINI, *Sovranità e potere nell'era digitale*, in T.E. Frosini, O. Pollicino, E. Apa e M. Bassini (curr.), *Diritti e libertà in Internet*, Milano, 2017, pp. 26 ss.

[7] La schiavitù è un fenomeno che affligge anche le società contemporanee. Tra gli elementi che ne connotano le manifestazioni odierne gli studiosi hanno individuato: la soggezione o, più enfaticamente, la «subordinazione dispotica» e l'esclusione, che si correla alla reificazione dell'individuo e a forme di discriminazione. In argomento v. A. CALORE, *Introduzione*, in A. Calore e P. De Cesari (curr.), *Schiavi. Passato e presente*, Torino, 2021, pp. 6-7. Con particolare riferimento alle nuove forme di schiavitù cui si sta assistendo in conseguenza della pervasività delle nuove tecnologie digitali v. R. BODEI, *Dominio e sottomissione. Schiavi, animali, macchine, Intelligenza Artificiale*, Bologna, 2019, in particolare pp. 297 ss. Sia infine consentito rinviare a V. MOLASCHI, *Algoritmi e nuove schiavitù* (28 luglio 2021), in *federalismi.it*, XVIII (2021).

«esternalizzazione di scelte umane» alle macchine[8]. Ciò è dovuto non solo alle loro capacità, ma anche a quello che è stato definito un "*anchoring effect*", ossia un effetto di ancoraggio, osservato da ricerche in tema di psicologia comportamentale, in virtù del quale gli algoritmi vengono ritenuti più affidabili degli esseri umani[9]. Tra le ragioni alla base di questa fiducia vi è la convinzione che gli algoritmi siano non solo più "bravi" ma anche oggettivi e neutrali e quindi, per quanto qui interessa, non inclini a pregiudizi fonti di discriminazione.

Sennonché tanto il mito dell'infallibilità quanto quello della neutralità algoritmica si sono rivelati fallaci[10]. Gli algoritmi "sbagliano"[11] e, come più in generale la tecnologia, sono anche una "costruzione sociale", che riflette visioni, interessi nonché dinamiche e assetti di potere[12].

Le operazioni di raccolta, analisi e correlazione dei dati che gli algoritmi compiono, sulla base di elementi quali rilevanza o similarità dei contenuti, non sono affatto neutrali. Il codice che li governa può essere il prodotto di ben precise dimensioni culturali o ricollegarsi a interessi economici o politici specifici. I risultati cui addivengono possono essere espressione di una certa immagine o "discorso" sulla società[13], il che può portare a cristallizzare dinamiche sociali, economiche e politiche che restituiscono una realtà alterata e talora discriminatoria.

Ecco, quindi, algoritmi aventi pregiudizi razziali come *Compas* (*Correctional Offender Management Profiling for Alternative Sanctions*), utilizzato dalla giustizia penale americana, su cui si avrà modo di tornare, o razzisti, antisemiti e misogini come Tay, *bot* di *Microsoft* impersonante un utente virtuale adolescente, avente pure simpatie naziste, peraltro silenziato

[8] Per tale espressione v. M. AIROLDI e D. GAMBETTA, *Sul mito della neutralità algoritmica*, in *The Lab's Quaterly*, XX, (2018), 4, p. 27.

[9] Questo aspetto è ben messo in evidenza da B. MARCHETTI, *La garanzia dello Human in the loop alla prova della decisione amministrativa algoritmica*, in *Biolaw journal*, 2/2021, p. 8. In tema, v. J.M. LOGG, J.A. MINSON e A. MOORE, *Algorithm appreciation: people prefer algorithmic to human judgment*, in *Organizational Behavior and Human Decision Processes*, CLI (2019), p. 90 ss.

[10] Per una critica a tale mito v. M. AIROLDI e D. GAMBETTA, *Sul mito della neutralità algoritmica*, cit., pp. 26 ss.

[11] È interessante, da questo punto di vista, l'esempio formulato da M.T. RIBEIRO, S. SINGH e C. GUESTRIN, *Why Should I Trust You?: Explaining the Predictions of Any Classifier*, arXiv:1602.04938, 2016, par. 6.4 . Gli autori ipotizzano l'elaborazione di uno strumento di classificazione in grado di distinguere foto di husky da foto di lupi. Se tutti i lupi sono rappresentati (inconsapevolmente o scientemente come nello studio) in presenza di neve e gli husky in assenza di questa, si ottiene un algoritmo che può classificare come lupo anche un husky se sullo sfondo c'è neve.

[12] A questa prospettiva è dedicato il numero monografico di *The Lab's Quaterly*, XX (2018), 4, *Gli algoritmi come costruzione sociale*, cit.

[13] Pone l'attenzione sull'algoritmo come "discorso", richiamandosi agli studi di Foucault, D. BEER, *The social power of algorithms*, in *Information, Communication and Society*, XX (2017), pp. 1 ss.

in tempi brevi a causa di tali impreviste derive[14]. Sono queste le c.d. *AI-driven discriminations*[15].

Proprio in ragione della vastità e gravità delle problematiche discriminatorie correlate agli algoritmi si è parlato di essi addirittura come di «armi di distruzione matematica» in grado di aumentare il livello di disuguaglianza nelle nostre società[16].

Casi di algoritmi discriminatori si sono verificati in molteplici ambiti: si pensi al settore creditizio, ove gli algoritmi "filtrano", attraverso i meccanismi di *credit scoring*, la concessione o meno di finanziamenti a individui e a imprese[17]; a quello assicurativo, in cui essi determinano condizioni e costi delle polizze; al mercato del lavoro, in cui la profilazione viene utilizzata per la valutazione del rendimento professionale, ai fini di assunzioni e licenziamenti.

Proprio in materia di lavoro è nota la vicenda dell'algoritmo impiegato dalla piattaforma *Deliveroo* per regolamentare la prenotazione dei turni di lavoro da parte dei *riders*. Il sistema di profilazione adottato, basato sui due parametri dell'affidabilità e della partecipazione, trattava nello stesso modo chi non cancellava nei termini la propria partecipazione alla sessione di lavoro prenotata per motivi futili e chi non lo faceva per l'adesione a uno sciopero o per malattia, disabilità (ivi compresa l'assistenza prestata a un portatore di handicap), cura di un minore malato, ecc. Ne derivava la discriminazione di chi non poteva rendersi disponibile per ragioni legittime, con conseguente emarginazione dal gruppo prioritario nella scelta dei turni e riduzione delle future opportunità di accesso al lavoro[18].

---

[14] Tay era un *bot*, vale a dire una personalità artificiale in grado di interagire sulla base di algoritmi di apprendimento. Il *software* di *Microsoft* avrebbe dovuto evolvere sulla base del patrimonio di informazioni acquisito dagli scambi con utenti reali. L'esperimento è stato "compromesso" dall'intervento di navigatori che hanno iniziato a "riempire la testa" di Tay con pregiudizi razziali, sessisti, negazionisti, ecc. Sulla "storia" di Tay v. le notizie reperibili sul sito dell'AGI, *Tay, utente virtuale Microsoft diventa ninfomane e nazista* (25 marzo 2016, 17:25), https://www.agi.it/lifestyle/tay_utente_virtuale_microsoft_diventa_ninfomane_e_nazista-643587/news/2016-03-25/.

[15] In argomento v. F.Z. BORGESIUS, *Discrimination, Artificial Intelligence and Algorithmic Decision-Making*, Council of Europe, Strasburgo, 2018, che fa una ricognizione di alcuni dei settori di attività pubblici e privati a maggior rischio di discriminazione.

[16] Il fatto che gli algoritmi, lungi dall'essere oggettivi, incorporino spesso pregiudizi o siano espressione della fallibilità di chi li ha ideati è il filo conduttore del volume di C. O'NEIL, *Armi di distruzione matematica. Come i Big Data aumentano la disuguaglianza e minacciano la democrazia*, tr. it., Milano, 2017.

[17] In tema v. F. MATTASSOGLIO, *Algoritmi e regolazione. Circa i limiti del principio di neutralità tecnologica*, *in Rivista della Regolazione dei mercati*, 2018, pp. 226 ss.

[18] La questione è stata decisa dal Tribunale di Bologna con una ordinanza del 31 dicembre 2020. Ne riferisce, tra gli altri, l'Osservatorio sullo Stato digitale dell'IRPA: v. N. CENTOFANTI, *Il caso Deliveroo: l'algoritmo FRANK e la discriminazione by Design* (11 febbraio 2021), in https://www.irpa.eu/il-caso-deliveroo-lalgoritmo-frank-e-la-discriminazione-by-design/ (ultimo accesso: 3 giugno 2022).

In questo contributo maggiore attenzione verrà dedicata ai casi di algoritmi discriminatori utilizzati in seno a poteri ed apparati dello Stato, che più di qualsiasi altro soggetto dovrebbero improntare la propria azione allo «sviluppo della persona umana», per riprendere l'art. 3 Cost., ai principi di uguaglianza e imparzialità e alla tutela dei diritti degli individui.

Una volta esemplificata la questione, verrà formulata qualche considerazione sul "come" gli algoritmi discriminatori vengono generati, per quanto i meccanismi non siano del tutto perspicui e tale mancanza di chiarezza sia essa stessa parte del problema.

Infine, lo scritto si concentrerà sulle prime risposte o tentativi di risposta alla questione delle discriminazioni algoritmiche dati da parte del diritto.

Esula dalla trattazione un altro aspetto della discriminazione, che qui si può solamente accennare: l'esclusione di coloro che non hanno accesso o non sono in grado di usare consapevolmente le tecnologie su cui si basano le infrastrutture digitali da cui dipende ormai quasi ogni aspetto della nostra esistenza. É questa la questione del c.d. *digital divide*, che può dipendere da molteplici fattori - geografici, economici, di genere, di cultura, di religione, di lingua e generazionali - spesso reciprocamente influenzantisi.

Al riguardo si può peraltro osservare come sussista una tragica circolarità: la sussistenza di situazioni di svantaggio è alla base del *digital divide*, il quale a sua volta peggiora le disparità esistenti[19]. Si pensi al caso di soggetti anziani e poveri (spesso, peraltro, le due situazioni coincidono) o di ragazzi che vivono in contesti disagiati: età, fragilità, mancanza di risorse sono alla base dell'impossibilità o difficoltà di accesso al digitale e a tutta una serie di servizi *on line*, il che si riverbera negativamente sulla condizione della persona aggravandone le problematiche[20].

Occorre infine considerare anche il caso di coloro che non aderiscono al modello di società digitale imperante, fatto di informazione, spazi virtuali, interconnessione e condivisione, che nella sua dimensione totalizzante incide

[19] Spunti in tal senso possono leggersi in E. Mª MENÉNDEZ SEBASTIÁN, JAVIER BALLINA DÍAZ, *Digital Citizenship: Fighting the Digital Divide*, in, 2/2021, Issue 1, p. 155.

[20] Un esempio particolarmente interessante è rappresentato dal fascicolo sanitario elettronico, strumento su cui v., *ex multis*, N. POSTERARO, *La digitalizzazione della sanità in Italia: uno sguardo al Fascicolo Sanitario Elettronico (anche alla luce del Piano Nazionale di Ripresa e Resilienza)*, in *federalismi.it*, Osservatorio di Diritto sanitario, 17 novembre 2021. Innovazione di grande importanza per la tutela del diritto alla salute, rischia di generare forme di esclusione proprio di quelle fasce di popolazione, maggiormente afflitte da preoccupazioni sanitarie (si pensi agli anziani o ai disabili), che dovrebbero invece beneficiarne. Emblematico è inoltre il caso degli studenti appartenenti a famiglie a basso reddito spesso rimasti esclusi dalla didattica a distanza (DAD), resa necessaria dalla pandemia, malgrado le risorse pubbliche stanziate. Il rapporto BES Istat 2021 ha evidenziato come l'8% degli alunni provenienti da famiglie svantaggiate sia rimasto escluso dalla DAD durante la pandemia Covid-19. Il dato sale al 23% se si considerano gli studenti che soffrono di disabilità.

sulla libertà di tutti. Anche in questo caso vi sono conseguenze in termini di esclusione e discriminazione[21].

2. Alcuni esempi: quando a discriminare è lo Stato…

Gli algoritmi sono sempre più utilizzati anche nell'ambito di processi decisionali pubblici[22].

La convinzione circa l'infallibilità e l'oggettività degli algoritmi, considerati scevri di quei pregiudizi che invece compromettono le decisioni umane, ha portato ad applicarli in ambiti estremamente delicati e critici, dominati da forti istanze di imparzialità, come ad esempio quello giudiziario[23]. Sennonché, proprio in tale settore si sono avuti esiti assai discutibili.

Tra gli esempi più noti e commentati di algoritmi discriminatori deve essere senz'altro ricordato il caso *Compas* [24]. *Compas* è un sistema algoritmico di predizione del rischio di recidiva, utilizzato dalla giustizia penale americana, in particolare nello stato del Wisconsin. Le informazioni fornite dal *software* sono funzionali alle decisioni su entità e modalità di esecuzione delle sanzioni penali, soprattutto per quanto riguarda la possibilità di accesso a misure alternative alla detenzione.

Lo strumento è stato elaborato da una società privata, il che ha implicazioni dal punto di vista della trasparenza.

Il rischio di recidiva è valutato sulla base di una serie di elementi quali i precedenti giudiziari del soggetto, vari dati statistici, la somministrazione di un questionario ed altri fattori non conosciuti in quanto coperti da proprietà intellettuale della società. Per questa stessa ragione non è noto come i diversi elementi siano ponderati e in che maniera i tassi di pericolosità sociale siano calcolati.

---

[21] Richiamando un esempio particolarmente studiato, come può "sopravvivere" o comunque qual è la qualità di vita in una *smart city* di chi non è abbastanza *smart*? Sui rischi di esclusione di chi non aderisce al modello di condivisione che è alla base del fenomeno delle *smart cities* v. F. FRACCHIA, P. PANTALONE, *Smart City: condividere per innovare (e con il rischio di escludere?)* (25 novembre 2015), in *federalismi.it*, XXII (/2015), in particolare pp. 23 ss.

[22] Per una panoramica dei settori di utilizzo degli algoritmi da parte delle pubbliche amministrazioni v. G. AVANZINI, *Decisioni amministrative e algoritmi informatici. Predeterminazione, analisi predittiva e nuove forme di intellegibilità*, Napoli, 2019, 35 ss.

[23] Sull'impatto dell'intelligenza artificiale sulla giustizia, quale ambito indicativo, assieme al lavoro, delle problematiche sociali sollevate dalle nuove tecnologie digitali v., *ex multis*, A. VENANZONI, *La valle del perturbante: il costituzionalismo alla prova delle intelligenze artificiali e della robotica*, in *Politica del diritto*, 2/2019, in particolare 245 ss.

[24] Si tratta del caso 'State of Wisconsin v. Eric Loomis', n. 2015AP157-CR, deciso con sentenza del 13 luglio 2016. Per un commento v. HAN-WEI, L. CHING-FU e C. YU-JIE, *Beyond State v. Loomis: artificial intelligence, government, algorithmization, and accountability*, in *International Journal of Law and Information Technology*, 27, 2019, 122 ss. Si v. altresì S. CARRER, *Se l'*amicus curiae *è un algoritmo: il chiacchierato caso Loomis alla Corte Suprema del Wisconsin*, in *Giurisprudenza Penale Web*, 4/2019.

La Corte Suprema del Wisconsin ha ritenuto che il ricorso a *Compas*, se condotto in modo appropriato, rispettando specifici limiti e accortezze indicati dalla Corte stessa, non violi il diritto ad un giusto processo. Le ombre emerse in relazione al suo utilizzo, però, non sono affatto marginali.

Successivi studi volti a verificare la correttezza delle sue previsioni hanno evidenziato la classificazione di criminali appartenenti a minoranze, in specie di colore, come ad alto rischio di recidiva secondo percentuali sproporzionatamente maggiori rispetto a delinquenti bianchi, che il sistema tende invece a indicare come a basso rischio[25]. Se ne deduce una chiara correlazione del rischio di recidiva a fattori quali la razza, il che spiega la sovrastima della pericolosità sociale dei soggetti neri, soprattutto nelle ipotesi di recidiva violenta, e la sottovalutazione di quella dei bianchi.

Sempre in materia di amministrazione della giustizia, problematiche analoghe si sono manifestate in relazione ad un altro *software*, il c.d. P.S.A, ossia il *Public Safety Assessment*, utilizzato dalle Corti di diversi Stati americani per determinare il periodo di custodia cautelare degli imputati o l'importo della cauzione. I giudici tendono ad adeguarsi a punteggi stabiliti dal sistema informatico sulla base di una banca dati contenente casi simili, precedenti giudiziari e dati anagrafici dei soggetti coinvolti.

La consegna della pace sociale agli algoritmi si riscontra anche nel campo della prevenzione dei reati, ove sono sempre più frequenti gli esperimenti di "polizia predittiva", con cui si cerca di identificare le zone cittadine e gli orari della giornata in cui si possono verificare più crimini affinché le forze dell'ordine possano intervenire anticipatamente o comunque tempestivamente[26]. Gli algoritmi operano come i *Precog* del film *Minority Report* (2002), liberamente tratto dall'omonimo racconto di fantascienza di Philip K. Dick.

Tali tecniche, oltre a porre non poche questioni di compatibilità con la protezione dei dati personali e la tutela della *privacy*[27], possono degenerare in esiti discriminatori, prendendo di mira soggetti e quartieri svantaggiati ma

[25] V. l'analisi di J. LARSON, S. MATTU, L. KIRCHNER e J. ANGWIN, *How We Analyzed the COMPAS Recidivism Algorithm* (23 maggio 2016), in https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm. Lo studio ha preso in considerazione più di 10.000 criminali nell'area di Broward County in Florida, comparando il tasso di recidiva previsto con quello effettivamente riscontrato dopo due anni.

[26] Quale esperimento di 'polizia predittiva' in Italia v. *eSecurity.Trento*, realizzato nell'ambito del progetto europeo *eSecurity – ICT for knowledge-based and predictive urban security*, coordinato dal gruppo di ricerca *eCrime* del Dipartimento di Giurisprudenza dall'Università degli Studi di Trento con il coinvolgimento di Questura di Trento, Centro ICT della Fondazione Bruno Kessler, Comune di Trento: http://www.esecurity.trento.it/.

[27] Per un'analisi di tali questioni v. A. BONFANTI, *'Big data' e polizia predittiva: riflessioni in tema di protezione del diritto alla privacy e dei dati personali* (24 ottobre 2018), in *mediaLAWS*, 3/2018, pp. 206 ss.

non necessariamente criminogeni o comunque nei quali l'intervento pubblico dovrebbe essere ben altro rispetto a quello di polizia.

Si immagini peraltro cosa potrebbe succedere se giustizia ed apparati repressivi dello Stato facessero propria una recente ricerca, molto controversa, secondo cui si può distinguere tra criminali e non con il 90% di probabilità attraverso l'analisi automatica di immagini di volti umani[28]. Ci si troverebbe ad assistere alla "rivincita algoritmica" di Lombroso.

Spesso, come anticipato nelle premesse di questo lavoro, gli algoritmi vengono utilizzati dalle pubbliche amministrazioni per l'individuazione dei beneficiari di emolumenti e servizi e per scovare eventuali truffe: anche in questo campo si sono verificati non pochi episodi di algoritmi discriminatori.

Un caso piuttosto noto è avvenuto in Olanda, ove nel 2014 il Ministero per gli Affari Sociali e il Lavoro aveva ideato SyRI (*System Risk Indicator*), un sistema di controllo basato su algoritmi finalizzati a prevedere quali individui fossero più inclini a commettere frodi in campo assistenziale[29]. SyRI, che incrociava dati provenienti da variegate fonti, è stato utilizzato specificamente in quartieri con popolazione a basso reddito in buona parte composta da minoranze.

Nel 2018 lo Stato olandese è stato citato in giudizio da una vasta coalizione di organizzazioni impegnate nella tutela dei diritti civili e sociali e nel 2020 la Corte distrettuale dell'Aja ha messo la parola fine all'uso dell'algoritmo, ravvisando nel ricorso a tale strumento una violazione di norme a protezione dei diritti umani[30] e dello stesso Regolamento UE 2016/679 per la protezione dei dati, c.d. GDPR.

Gli aspetti che avevano destato preoccupazione sono non pochi: la circostanza che il sistema avesse preso di mira i quartieri disagiati, dove oltretutto vive un maggior numero di persone di origine straniera, rappresentava una discriminazione basata sulle condizioni socio-economiche e l'etnia delle persone. Evidente era inoltre il *vulnus* alla tutela della *privacy*, in danno soprattutto degli abitanti più poveri.

Si segnalano peraltro episodi in cui la discriminazione che colpisce i soggetti più svantaggiati è involontaria, costituendo quasi una eterogenesi dei fini. Un esempio studiato è quello di *Street Bumb*, esperimento condotto a Boston nel 2011, che aveva utilizzato l'intelligenza artificiale per raccogliere

[28] Si tratta del controverso lavoro di X. WU, X. ZHANG, *Automated inference on criminality using face images* (26 maggio 2017), in https://arxiv.org/abs/1611.04135v1.

[29] Ne riferisce il Centro di Ateneo per i Diritti Umani 'Antonio Papisca' dell'Università degli Studi di Padova: *Corte Distrettuale dell'Aja: sentenza storica sullo spionaggio digitale dei poveri olandesi*, in https://unipd-centrodirittiumani.it/it/news/Corte-Distrettuale-dellAja-sentenza-storica-sullo-spionaggio-digitale-dei-poveri-olandesi/5088.

[30] Si v. al riguardo le problematiche evidenziate da P. ALSTON, *Special Rapporteur on extreme poverty and human rights, brief as amicus curiae before the District Court of the Hague on the case of NJCM c.s./De Staat der Nederlanden (SyRI)*, case No. C/09/550982/ HA ZA 18/388, September 2019.

informazioni sul manto stradale delle varie zone cittadine attraverso le segnalazioni fatte tramite *smartphone* degli abitanti[31]. Nei quartieri più a basso reddito, in cui le condizioni delle strade erano decisamente più deteriorate, vennero fatte meno segnalazioni, in quanto la popolazione ivi residente disponeva di un minor numero di cellulari. Paradossalmente, quindi, il sistema favorì gli interventi di manutenzione nei quartieri "ricchi" ove la condizione del manto stradale era invece migliore.

Una delle più discusse frontiere dell'utilizzo degli algoritmi, in cui maggiore è il rischio di derive discriminatorie, è rappresentato dal c.d. *social credit scoring*, attuato in Cina[32]. Si tratta di un sistema che valuta l'affidabilità sociale dei cittadini sulla base di caratteristiche della personalità, dei comportamenti tenuti, della rete di rapporti o di altri elementi, talora conosciuti talora soltanto previsti, cui è correlato un punteggio sociale. Tale punteggio determina e modula l'accesso a contributi economici o servizi.

Il Consiglio di Stato cinese ravvisa nel *social credit scoring* una componente essenziale dell'economia di mercato socialista e della *governance* sociale: esso, infatti, contribuisce a sviluppare una «cultura della sincerità», a incoraggiare la «fiducia» e a costruire una «armoniosa società socialista»[33].

Non si può non vedere come si tratti di una sorveglianza penetrante e invasiva, che può toccare ogni aspetto della vita umana, ridotto ad un *rating*, con conseguenti pesanti limitazioni della libertà degli individui. Per quanto qui rileva, tale sistema può essere all'origine di forme di discriminazione ed esclusione di chi rifiuta o si allontana dai canoni e dagli *standard* imposti dagli interessi dello Stato.

3. Anatomia della discriminazione algoritmica.

Il termine per indicare i c.d. algoritmi discriminatori è *bias*. Tali "pregiudizi" caratterizzano sistemi informatici che discriminano sistematicamente e ingiustamente certi individui o gruppi di individui in favore di altri, negando

[31] Il caso è riferito da F.Z. BORGESIUS, *Discrimination, Artificial Intelligence and Algorithmic Decision-Making*, cit. p. 19.

[32] Per approfondimenti v. R. BOTSMAN, *Big data meets Big Brother as China moves to rate its citizens* (27 ottobre 2017), consultabile alla pagina www.wired.co.uk. In argomento v. altresì F. COSTANTINI, G. FRANCO, *Decisione automatizzata, dati personali e pubblica amministrazione in Europa: verso un "Social credit system"?*, in *Istituzioni del Federalismo*, 2019, pp. 715 ss.

[33] Si v. il documento del Consiglio di Stato cinese «Planning Outline for the Construction of a Social Credit System (2014-2020)» (2014), la cui traduzione può essere letta alla pagina https://chinacopyrightandmedia.wordpress.com/2014/06/14/planning-outline-for-the-construction-of-a-social-credit-system-2014-2020/ (ultimo accesso: 3 giugno 2022).

opportunità o beni ovvero attribuendo un risultato indesiderato sulla base di motivazioni irragionevoli o inappropriate[34].

Le discriminazioni si verificano soprattutto nel caso di processi algoritmici correlati ad attività di profilazione, ove per profilazione si intende, ai sensi del GDPR, «qualsiasi forma di trattamento automatizzato di dati personali consistente nell'utilizzo di tali dati personali per valutare determinati aspetti personali relativi a una persona fisica, in particolare per analizzare o prevedere aspetti riguardanti il rendimento professionale, la situazione economica, la salute, le preferenze personali, gli interessi, l'affidabilità, il comportamento, l'ubicazione o gli spostamenti di detta persona» (art. 4, comma 4).

I semi della discriminazione possono annidarsi in vari momenti delle procedure algoritmiche[35].

La programmazione e la selezione delle informazioni rilevanti sono le fasi in cui tipicamente possono generarsi discriminazioni, in quanto si crea il *set* di dati su cui "lavorerà" l'algoritmo e gli si insegna a creare correlazioni nell'ambito di essi, correlazioni che possono dare origine, come si è visto, ad effetti distorsivi.

Questi ultimi possono essere anzitutto dovuti al modo di ragionare, ai valori o proprio ai pregiudizi del programmatore, che possono dipendere sia dai suoi preconcetti personali quanto dalla stessa organizzazione di cui fa parte. Lo sviluppatore può cioè condizionare lo schema dell'algoritmo, selezionando, vale a dire includendo o escludendo, caratteristiche tipiche di una categoria di persone a rischio di discriminazione. Si pensi a caratteri quali la razza, il colore della pelle ovvero il sesso, la lingua, la religione, le opinioni politiche, ecc.: sono queste "categorie algoritmiche sospette", che possono dare luogo a classificazioni discriminatorie, inficiando i circuiti decisionali dell'intelligenza artificiale e delle organizzazioni che ne fanno uso[36].

I dati di partenza sono determinanti nella generazione di disparità e forme di esclusione anche a prescindere da uno specifico intervento umano rivolto o meno scientemente al raggiungimento di un risultato discriminatorio. Si

[34] Si v. la definizione di B. FRIEDMAN, H. NISSEMBAUM, *Bias in Computer Systems*, *ACM Transactions on Information Systems*, 3/1996, 332*: «*Accordingly, we use the term bias to refer to computer systems that systematically and unfairly discriminate against certain individuals or groups of individuals in favor of others. A system discriminates unfairly if it denies an opportunity or a good or if it assigns an undesirable outcome to an individual or group of individuals on grounds that are unreasonable or inappropriate».

[35] Per un'analisi delle fasi delle decisioni algoritmiche in cui possono porsi le premesse per le c.d. *AI-driven discriminations* v. P. ZUDDAS, *Intelligenza artificiale e discriminazioni*, Liber amicorum per Pasquale Costanzo, 16 marzo 2020, in particolare pp. 5 ss.

[36] In argomento, con particolare riferimento all'uso di algoritmi da parte delle amministrazioni pubbliche, v. D.U. GALETTA, J.G. CORVALÁN, *Intelligenza Artificiale per una Pubblica Amministrazione 4.0? Potenzialità, rischi e sfide della rivoluzione tecnologica in atto* (6 febbraio 2019), in *federalismi.it*, pp. 21-22.

consideri il caso di dati obsoleti, inesatti o anche incompleti. È questo il principio del c.d. *garbage in – garbage out*.

Il citato esempio di *Street Bump* dimostra quanto sia rilevante il metodo di raccolta dei dati e le conseguenze paradossali che possono verificarsi a fronte di una raffigurazione parziale della realtà, in cui vi sono gruppi sociali od esigenze sovra-rappresentate o sotto-rappresentate[37].

L'insieme dei dati su cui opera l'algoritmo può essere più o meno definito, a seconda che sia "chiuso", come nel caso di archivi digitali e banche dati *off line* ovvero "aperto", come nell'ipotesi di dati tratti da *internet*.

Non necessariamente i set di dati chiusi sono più immuni alle discriminazioni. Si pensi al caso *Compas*, in cui l'algoritmo "ragiona" muovendo da una serie di elementi tra cui quelli contenuti in archivi giudiziari. In ipotesi di questo genere, anzi, c'è il rischio che i pregiudizi precedenti si cristallizzino e perpetuino.

Sicuramente, comunque, l'acquisizione di informazioni dalla rete, anche se può essere limitata a siti e piattaforme, è ancor meno controllata, fino ad arrivare a trasfondere, come efficacemente affermato, i pregiudizi diffusi nella società globale[38].

Quelli sin qui evidenziati sono casi di pregiudizi digitali "derivati" ma ci sono anche pregiudizi che potrebbero definirsi, pur con qualche accortezza, "autonomi"[39].

Le discriminazioni possono essere generate dai meccanismi di apprendimento dell'intelligenza artificiale: è lo stesso algoritmo che crea correlazioni tra caratteristiche individuali anche apparentemente non rilevanti e porta a categorizzazioni in profili falsati e discriminatori.

È proprio nel contesto del c.d. *machine learning* che possono aversi più frequentemente episodi di discriminazione per associazione o *proxy discrimination*. Sono forme di discriminazione particolarmente subdole in quanto non derivanti da fattori di discriminazione tradizionale o comunque facilmente riconoscibili. La riconducibilità a un gruppo o categoria a rischio di discriminazione avviene infatti sulla base di altri dati − diversi, tanto per fare un esempio, da razza, etnia, sesso o genere − che però possono risultare egualmente associati a quello stesso insieme di individui. Si pensi al quartiere di residenza o al codice postale, ma anche a preferenze riguardanti siti *internet*, film e serie televisive, ecc.

Naturalmente questo tipo di discriminazioni deflagra quando già il *set* di dati di partenza presenta delle criticità, come nel caso dell'incorporazione di

[37] «*Biased training data can lead to biased AI systems*»: F.Z. BORGESIUS, *Discrimination, Artificial Intelligence and Algorithmic Decision-Making*, cit., p. 19.

[38] In questi termini P. ZUDDAS, *Intelligenza artificiale e discriminazioni*, cit., p. 5.

[39] Per la distinzione tra pregiudizi digitali "derivati" ed "autonomi" v., ancora, la classificazione proposta da P. ZUDDAS, *Intelligenza artificiale e discriminazioni*, cit., p. 5.

decisioni precedenti a loro volta contrassegnate da pregiudizi o della presenza di dati di scarsa qualità, non aggiornati, parziali, ecc.

Si tenga altresì presente che, nell'ambito del *machine learning*, artefice di rappresentazioni ed esiti distorsivi può essere la stessa *black box*[40]. È questo uno dei paradossi delle tecnologie digitali: gli algoritmi, che dovrebbero assicurare trasparenza e intelligibilità dei processi logici, presentano non poche "zone oscure", imperscrutabili anche per gli stessi programmatori. Una discriminazione, quindi, "fuori controllo".

4. La risposta del diritto. Alcune considerazioni sui principi in materia di decisioni algoritmiche del GDPR e sulla proposta di regolamento UE in materia di intelligenza artificiale.

Oggigiorno la tecnologia non è più soltanto strumento per realizzare quanto deciso dai soggetti agenti umani, ma è diventata essa stessa fautrice di decisioni rilevanti dal punto di vista della dignità e dei diritti degli individui[41].

Gli algoritmi impattano su molteplici dimensioni dell'uguaglianza, sia promuovendola che causando o consolidando discriminazioni ed esclusione, rispetto alle quali si rivela necessaria un'adeguata risposta giuridica: «*In sum, to be fair and equal, the algorithms must be regulated*»[42].

Nell'utilizzo di essi non si può quindi prescindere dal quadro generale di tutela dei diritti fondamentali sancito a livello europeo e interno[43].

Per quanto riguarda l'UE vengono in considerazione la Carta dei diritti fondamentali dell'Unione Europea, che dedica un intero capo all'uguaglianza (III) nonché alla dignità (I) e la Convenzione europea dei diritti dell'uomo, divenuta anch'essa fonte primaria dell'UE, di cui pure il principio di uguaglianza è uno dei pilastri (art. 14).

Non si può inoltre non rimarcare il ruolo guida che può giocare la nostra Costituzione: più autori hanno sottolineato come gli sviluppi dell'intelligenza artificiale debbano essere costituzionalmente orientati[44]; la strada da

[40] In tema, v. F. PASQUALE, *The black box society*, Cambridge Mass., 2015.

[41] Sul sovvertimento della relazione tra «agente» e «strumento» che caratterizza il rapporto tra uomini e algoritmi v. A. SIMONCINI, *L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà*, in *Biolaw Journal*, 1/2019, pp. 67 ss. In argomento cfr. V. MOLASCHI, *Algoritmi e nuove schiavitù*, cit., *passim.*

[42] Così G. DE MINICO, *Towards an "Algorithm Constitutional by Design"*, in *Biolaw Journal*, 1/2021, p. 402.

[43] In generale, sulla necessità di regolamentare l'intelligenza artificiale alla luce del quadro dei diritti fondamentali v. A. MANTELERO, *Regulating AI within the Human Rights Framework: A Roadmapping Methodology*, in P. Czech, L. Heschl, K. Lukas, M. Nowak e G. Oberleitner (curr.), *European Yearbook on Human Rights 2020*, Cambridge UK, 2020, pp. 478 ss.

[44] La configurazione di un'intelligenza artificiale «costituzionalmente orientata» è stata oggetto di approfondite riflessioni da parte di C. CASONATO, *Per un'intelligenza artificiale costituzionalmente orientata*, cit., *passim*, che ha studiato le potenzialità della Costituzione nel plasmare la disciplina dell'intelligenza artificiale. Sui rapporti tra intelligenza artificiale

perseguire, quindi, è quella di algoritmi "costituzionali *by design*", riprendendo un'efficace espressione utilizzata in dottrina[45].

Si pensi, al riguardo, al principio personalista (artt. 2 e 3 Cost.), che pone la persona e la sua dignità al centro dell'architettura costituzionale e quindi dell'ordinamento, o alla consacrazione dello stesso principio di uguaglianza, che la Carta costituzionale declina nella sua dimensione formale e sostanziale (art. 3 Cost.).

Quando a far uso degli algoritmi è la pubblica amministrazione, bisogna guardare ai principi dell'art. 97 Cost., tradotti nella l. 241/1990 sul procedimento amministrativo, di cui la giurisprudenza amministrativa ha offerto una lettura alla luce delle statuizioni sulle decisioni algoritmiche del GDPR[46], che si avrà modo di illustrare nel prosieguo della trattazione.

In questa sede si formulerà qualche riflessione sulle forme di tutela che possono rinvenirsi a livello di *data protection*, ma occorre tenere presente che la risposta giuridica non può prescindere dal diritto antidiscriminatorio, da declinare in chiave digitale[47].

Vi è anzitutto un livello di intervento, c.d. «etico-collettivo»[48], che si esprime, tra gli altri, in documenti di *soft law*[49]: in prospettiva europea si possono citare, senza pretesa di esaustività, la risoluzione del Parlamento europeo sui principi etici dell'IA, della robotica e della tecnologia correlata (2020)[50], che ha giocato un ruolo rilevante nell'elaborazione, da parte della Commissione europea, della proposta di regolamento UE in materia di

---

e diritto costituzionale v. inoltre A. SIMONCINI, S. SUWEIS, *Il cambio di paradigma nell'intelligenza artificiale e il suo impatto sul diritto costituzionale,* in *Rivista di filosofia del diritto*, 1/2019, pp. 87 ss.

[45] Sul tema v., ancora, G. DE MINICO, *Towards an "Algorithm Constitutional by Design"*, cit., in particolare pp. 398 ss.

[46] Ci si riferisce a Cons. Stato, 13 dicembre 2019, n. 8472, 8473, 8474 e 4 febbraio 2020, n. 881. La bibliografia in materia è estremamente ampia: *ex multis*, v. R. FERRARA, *Il giudice amministrativo e gli algoritmi. Note estemporanee a margine di un recente dibattito giurisprudenziale*, in *Dir. amm*., 2019, pp. 773 ss. Sui rapporti tra pubbliche amministrazioni e tecnologie digitali v., per tutti, R. CAVALLO PERIN, D.U. GALETTA (curr.), *Il diritto dell'amministrazione pubblica digitale*, Torino, 2020, di cui v., per i profili toccati in questo scritto, il contributo di A. SIMONCINI, *Amministrazione digitale algoritmica. Il quadro costituzionale*, pp. 1 ss.

[47] In tal senso v. G. RESTA, *Governare l'innovazione tecnologica: decisioni algoritmiche, diritti digitali e principio di uguaglianza*, cit., p. 236.

[48] Si v., al riguardo, A. SIMONCINI, *Sovranità e potere nell'era digitale*, cit., pp. 34 ss., il quale precisa che la risposta al potere digitale si manifesta in una stratificazione di interventi, di diversa tipologia e articolati in più livelli: un livello «morale-soggettivo», che per lo più sfugge al diritto, un livello «etico-collettivo» ed, infine, uno «giuridico-pubblicistico».

[49] Ne fanno parte anche la regolamentazione privata e l'autoregolamentazione: A. SIMONCINI, *Sovranità e potere nell'era digitale*, cit., p. 35.

[50] Si tratta della risoluzione del Parlamento europeo del 20 ottobre 2020 recante raccomandazioni alla Commissione concernenti il quadro relativo agli aspetti etici dell'intelligenza artificiale, della robotica e delle tecnologie correlate (2020/2012(INL)). La risoluzione dedica un capitolo apposito al tema della «Non distorsione e non discriminazione».

intelligenza artificiale[51] e gli Orientamenti etici per un'IA affidabile (2019)[52], di cui pure tale proposta ha beneficiato. Recentemente, inoltre, la Commissione ha proposto una «Dichiarazione europea sui diritti e i principi digitali per il decennio digitale»[53], destinata ad essere condivisa anche dal Parlamento europeo e dal Consiglio. La Dichiarazione ha quale obiettivo la promozione di una transizione digitale che «metta al centro le persone» ed è particolarmente sensibile ai valori della solidarietà e dell'inclusione[54].

Per quanto concerne specificamente l'Italia, si possono ricordare il Libro Bianco sull'Intelligenza Artificiale al servizio del cittadino, Versione 1.0. (marzo 2018) [55], che presta ai temi in esame un'attenzione maggiore e alcuni spunti, che forse avrebbero meritato di essere più ampiamenti sviluppati, presenti nella Strategia Nazionale per l'Intelligenza Artificiale (2020)[56] e nel successivo Programma Strategico Intelligenza Artificiale 2022-2024[57].

Pur senza sottacere l'importanza dei documenti citati, nelle righe seguenti ci si concentrerà sulle risposte di tipo «giuridico-pubblicistico», aventi carattere

[51] Con tale risoluzione il Parlamento europeo ha chiesto alla Commissione, in conformità alla procedura di cui all'art. 225 del Trattato sul funzionamento dell'Unione europea, di presentare una proposta di regolamento sui principi etici relativi allo sviluppo, alla diffusione e all'utilizzo dell'intelligenza artificiale, della robotica e delle tecnologie correlate, a norma dell'art. 114 del medesimo Trattato sul ravvicinamento delle legislazioni dell'UE.

[52] Gli Orientamenti etici per un'IA affidabile sono stati elaborati da un gruppo indipendente di esperti ad alto livello sull'intelligenza artificiale istituito dalla Commissione europea nel giugno 2018. Tra i principi etici che concorrono a garantire l'affidabilità dell'intelligenza artificiale vi è il principio di equità (n. 52). In argomento si v. altresì i riferimenti in tema di «Legalità dell'IA» (n. 23) e quelli nelle parti del documento dedicate a «I diritti fondamentali come base per un'IA affidabile» (n. 44), «Requisiti per un'IA affidabile» (n. 58-5), «Diversità, non discriminazione ed equità» (n. 80).

[53] V. le due Comunicazioni della Commissione: COM(2022) 27 final, «Comunicazione della Commissione al Parlamento Europeo, al Consiglio, al Comitato Economico e Sociale Europeo e al Comitato delle Regioni relativa alla definizione di una Dichiarazione europea sui diritti e i principi digitali»; COM(2022) 28 *final*, recante la «Dichiarazione europea sui diritti e i principi digitali per il decennio digitale» vera e propria.

[54] V. il cap. II: «Solidarietà e inclusione».

[55] «Prevenire le disuguaglianze» è una delle «sfide dell'IA al servizio del cittadino» (Sfida n. 7).

[56] La Strategia individua alcuni «Principi guida» che debbono governare il cambiamento del paradigma tecnologico. In particolare, nella parte dedicata ai principi dell'antropocentrismo, dell'affidabilità e della sostenibilità, precisa che «L'IA deve essere al servizio delle persone, garantendo una supervisione umana, prevenendo i rischi di inasprimento degli squilibri sociali e territoriali potenzialmente derivanti da un suo utilizzo inconsapevole o inappropriato». Il documento sottolinea inoltre come sia necessario «stabilire un insieme di regole, da sviluppare in ambito europeo, che garantiscano un IA a misura di cittadino - minimizzando rischi di discriminazione o di errore» (par. 3).

[57] Il Programma Strategico riprende l'approccio della Strategia Nazionale per l'Intelligenza Artificiale, che punta a configurare l'intelligenza artificiale italiana come «antropocentrica, affidabile e sostenibile» (par. 2.2., relativo ai «Principi guida»). Per quanto qui interessa, rileva la promozione di un'idea di intelligenza artificiale incentrata anche sull'inclusione economica e sociale.

vincolante[58]. Vengono anzitutto in considerazione le previsioni del GDPR e può essere di interesse soffermarsi sul futuro regolamento UE in materia di intelligenza artificiale, benché si tratti di un testo ancora in fase di discussione[59]. La proposta della Commissione europea si muove proprio nella direzione di promuovere un utilizzo dell'intelligenza artificiale nel rispetto dei diritti fondamentali dei cittadini dell'UE, oltre che di elevati standard di sicurezza.

Il GDPR reca i principi fondamentali che debbono guidare le decisioni algoritmiche, tanto dei soggetti privati, quanto di quelli pubblici. Vale la pena riportarne sinteticamente i contenuti in quanto tali principi presentano interazioni da valorizzare nell'ottica di assicurare una miglior tutela all'individuo anche dal punto di vista della lotta alle disuguaglianze.

In primo luogo vi è il principio di non esclusività della decisione algoritmica (art. 22, comma 1), espressione giuridica del c.d. *Human-in-the loop*, modello che, nell'ambito della *computer science*, richiede un'interazione uomo-macchina[60]. Ai sensi di tale principio qualora una decisione algoritmica produca effetti giuridici o incida significativamente sulla persona, l'interessato ha il diritto che questa non sia basata unicamente su un trattamento automatizzato, ivi compresa la profilazione; deve comunque esserci un intervento umano[61].

Altro principio importante è quello di conoscibilità: ognuno ha il diritto di conoscere l'esistenza di processi decisionali automatizzati, ancora una volta compresa la profilazione (art. 22, commi 1 e 4), che lo riguardino (art. 15, comma 1, lett. h)[62]. Il diritto alla conoscenza comprende anche l'importanza e le conseguenze del trattamento per l'interessato.

Il principio di conoscibilità è completato dal principio di comprensibilità, senza il quale il primo sarebbe in tutto o in parte vanificato: l'interessato ha anche il diritto di ottenere «informazioni significative sulla logica utilizzata» dalle procedure in questione, così come precisato dallo stesso GDPR (art. 15, comma 1, lett. h).

---

[58] È questo il terzo «ordine normativo» di intervento secondo A. SIMONCINI, *Sovranità e potere nell'era digitale*, cit., p. 35.

[59] Il testo è infatti oggetto di dibattito in seno al Parlamento europeo.

[60] Per approfondimenti su tale principio v. B. MARCHETTI, *La garanzia dello Human in the loop alla prova della decisione amministrativa algoritmica*, cit.

[61] La disposizione, peraltro, prevede al comma 2 alcune eccezioni: si tratta dei casi in cui la decisione: a) sia necessaria per la conclusione o l'esecuzione di un contratto tra l'interessato e un titolare del trattamento; b) sia autorizzata dal diritto dell'Unione o dello Stato membro cui è soggetto il titolare del trattamento, che precisa altresì misure adeguate a tutela dei diritti, delle libertà e dei legittimi interessi dell'interessato; c) si basi sul consenso esplicito dell'interessato.

[62] L'articolo in esame regolamenta il diritto di accesso dell'interessato, vale a dire il diritto di ottenere dal titolare del trattamento la conferma che sia o meno in corso un trattamento di dati personali che lo riguardano e, in tal caso, di ottenere l'accesso ai dati personali. Il diritto di accesso si estende a una serie di informazioni tra le quali rientra appunto l'esistenza di processi automatizzati.

Infine, vi è il principio di non discriminazione algoritmica, che è oggetto di riflessione in queste pagine. Il GDPR, invero, non contiene una previsione specifica, ma un'enunciazione rilevante in materia è data dal Considerando n. 71. In esso si sottolinea l'opportunità che il titolare del trattamento operi «secondo una modalità che tenga conto dei potenziali rischi esistenti per gli interessi e i diritti dell'interessato e impedisca, tra l'altro, effetti discriminatori nei confronti di persone fisiche sulla base della razza o dell'origine etnica, delle opinioni politiche, della religione o delle convinzioni personali, dell'appartenenza sindacale, dello status genetico, dello stato di salute o dell'orientamento sessuale, ovvero un trattamento che comporti misure aventi tali effetti».

Tale indicazione, poiché è contenuta in un Considerando, non ha valore prescrittivo, ma interpretativo, orientativo. Si può però ritenere che essa costituisca una precisazione del più generale principio di non discriminazione sancito dalla Carta dei diritti fondamentali dell'Unione Europea (artt. 20 e ss.) e dalla Convenzione europea dei diritti dell'uomo (art. 14), oltre che dalla nostra Costituzione, che tutela il principio di uguaglianza (art. 3).

Per arginare i rischi di discriminazione, nonché per garantire altre finalità quali la correttezza e la sicurezza dei dati, la norma richiama la necessità di adottare «misure tecniche e organizzative adeguate». Occorre quindi operare sia sul fronte della costruzione e implementazione degli algoritmi sia su quello dell'organizzazione del contesto e delle strutture in cui l'algoritmo è configurato e/o utilizzato. Per quanto riguarda gli aspetti organizzativi merita di essere segnalato uno spunto contenuto negli Orientamenti etici per un'intelligenza artificiale affidabile (2019) definiti dall'UE, secondo il quale andrebbe incoraggiata l'assunzione di personale «proveniente da contesti, culture e discipline diverse» così da garantire «la diversità di opinioni».

I principi citati, applicati nell'organizzazione e nelle dinamiche di funzionamento, vale a dire delle procedure, delle p.a., che si è visto essere artefici di non poche discriminazioni, possono contribuire a far emergere i fenomeni distorsivi e le ragioni che ne stanno alla base, consentendo quindi di intervenire, eliminandoli o limitandoli. Può giocare un ruolo rilevante in tal senso il responsabile del procedimento[63], in cui può ravvisarsi una concretizzazione del principio *Human-in-the-loop*. Si pensi inoltre alla virtuosa relazione di reciprocità che può instaurarsi tra trasparenza amministrativa e trasparenza algoritmica[64].

[63] Sul rapporto tra responsabile del procedimento e intelligenza artificiale v. D.U. GALETTA, J.G. CORVALÁN, *Intelligenza Artificiale per una Pubblica Amministrazione 4.0? Potenzialità, rischi e sfide della rivoluzione tecnologica in atto*, cit., p. 19.

[64] In tema, v. A.G. OROFINO, *The Implementation of the Transparency Principle in the Development of Electronic Administration*, in *Erdal*, I (2020), 1-2, 123 ss. Sulla trasparenza informatica v. altresì ID., *La trasparenza oltre la crisi. Accesso, informatizzazione e controllo civico*, Bari, 2020, in particolare pp. 193 ss.

La proposta di Regolamento in materia di intelligenza artificiale diversifica il regime regolatorio applicabile alle attività di intelligenza artificiale sulla base di un approccio basato sul rischio che queste comportano. Nell'ottica del discorso che si sta conducendo risulta importante sottolineare come questo sia strettamente correlato, oltre che a questioni di salute e di sicurezza, proprio all'impatto che tali tecnologie hanno sui valori dell'Unione e sui diritti fondamentali[65].

In questo quadro nella proposta alle problematiche della discriminazione è dedicato uno spazio sicuramente maggiore rispetto al GDPR. Essa, infatti, si prefigge di integrare «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» (Relazione, par. 1.2).

Anzitutto, tra le pratiche di intelligenza vietate vi sono i sistemi c.d. di *social credit scoring* (art. 5, comma 1, lett. c) [66], utilizzati, come si è visto, in Cina, che determinino uno o entrambi dei seguenti effetti distorsivi: un trattamento pregiudizievole o sfavorevole di talune persone fisiche o di interi gruppi di persone fisiche in contesti sociali che non hanno alcun rapporto con quelli in cui i dati sono stati originariamente generati o raccolti; un trattamento con le stesse caratteristiche che sia sproporzionato rispetto alla gravità del comportamento sociale di tali individui o gruppi.

Il testo invoca in più parti una specifica attenzione alle problematiche discriminatorie, come pure di violazione di altri diritti fondamentali, che possono insorgere in ambiti peculiari, da qualificarsi quindi come «ad alto rischio» (art. 6, par. 2 e all. III). Può essere di interesse osservare come molti di questi riguardino settori di attività pubblica. In larga parte si tratta proprio degli ambiti cui si riferiscono gli episodi discriminatori citati nelle pagine precedenti: giustizia, sicurezza pubblica e polizia, erogazione di servizi essenziali quali l'assistenza pubblica; a questi vanno aggiunti istruzione e formazione professionale, gestione dei fenomeni migratori e controllo alle frontiere, ecc.

Tra sistemi più controversi, rispetto ai quali si auspicherebbero in seno al dibattito previsioni ancora più restrittive di quelle attualmente prospettate e finanche la completa messa al bando, vi sono quelli di identificazione biometrica e di polizia predittiva, proprio in quanto forieri di possibili pesanti discriminazioni e di una pericolosa sorveglianza di massa[67].

---

[65] Si distingue pertanto tra applicazioni aventi un rischio inaccettabile e quindi vietate e sistemi ad alto rischio e a rischio basso o minimo.

[66] L'articolo parla di immissione sul mercato, messa in servizio o uso di sistemi di IA da parte delle autorità pubbliche o per loro conto «ai fini della valutazione o della classificazione dell'affidabilità delle persone fisiche per un determinato periodo di tempo sulla base del loro comportamento sociale o di caratteristiche personali o della personalità note o previste».

[67] Per alcune osservazioni sul dibattito che sta avendo luogo al Parlamento europeo v. A.D. SIGNORELLI, *Il regolamento europeo sull'intelligenza artificiale inizia a prendere forma*, in *La Repubblica*, *Italian Tech*, 28 aprile 2022.

I sistemi di IA ad alto rischio sono quelli in cui, come accennato, maggiori sono le preoccupazioni per la sicurezza e il rispetto dei diritti fondamentali, per garantire i quali il futuro regolamento prescrive una serie di accorgimenti tecnici, relativi ai meccanismi c.d. "interni" della decisione algoritmica, che possono operare anche in senso antidiscriminatorio[68]. Tali prescrizioni riguardano la progettazione degli algoritmi, la qualità dei dati utilizzati, l'addestramento dei modelli, le fasi di convalida e prova, ecc.

Si è avuto modo di osservare come spesso i *bias* derivino dallo stesso *machine learning*. In questi casi non sempre l'adozione preventiva di misure ed accortezze sul piano tecnico è sufficiente ad arginare il problema: le discriminazioni prodotte in seno alla *black box* possono essere sconosciute, inintelligibili quanto questa, il che, si è avuto modo di osservare, ne rende oscura la genesi agli stessi programmatori.

Per evitare o limitare gli effetti discriminatori delle decisioni algoritmiche sono quindi necessari anche controlli ed interventi *ex post*, a "valle" del processo algoritmico, effettuati dall'uomo. A venire in considerazione è proprio l'accennato principio di non esclusività algoritmica, che può quindi giocare di sponda rispetto al principio di non discriminazione.

Con riferimento ai sistemi di IA ad alto rischio, tali anche per i fenomeni discriminatori che possono interessarli, la proposta di regolamento traduce il c.d. *Human in the loop* in uno specifico articolo dedicato alla «Sorveglianza umana» (art. 14).

La previsione muove proprio dalla consapevolezza del fatto che rischi per la sicurezza e i diritti fondamentali possano emergere «nonostante» l'applicazione di altri requisiti volti a rendere affidabili i sistemi in esame. Tra le possibili azioni contemplate dalla norma vi è anche l'*extrema ratio* del pulsante di "arresto".

Anche i principi di conoscibilità e di comprensibilità possono concorrere a rafforzare il principio di non discriminazione, permettendo di verificare se un algoritmo operi in maniera discriminatoria e, nel caso, per quali ragioni e meccanismi, fatti salvi gli evidenziati limiti concernenti la *black box*[69].

Da questo punto di vista, la proposta di regolamento si occupa specificamente dell'opacità algoritmica[70] e la trasparenza è concepita come uno strumento di tutela dei diritti fondamentali (Relazione, par. 3.5), specialmente nel caso di sistemi di IA ad alto rischio (Relazione, par. 5.2.3).

Le prime "sentinelle" rispetto al verificarsi di distorsioni algoritmiche sono indubitabilmente gli utenti: i sistemi di IA ad alto rischio devono essere progettati e sviluppati in modo tale da garantire che il loro funzionamento sia

[68] Distingue tra profili "interni" ed "esterni" della decisione algoritmica P. ZUDDAS, *Intelligenza artificiale e discriminazioni*, cit., pp. 12 ss.

[69] Così A. SIMONCINI, *Amministrazione digitale algoritmica. Il quadro costituzionale*, cit., p. 37.

[70] In argomento v. E. CAMPO, A. MARTELLA e L. CICCARESE, *Gli algoritmi come costruzione sociale. Neutralità, potere e opacità*, cit., pp. 7 ss.

sufficientemente trasparente così da consentire a chi li usa di interpretarne i risultati e utilizzarli in modo adeguato (art. 13). Gli utenti debbono essere anche informati dei possibili rischi di violazione dei propri diritti fondamentali (art. 13, comma 2, iii).

5. La rilevanza di una tutela *by education.*

L'analisi che si è condotta mostra come il tema delle discriminazioni algoritmiche sia affrontato con un crescendo di consapevolezza sia dal GDPR che dalla proposta di regolamento della Commissione europea in tema di IA, che cerca di bilanciare il sostegno alle potenzialità di quest'ultima con la minimizzazione dei rischi per i diritti degli individui che possono derivare dal suo impiego.

I sistemi di regolazione attuale, così come quelli che si prospettano nel futuro, hanno però dei limiti intrinseci, dovuti alla velocità e anche all'imprevedibilità degli sviluppi tecnologici, rispetto ai quali il diritto non sempre è in grado di fornire una risposta pronta e congrua. Difficile è inoltre disciplinare ciò che non si conosce appieno: si pensi alle problematiche che scaturiscono dalla *black box*, il cui funzionamento resta sconosciuto agli stessi programmatori.

In un quadro di questo tipo, dominato da complessità e incertezza, né le strumentazioni e la progettazione di natura tecnica né la risposta giuridica sono sufficienti. Occorre intervenire anche a livello c.d. «morale-soggettivo»[71], ossia delle conoscenze, della cultura e dei codici di comportamento degli individui. Si rivela quindi determinante l'azione sul piano educativo: gli studiosi parlano di una tutela *by education*[72].

L'educazione cui ci si riferisce non è solo quella tecnica; occorre che essa assuma una dimensione sociale ed etica. Negli Orientamenti etici dell'UE per un'intelligenza artificiale affidabile (2019), significativamente, si propone di «formare sistematicamente una nuova generazione di esperti in etica dell'IA».

Sia chi progetta gli algoritmi che chi li usa deve avere coscienza dei valori implicati dalla relativa diffusione: dignità, libertà, uguaglianza, per citare i principali.

Affinché chi si occupa di tecnologie digitali e algoritmi faccia propri tali principi e li trasferisca nel proprio lavoro è importante che questi divengano parte integrante della formazione: di qui il ruolo fondamentale rivestito da università, agenzie formative e associazioni professionali.

---

[71] È questo il primo dei livelli di intervento volti al governo della "rete" secondo la già accennata ricostruzione di A. SIMONCINI, *Sovranità e potere nell'era digitale*, cit., pp. 34 ss.
[72] Sono di questo avviso A. SIMONCINI, S. SUWEIS, *Il cambio di paradigma nell'intelligenza artificiale e il suo impatto sul diritto costituzionale,* cit., pp. 103 ss.

All'educazione dei programmatori, che disegnano gli algoritmi, deve accompagnarsi quella degli utenti che possono esserne vittima, che debbono avere le necessarie basi per capire le nuove tecnologie, utilizzarle con "saggezza"[73] e – vale la pena sottolinearlo − per rendersi conto degli eventuali fenomeni distorsivi. A tal proposito non bisogna peraltro dimenticare che gli stessi utenti possono essere consapevoli o ignari autori di stigmatizzazioni e discriminazioni, il che rende tanto più irrinunciabile un'adeguata educazione ai valori.

Proprio in un'ottica più generale, che guarda alla società digitale nel suo complesso, si è evidenziata l'importanza di ascrivere digitalizzazione e intelligenza artificiale ai contenuti del diritto/dovere di istruzione[74]. I passi in tale direzione, che vedono tra le numerose materie del reintrodotto insegnamento dell'educazione civica anche l'educazione alla cittadinanza digitale, sono però ancora troppo timidi[75].

[73] In tema di "destrezza" e "saggezza" digitale v. G. PEDRAZZI, *La cittadinanza digitale: educazione, partecipazione e inclusione*, in A. Calore e F. Mazzetti (curr.), *I confini mobili della cittadinanza*, Torino, 2019, in particolare pp. 191 ss.

[74] In tal senso v. C. CASONATO, *Per un'intelligenza artificiale costituzionalmente orientata*, cit., pp. 105-106.

[75] Il riferimento è alla l. 20 agosto 2019, n. 92, che ha previsto l'insegnamento dell'educazione civica, a lungo assente dai programmi scolastici, ricomprendendo, tra le varie tematiche che ne fanno parte, l'educazione alla cittadinanza digitale (art. 3, comma 1, lett. c) e art. 5). Le materie trattate, però, sono troppe e varie per dedicare ad esse un tempo adeguato: vanno dalla digitalizzazione allo sviluppo sostenibile, alla lotta alle mafie, all'educazione stradale.