

Decisioni algoritmiche e discriminazioni: lo stato dell'arte*

Dora Trombella

Abstract

Alcuni sistemi di Intelligenza Artificiale funzionano mediante dei meccanismi di *machine learning*, ossia apprendono e migliorano le proprie performance sulla base dei dati che utilizzano. La macchina algoritmica, estranea a qualsiasi valutazione di tipo etico, procede in via automatica a rielaborare le informazioni razionalizzando tutto ciò che è reale e, altresì, perpetrando i *biases* radicati nella società. Tuttavia, l'AI Act potrebbe fornire valide soluzioni con riguardo a questa problematica, in quanto, anzitutto, mira ad assicurare che i sistemi di IA immessi sul mercato europeo rispettino i diritti fondamentali e i valori dell'Unione.

Some Artificial Intelligence systems function through machine learning mechanisms, which means that they learn and improve their performance based on the data they collect. The algorithm, with no ethical evaluation, automatically proceeds to reprocess information by rationalising all that is real and, likewise, perpetuating the ingrained biases in society. However, about this issue, the AI Act could be valuable because at its core it aims to ensure that AI systems in the European market respect the fundamental rights and the values of the Union.

Sommario

1. Da informazioni distorte ad esiti pregiudizievoli nei settori “ad alto rischio” – 2. L'opacità delle procedure decisionali. – 3. Gli strumenti forniti dal GDPR. – 4. AI Act: un argine alla discriminazione algoritmica? – 5. Qualche considerazione conclusiva.

Keywords

intelligenza artificiale – algoritmo – apprendimento automatico – decisioni algoritmiche – discriminazioni algoritmiche

* Su determinazione della direzione, il contributo è stato sottoposto a referaggio anonimo in conformità all'art. 15 del regolamento della Rivista

1. Da informazioni distorte ad esiti pregiudizievoli nei settori “ad alto rischio”

Con la nozione di *machine learning* ci si riferisce ad una procedura automatizzata volta ad individuare delle correlazioni tra più variabili all'interno di un set di dati al fine di effettuare previsioni o stime¹.

La raccolta e l'elaborazione dei dati sono fondamentali per la realizzazione delle applicazioni di *machine learning*, in quanto la qualità delle informazioni immesse nel sistema incide notevolmente sulle sue prestazioni².

Infatti, nelle specifiche fasi di programmazione e di apprendimento possono introdursi nel modello delle informazioni idonee a generare un risultato discriminatorio ed escludente rispetto a individui o a gruppi di individui³.

È necessario ricordare che il diritto antidiscriminatorio individua due tipologie di discriminazione: la discriminazione diretta che sussiste nel caso in cui per la razza o l'origine etnica, il sesso, la religione, le convinzioni personali, gli handicap, l'età, l'orientamento sessuale «una persona è trattata meno favorevolmente di quanto sia, sia stata o sarebbe trattata un'altra in situazione analoga»; mentre la discriminazione indiretta si verifica «quando una disposizione, un criterio, una prassi, un atto, un patto o un comportamento apparentemente neutri possono mettere le persone», per le caratteristiche personali suddette, in «una situazione di particolare svantaggio rispetto ad altre persone»⁴.

La definizione di discriminazione diretta mal si adatta alle forme di discriminazione algoritmica⁵, in quanto, la macchina, estranea a qualsiasi giudizio morale e operando unicamente su un piano logico-matematico, eredita e riproduce automaticamente le scelte preliminari effettuate dall'essere umano che l'ha programmata. Peraltro, è altamente improbabile che la decisione algoritmica si basi sulla esclusione di caratteri che identificano la categoria protetta in via immediata, discendendo (l'esclusione), piuttosto, da caratteristiche inestricabilmente connesse al fattore di discriminazione⁶.

La discriminazione indiretta, invece, che crea disparità di trattamento non intenzionali, pare, almeno da un punto di vista concettuale, più confacente al fenomeno oggetto di

¹ D.Lehr - P.Ohm, *Playing with the data: What legal scholars should learn about machine learning*, in *University of California Davis Law Review*, 51, 2017, 671.

² T. Wang - B. Li - M. Chen - S. Yu, *Machine Learning Empowered Intelligent Data Center Networking: Evolution, Challenges and Opportunities*, Singapore, 2023, 10.

³ S. Barocas - A.D. Selbst, *Big data disparate impact*, in *California Law Review*, 104, 2016, 671 ss. indicano le cinque circostanze in cui possono verificarsi eventuali discriminazioni: nella fase di individuazione delle *class labels* per la definizione dei risultati; nella fase di addestramento dei dati; durante la selezione delle caratteristiche rilevanti per il modello; a partire dai *proxies* prescelti o dalla discriminazione intenzionale inserita dai programmatori.

⁴ Il riferimento è alle seguenti disposizioni: Art. 2, Dir. 2000/43/CE sulle discriminazioni per razza o origine etnica; Art. 2, Dir. 2000/78/CE sulle discriminazioni per religione, convinzioni personali, handicap, età, tendenze sessuali; Art. 2, Dir. 2006/54/CE sulle discriminazioni di genere.

⁵ D. Morondo Taramundi, *Le sfide della discriminazione algoritmica*, in *Genius, Rivista di studi giuridici sull'orientamento sessuale e l'identità di genere*, 1, 2022, 7.

⁶ *Ivi*, 8.

trattazione⁷.

Ebbene, tali potenziali esiti pregiudizievole si sono, di fatto, riscontrati in ambiti particolarmente sensibili: la giustizia, la salute, il lavoro e l'accesso al credito⁸.

Con riguardo al primo settore, occorre precisare che con “giustizia predittiva” si intende la possibilità per l'algoritmo di prevedere l'esito di una controversia mediante l'elaborazione di dati normativi e giurisprudenziali, previa consultazione di banche dati, raccolte di giurisprudenza e opere edite in formati accessibili al sistema⁹.

Infatti, le macchine adibite a questo scopo possiedono delle capacità di memorizzazione e computazione, impensabili per l'essere umano, che consentono la formulazione immediata della decisione a seguito dell'analisi di migliaia di dati.

Sul tema è emblematica la vicenda relativa al celebre sistema COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*), ampiamente utilizzato negli USA, che quantifica mediante un algoritmo il rischio che un imputato possa delinquere nuovamente. In particolare, tale sistema valuta la probabilità di recidiva sia sulla base di quanto contenuto nel fascicolo processuale sia “alla luce” delle informazioni assunte a seguito di un test composto da 137 domande a cui l'imputato viene sottoposto e che riguardano l'età, l'attività lavorativa svolta, il grado di istruzione, i legami affettivi, l'uso di droghe, le opinioni personali e il percorso criminale.

Come noto, la Suprema Corte del Wisconsin nel 2016 si è pronunciata in merito al caso del Sig. Eric L. Loomis a cui il sistema COMPAS, nel giudizio di primo grado, aveva comminato una pena di sei anni di reclusione per ricettazione e resistenza a pubblico ufficiale¹⁰.

Il Sig. Loomis, sosteneva che il sistema funzionasse in maniera del tutto ignota alla difesa col rischio che l'imputato potesse incorrere in valutazioni discriminatorie e, altresì, che, nel suo caso, non vi fosse stata una pronuncia personalizzata, basandosi l'algoritmo sulla rielaborazione di informazioni relative al gruppo etnico in cui l'imputato era incluso.

La Corte Suprema ha dichiarato, all'unanimità, la legittimità della procedura automatizzata che aveva condotto alla sentenza impugnata, in quanto, il sistema COMPAS, a detta della Corte, rappresenta unicamente un ausilio rispetto all'attività degli organi giudiziari che devono valutare, anche in base alle informazioni fornite dalla macchina,

⁷ Per una trattazione dettagliata della problematica si rinvia a C. Nardocci, *Intelligenza Artificiale e discriminazioni*, in *Rivista del Gruppo di Pisa*, 3, 2021.

⁸ Con riguardo all'impatto delle nuove tecnologie sul diritto costituzionale si veda S. Simoncini, *L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà*, in *BioLaw Journal*, 1, 2019, 63 ss.; C. Colapietro, *Intelligenza artificiale e discriminazioni*, in *Studi parlamentari e di politica costituzionale*, 2022, I, 9 ss.; M. D'amico, *Una parità ambigua. Costituzione e diritti delle donne*, Milano, 2020.

⁹ Approfondisce la tematica U. Ruffolo, *La machina sapiens come “avvocato generale” ed il primato del giudice umano: una proposta di interazione virtuosa* in U. Ruffolo, (a cura di), *XXVI lezioni di diritto dell'intelligenza artificiale*, Torino, 2021, 205 ss evidenziando che la macchina potrebbe funzionare essenzialmente come *bouche de la loi*, optando la stessa per la soluzione «oggettivamente esatta in quanto rispondente al comando generale ed astratto della legge».

¹⁰ Per una trattazione più dettagliata della vicenda: F. Lagioia - G. Sartor, *Il sistema compas: algoritmi, previsioni, iniquità*, in U. Ruffolo (a cura di), *XXVI lezioni di diritto dell'intelligenza artificiale*, Torino, 2021, 226 ss.

il rischio di recidiva dell'imputato¹¹.

Inoltre, la Corte ha escluso che il sistema possa definirsi discriminatorio, in quanto l'assegnazione di un punteggio più alto agli imputati di sesso maschile è da ritenersi compatibile con le statistiche che mostrano che le donne commettono meno crimini violenti e sono meno recidive rispetto agli uomini. Per quanto riguarda l'eventuale distorsione generata dal sistema che attribuisce automaticamente agli uomini neri una valutazione più elevata rispetto ai bianchi, la Corte si è limitata ad invitare gli utilizzatori della macchina ad informarsi maggiormente sulla problematica.

Nell'ordinamento italiano, indubbiamente, sarebbe precluso un processo decisionale automatizzato in chiave predittiva siffatto sia in virtù di quanto previsto dall'art. 22, par. 1, GDPR (come si dirà meglio in seguito) ma anche a norma dell'art. 220, c. 2, del Codice di procedura penale che proibisce ogni perizia volta a stabilire il carattere o la personalità dell'imputato. Per giunta, il sistema risulterebbe lesivo del diritto all'equo processo nonché del diritto di difesa dell'imputato tutelati dalla Carta costituzionale. Tuttavia, anche in Italia, sono stati avviati diversi progetti nel settore giustizia che relegano però l'algoritmo ad un ruolo ausiliario rispetto all'operato dell'autorità giudiziaria¹².

I sistemi di IA, come si anticipava, sono, altresì, impiegati per assumere decisioni in campo sanitario al fine di predire il successo di una determinata terapia, la diffusione di una malattia o lo sviluppo di specifiche patologie o, ancora, per determinare il diritto all'accesso ai servizi sanitari.

Invero, i più moderni sistemi di intelligenza artificiale provvedono a rielaborare una moltitudine di informazioni concernenti i casi clinici pregressi e le caratteristiche dei pazienti coinvolti, nonché la sintomatologia, la diagnosi, la cura applicata restituendo automaticamente, sulla base di quanto appreso, un responso al caso sottoposto al loro esame¹³.

Ad esempio, i sistemi sanitari statunitensi si affidano ad algoritmi di previsione "commerciali" al fine di aiutare i pazienti con esigenze sanitarie particolarmente complesse. È interessante il caso concernente uno di questi modelli che impiegava i costi sanitari come *proxies* per determinare il bisogno di cure. Nella specie, i pazienti neri, a parità di condizioni di salute, risultava spendessero meno rispetto ai bianchi; dunque, il sistema concludeva erroneamente che i pazienti neri fossero più sani degli altri e, pertanto,

¹¹ Supreme Court of Wisconsin, *State of Wisconsin v. Eric L. Loomis*, 13 July 2016, Case no. 2015 AP157-CR.

¹² M. Martorana, *Polizia e giustizia predittive: cosa sono e come vengono applicate in Italia*, in *Agenda Digitale*, 27 gennaio 2021; C. Morelli, *Giustizia predittiva: il progetto (concreto) della Corte d'appello di Brescia*, in *Altalex.com*, 8 aprile 2019; C. Castelli, *Giustizia predittiva: i progetti in corso in Italia*, in *Agenda Digitale*, 2 agosto 2023.

¹³ Sull'utilizzo dell'IA in medicina si segnalano i seguenti contributi: D. Pacini - G. Folesani, *Il ruolo dell'Intelligenza Artificiale in cardiocirurgia*, in U. Ruffolo - M. Gabrielli, (a cura di), *Intelligenza Artificiale, dispositivi medici e diritto. Un dialogo fra saperi: giuristi, medici e informatici a confronto*, Torino, 2023, 43 ss; A. Zini, *Intelligenza artificiale e patologie neurologiche e cerebrovascolari*, in U. Ruffolo - M. Gabrielli, (a cura di), *Intelligenza Artificiale, dispositivi medici e diritto. Un dialogo fra saperi: giuristi, medici e informatici a confronto*, Torino, 2023, 49 ss; G. Pipino, *Intelligenza artificiale in ortopedia*, in U. Ruffolo - M. Gabrielli, (a cura di), *Intelligenza Artificiale, dispositivi medici e diritto. Un dialogo fra saperi: giuristi, medici e informatici a confronto*, Torino, 2023, 57 ss.

destinava loro meno risorse economiche¹⁴.

Un'altra problematica del settore concerne la mancata inclusione nei *data set* di informazioni riguardanti individui differenti da maschi bianchi.

Tale distorsione potrebbe, infatti, reiterare i gravissimi esiti a cui, in passato, si è giunti non effettuando nessun test clinico su persone di sesso femminile nella fase di sperimentazione di un farmaco o di una terapia¹⁵.

Della questione si è interessato il Comitato Nazionale per la Bioetica con il parere «La sperimentazione farmacologica sulle donne» approvato nella seduta plenaria del 28 novembre 2008 in cui viene evidenziato che «sebbene le donne siano le maggiori consumatrici di farmaci, la sperimentazione tende a non tenere in sufficiente considerazione la loro specificità e il cambiamento delle condizioni di salute femminile, con un conseguente incremento di danni avversi all'assunzione di farmaci». Invero, «La donna non può essere assimilata all'uomo, come una mera variabile, ma ha una specificità che la sperimentazione è chiamata a tenere in considerazione per promuovere una medicina che riconosca adeguatamente le pari opportunità uomo/donna».

La parzialità dei set di dati utilizzati a fini di ricerca scientifica ha caratterizzato, di recente, alcune app, ad esempio *Derm Assist* e *SkinVision*, che visualizzando le fotografie scattate con lo *smartphone* possono agilmente individuare una serie di patologie cutanee. In particolare, queste tecnologie provvedono a segnalare le escrescenze come innocue o “ad alto rischio” e, dunque, consigliano all'utente se rivolgersi o meno alle cure. Tuttavia, è prontamente emerso che le applicazioni suddette utilizzano algoritmi non adeguatamente addestrati e producono pregiudizi sistematici relativi alla razza, all'età e, addirittura, al tipo di assicurazione stipulata in quanto sono stati sviluppati a partire da immagini riferibili pressoché unanimemente a pazienti anziani, maschi e bianchi¹⁶.

Anche le decisioni automatizzate assunte in ambito lavorativo hanno prodotto analoghi effetti discriminatori. Il caso più celebre ha ad oggetto l'algoritmo che Amazon aveva progettato per automatizzare la procedura di reclutamento del personale. Tale sistema, infatti, era stato allenato attraverso i curricula ricevuti dalla società nell'arco dei dieci anni precedenti che provenivano, in larga parte, da individui di sesso maschile. Seppur non fosse stato inserito il sesso come criterio selettivo, l'algoritmo era riuscito a riconoscerlo da altre informazioni e, realizzando una discriminazione di tipo “intersezionale”, favoriva nella valutazione i termini presenti in maggior numero all'interno dei curricula degli uomini¹⁷.

¹⁴ Sulle decisioni pregiudizievoli in sanità si rinvia a: Z. Obermeyer - B. Powers - C. Vogeli - S. Mullainathan, *Dissecting racial bias in an algorithm used to manage the health of populations*, in *Science*, 366,6464, 2019, 447 ss.; N.Norori - Q.Hu - F. M. Aellen - F. D. Faraci - A. Tzovara, *Addressing bias in big data and AI for health care: A call for open science*, in *Patterns*, 2(10), 2021, 1 ss.; R.B.Parikh - S.Teeple - A.S. Navathe, *Addressing bias in artificial intelligence in health care*, in *JAMA*, 322, 2377-2378.

¹⁵ Sul tema F. Franconi - I. Campesi, *Pharmacogenomics, pharmacokinetics and pharmacodynamics: interaction with biological differences between men and women*, in *British Journal of Pharmacology*, 171(3), 2014, 580 ss.; v. anche A. Carnevale - E. A. Tangari - A. Iannone - E. Sartini, *Will Big Data and personalized medicine do the gender dimension justice?*, in *AI & Society*, 38(2), 2023, 829 ss.

¹⁶ J. Madhusoodanan, *These apps say they can detect cancer. But are they only for white people?*, in *The Guardian*, 28 agosto 2021.

¹⁷ J. Destin, *Amazon scraps secret AI recruiting tool that showed bias against women*, in *Reuters*, 11 ottobre 2018.

Sulla cecità discriminatoria dell'algoritmo a discapito dei lavoratori è interessante rammentare il contenuto dell'ordinanza del Tribunale di Bologna (sez. lavoro) del 31 dicembre 2020¹⁸ che ha accolto il ricorso presentato da alcune associazioni sindacali contro una nota società di *food delivery* poiché riteneva che il sistema con cui venivano gestite le prenotazioni dei turni dei *riders*, basato sul cosiddetto Algoritmo *Frank*, fosse discriminatorio. Infatti, tale modello assegnava un determinato punteggio in base alla effettiva partecipazione ai turni prescelti oppure in base alla tempestiva disdetta comunicata con un preavviso di ore ventiquattro. Dunque, l'algoritmo penalizzava allo stesso modo sia il *rider* che non aveva partecipato al turno prescelto per scarsa professionalità sia quello che non vi aveva potuto partecipare perché, ad esempio, aveva deciso di esercitare il proprio diritto di sciopero.¹⁹

Con riguardo ad una vicenda simile, la Sezione Lavoro del Tribunale di Palermo con sentenza resa in data 31 marzo 2023 ha ritenuto che «è antisindacale e pertanto va repressa con gli strumenti propri della procedura di cui all'art. 28 Stat. Lav. la condotta posta in essere dal datore di lavoro che, servendosi di *riders* per garantire la fornitura dei beni e servizi prodotti, si avvalga di una piattaforma digitale il cui meccanismo di funzionamento, segreto, appaia lesivo oltre che discriminatorio nella scelta del ciclofattorino cui affidare la commessa», peraltro, secondo il Giudice, il datore di lavoro avrebbe l'onere di «palesare le modalità di funzionamento dello strumento elettronico in questione».

Se in Italia i limiti all'utilizzo di piattaforme online per garantire la tutela dei lavoratori contro le discriminazioni algoritmiche sono da rintracciarsi nelle varie pronunce giurisprudenziali, altri ordinamenti europei sono stati "tempestivi" nella regolazione del fenomeno. Primo fra tutti il legislatore spagnolo che, col Real Decreto ley n. 9/2021 convertito nella ley n. 12 del 28 settembre 2021, è intervenuto sullo *Estatuto de los trabajadores* modificandone l'art. 64, relativo ai diritti di informazione e consultazione della rappresentanza legale dei lavoratori. Al nuovo par. 4, lett. d), la norma dispone che il comitato aziendale deve essere informato dall'azienda dei parametri, delle regole e delle istruzioni su cui si basano gli algoritmi o i sistemi di intelligenza artificiale che influiscono sul processo decisionale che può avere un impatto sulle condizioni di lavoro, sull'accesso e sul mantenimento dell'impiego, compresa la profilazione.

Le tecniche di *machine learning* trovano vasto impiego anche nella valutazione del rischio di credito, poiché a differenza degli approcci statistici tradizionali detengono un livello di accuratezza superiore e sono in grado di elaborare un'enorme quantità di dati in volume (numero di osservazioni) e ricchezza (numero di variabili, tipologie di dato)²⁰.

¹⁸ Per una analisi puntuale dell'argomento: S. Borelli - M. Ranieri, *La discriminazione nel lavoro autonomo. Riflessioni a partire dall'algoritmo Frank*, in *Labour & Law Issues*, 7(1), 2021; M. Borzaga - M. Mazzetti, *Discriminazioni algoritmiche e tutela dei lavoratori: riflessioni a partire dall'Ordinanza del Tribunale di Bologna del 31 dicembre 2020*, in *BioLaw Journal*, 1, 2022; A. Perulli, *La discriminazione algoritmica: brevi note introduttive a margine dell'Ordinanza del Tribunale di Bologna*, in *Lavoro Diritti Europa*, 1, 2021.

¹⁹ Sulle differenze di funzionamento degli algoritmi c.d. *rule-based* (es. l'algoritmo Frank) rispetto a quelli di *machine learning* (utilizzato nel caso Amazon) si rinvia a G. Gaudio, *Le discriminazioni algoritmiche*, in *Lavoro Diritti Europa*, 1, 2024, 5. Più diffusamente sul tema, M. Barbera, *Discriminazioni algoritmiche e forme di discriminazione*, in *Labour & Law Issues*, 7(1), 2021.

²⁰ E. Bonaccorsi Di Patti - F. Calabresi - B. De Varti - F. Federico - M. Affinito - M. Antolini - F. Lorzio - S. Marchetti - I. Masiani - M. Moscatelli - F. Privitera - G. Rinna, *Intelligenza artificiale nel credit scoring*.

In sostanza, quando un istituto di credito riceve una richiesta di finanziamento da parte di un cliente il cosiddetto *credit scoring* valuta tutti i suoi dati personali (regolarità nei pagamenti, morosità, debiti, etc.) e, di conseguenza, decide se concedere o meno il prestito richiesto.

In questo settore²¹ è diffusa la configurazione di *biases* “storici”,²² ossia: i soggetti appartenenti ad uno specifico genere, gruppo etnico o sociale a cui in passato era stato negato l’accesso al credito vengono sottorappresentati nei dati di riferimento dell’algoritmo a vantaggio delle categorie cui storicamente è stato concesso.

A titolo d’esempio si ricorda la vicenda²³ giunta sino alla Corte di giustizia europea avente ad oggetto le controversie di un’agenzia privata di informazione creditizia, la Schufa.

Nella specie, una cittadina tedesca si era vista negare la concessione di un mutuo da parte della banca a fronte di un *credit scoring* negativo fornito dall’agenzia che si era rifiutata, altresì, di fornire una puntuale motivazione con riguardo alla decisione assunta invocando il segreto commerciale sulle informazioni relative al metodo di calcolo utilizzato. Dinnanzi al Giudice europeo la Schufa ha costruito la propria difesa sulla estraneità della agenzia a qualsiasi responsabilità in quanto si sarebbe limitata a fornire una valutazione all’istituto di credito che, in seguito, avrebbe adottato materialmente la decisione.

La Corte ha osservato che la lesione del diritto della ricorrente si è realizzata nel momento in cui la Schufa ha espresso il proprio giudizio negativo sull’affidabilità creditizia che va considerato come vera e propria “decisione” così come intesa nella fattispecie di cui all’art. 22 del GDPR. Infatti, se l’attività di *scoring* venisse relegata a mero momento preparatorio della decisione dell’istituto di credito, l’interessato non potrebbe esercitare il proprio diritto di accesso nei confronti della agenzia che elabora la valutazione di affidabilità creditizia né potrebbe ottenere eventuali informazioni dall’istituto mutuante che generalmente non ne dispone.

Un ricorso²⁴ simile presentato dinnanzi alla Corte di giustizia riguarda una consumatrice austriaca che si è vista negare il rinnovo di un abbonamento di telefonia mobile del costo di dieci euro mensili a seguito di una valutazione di inaffidabilità finanziaria fornita in modo automatizzato (e non conosciuto dalla ricorrente) dalla Società B&D. Tuttavia, sugli aspetti prettamente giuridici delle ultime due vicende giudiziarie citate ci si soffermerà al terzo paragrafo.

Analisi di alcune esperienze nel sistema finanziario italiano, in *Questioni di economia e finanza (occasional papers)*, 721, 2022, 30.

²¹ Per una panoramica completa sul tema si veda: G. Curcurutu - P. Inturri, *Discriminazioni algoritmiche e tutela dei consumatori vulnerabili nell’accesso al credito*, in *BioLaw Journal*, 1, 2024, 317 ss.

²² Ivi, 34 ss. per tutte le tipologie di distorsioni configurabili in questo settore.

²³ CGUE, C-634/21, *OQ v Land Hessen* (2023).

²⁴ CGUE, C-203/22, *Dun & Bradstreet Austria*.

2. L'opacità delle procedure decisionali

Tra le problematiche «identificate ma non risolte»²⁵ che investono le applicazioni di *machine learning* vi è anche la cosiddetta *black box*, che consiste in «un sistema il cui funzionamento è misterioso; possiamo osservare i suoi ingressi e le sue uscite, ma non sappiamo come l'uno si trasformi nell'altro»²⁶.

In altri termini, se si verifica la *black box* non è possibile tracciare l'iter logico seguito dalla macchina per raggiungere l'obiettivo assegnato ed è, altresì, impossibile comprendere come e perché, sulla base del set di dati elaborati, il sistema sia giunto a determinati risultati²⁷.

Tali applicazioni di intelligenza artificiale sono state assimilate ad «oracoli che fanno pronostici»²⁸ in quanto non sono in grado di accompagnare il dispositivo ad una motivazione logica e, di conseguenza, rendono maggiormente complicate la prevenzione e la repressione di eventuali effetti discriminatori.

A ben vedere le implicazioni derivanti dall'esponenziale propagazione dei sistemi di intelligenza artificiale e i timori di eventuali ripercussioni sui diritti fondamentali risalgono a tempi assai remoti, atteso che già nel 1950, Norbert Wiener, fondatore della cibernetica, faceva menzione nella sua opera²⁹ del racconto *The Monkey's Paw* di William Wymark Jacobs pubblicato in Inghilterra nel 1902 in cui si narra di due coniugi che, dopo aver espresso il desiderio di ricevere (a tutti i costi) duecento sterline dinnanzi ad un talismano, apprendono la tragica notizia della morte del figlio a seguito di un incidente sul lavoro. In seguito, i genitori, come risarcimento del danno, percepiscono proprio una somma di denaro di importo pari a duecento sterline. Wiener, non a caso, ha menzionato questa storia per rappresentare i ciechi meccanismi di funzionamento dei dispositivi di intelligenza artificiale che perseguono pedissequamente gli obiettivi dettati loro dagli esseri umani non tenendo in considerazione le conseguenze negative derivanti dai procedimenti decisionali.

L'effetto *black box* costituisce una problematica “tecnica” del sistema, poiché discende dalla complessità dei calcoli e dalla natura non lineare dei processi automatici di elaborazione, talvolta imperscrutabili agli stessi programmatori³⁰. Infatti, spesso, procedere con tecniche di *reverse engineering* risulta impossibile o eccessivamente oneroso³¹.

²⁵ A. Longo - G. Scorza, *Intelligenza artificiale. L'impatto sulle nostre vite, diritti, libertà*, Milano, 2020, 206.

²⁶ F. Pasquale, *The Black Box Society. The Secret Algorithms that Contain Money and Information*, Cambridge, 2016.

²⁷ G. Lo Sapio, *La black box: l'esplicabilità delle scelte algoritmiche quale garanzia di buona amministrazione*, in *federalismi.it*, 16, 2021, 117. Sugli effetti positivi dell'opacità nei sistemi di intelligenza artificiale si veda L. Floridi, *Etica dell'intelligenza artificiale. Sviluppi, opportunità, sfide*, Milano, 2022, 153 ss.

²⁸ T. Numerico, *Big data e algoritmi. Prospettive critiche*, Roma, 2021, 133. Sull'utilizzo del termine oracolo per definire i sistemi di *machine learning* si rimanda a G. Finocchiaro, *Intelligenza artificiale: quali regole?*, Bologna, 2024, 21 ss.

²⁹ N. Wiener, *The human use of human beings: Cybernetics and society*, Boston, 1950. Sul «ritorno della zampa di scimmia» v. N. Cristianini, *La scoriatoia. Come le macchine sono diventate intelligenti senza pensare in modo umano*, Bologna, 2023, 96 ss.

³⁰ G. Lo Sapio, *La black box: l'esplicabilità delle scelte algoritmiche quale garanzia di buona amministrazione*, cit., 2021, 117.

³¹ Sulla possibilità di un controllo controfattuale v. F. Donati, *Intelligenza artificiale e giustizia*, in *Rivista*

Proprio sulla trasparenza e sulla comprensibilità della procedura decisionale automatizzata è intervenuta la più autorevole giurisprudenza amministrativa italiana³² al fine di colmare una lacuna legislativa «con l'apparato normativo del diritto pubblico»³³ in merito all'utilizzo di un algoritmo nella formulazione delle graduatorie degli insegnanti vincitori di un concorso pubblico e assegnati alle sedi sulla base di criteri non noti e non trasparenti.³⁴ Sul punto va rammentato che dopo l'entrata in vigore della legge n. 107/2015, nota come la "Buona scuola", che prevedeva un piano straordinario di assunzioni a tempo indeterminato e di mobilità su scala nazionale, il MIUR, al fine di gestire più agilmente l'ingente numero di assegnazioni, aveva individuato come soluzione l'utilizzo di un *software*.

Il Consiglio di Stato nella sentenza n. 2270 dell'8 aprile 2019 ha accolto il ricorso presentato da alcuni insegnanti, i quali lamentavano il fatto che la procedura di assunzione fosse gestita interamente da un algoritmo e che tale meccanismo avesse dato luogo a provvedimenti privi di qualsiasi motivazione senza individuare alcun funzionario che si occupasse di valutare le singole situazioni e le preferenze indicate nonché di esternare le relative determinazioni provvedimentali. Nel caso di specie, infatti, in maniera del tutto illogica, ai candidati meglio posizionati in graduatoria erano state assegnate, ai fini dell'individuazione delle sedi di servizio, province lontane da quelle di residenza, mentre i candidati che avevano ottenuto un punteggio inferiore avevano potuto beneficiare di posti nella provincia di residenza, nella disciplina e nell'ordine di scuola espressi nella domanda di assunzione.

Il Giudice amministrativo, nella suddetta pronuncia, ha rilevato che un più elevato livello di digitalizzazione dell'amministrazione pubblica è fondamentale per migliorare la qualità dei servizi resi ai cittadini, in particolare, dall'automazione del processo decisionale della pubblica amministrazione derivano indiscutibili vantaggi in relazione

AIC, 1, 2020, 428.

³² Più nel dettaglio sulla decisione amministrativa automatica: P. Otranto, *Riflessioni in tema di decisione amministrativa, intelligenza artificiale e legalità*, in *federalismi.it*, 7, 2021; Fulvio Costantino, *Rischi e opportunità del ricorso delle amministrazioni alle predizioni dei big data*, in *Diritto pubblico, Rivista fondata da Andrea Orsi Battagliani*, 1, 2019, 43 ss.; F. Patroni Griffi, *La decisione robotica e il giudice amministrativo*, in *Giustizia Amministrativa*, 28 agosto 2018; G. Carullo, *Decisione amministrativa e intelligenza artificiale*, in *Diritto dell'informazione e dell'informatica*, 3, 2021; E. Prosperetti, *Accesso al software e al relative algoritmo nei procedimenti amministrativi e giudiziari. Un'analisi a partire da due pronunce del TAR Lazio*, in *Diritto dell'informazione e dell'informatica*, 4-5, 2019; M. C. Cavallaro - G. Smorto, *Decisione pubblica e responsabilità dell'amministrazione nella società dell'algoritmo*, in *federalismi.it*, 16, 2019; R. Ferrara, *Il giudice amministrativo e gli algoritmi. Note estemporanee a margine di un recente dibattito giurisprudenziale*, in *Diritto amministrativo*, 4, 2019; M. Timo, *Algoritmo e potere amministrativo*, in *Il diritto dell'economia*, 1, 2020; F. Laviola, *Algoritmico, troppo algoritmico: decisioni amministrative automatizzate, protezione dei dati personali e tutela della libertà dei cittadini alla luce della più recente giurisprudenza amministrativa*, in *BioLaw Journal*, 3, 2020; A. Simoncini, *Profili costituzionali della amministrazione algoritmica*, in *Rivista trimestrale di diritto pubblico*, 4, 2019.

³³ M. Palmirani, *Interpretabilità, conoscibilità, spiegabilità dei processi decisionali automatizzati*, in U. Ruffolo (a cura di), *XXVI lezioni di diritto dell'intelligenza artificiale*, Torino, 2021, 74. Sul tema vedi anche B. Marchetti, *La garanzia dello human in the loop alla prova della decisione amministrativa algoritmica*, in *BioLaw Journal*, 2, 2021, 368; N. Paolantonio, *Il potere discrezionale della pubblica automazione. Sconcerto e stilemi. (Sul controllo giudiziario delle "decisioni algoritmiche")*, in *Diritto Amministrativo*, 4, 2021, 820.

³⁴ Come fa notare G. Lo Sapia, *La black box: l'esplicabilità delle scelte algoritmiche quale garanzia di buona amministrazione*, cit., 117 il termine *black box* «richiama, immediatamente, e per contrapposizione, la più nota e risalente metafora dell'amministrazione come "casa di vetro"».

ai canoni di efficienza ed economicità dell'azione amministrativa, i quali, declinando quanto previsto dall'art. 97 Cost. «impongono all'amministrazione il conseguimento dei propri fini con il minor dispendio di mezzi e risorse e attraverso lo snellimento e l'accelerazione dell'iter procedimentale». Pertanto, nei casi come quello de quo, implicanti l'elaborazione di ingenti quantità di istanze, relativi «ad una procedura di assegnazione di sedi in base a criteri oggettivi» e, comunque, se è necessario svolgere «operazioni meramente ripetitive e prive di discrezionalità»³⁵, l'utilizzo dell'algoritmo risulta essere vantaggioso poiché evita qualsiasi negligenza e/o dolo del funzionario ed assicura una maggior garanzia di imparzialità della decisione automatizzata.

Tuttavia, il Giudice precisa che l'algoritmo deve essere considerato come un «atto amministrativo informatico» e, quindi, deve soggiacere ai principi generali dell'attività amministrativa, quali quelli di pubblicità e trasparenza, di ragionevolezza e di proporzionalità; all'algoritmo non devono essere lasciati spazi applicativi discrezionali, ma «deve prevedere con ragionevolezza una soluzione per tutti i casi possibili, anche i più improbabili»; l'amministrazione deve, inoltre, compiere *ex ante* un ruolo di composizione di interessi «anche per mezzo di costanti test, aggiornamenti e modalità di perfezionamento dell'algoritmo»; il giudice può «per la prima volta sul piano “umano”» valutare la correttezza del processo automatizzato in tutte le sue componenti.

Nello specifico il Collegio ha sottolineato che l'algoritmo deve poter essere conoscibile in tutti i suoi aspetti: «Dai suoi autori al procedimento usato per la sua elaborazione, al meccanismo di decisione, comprensivo delle priorità assegnate nella procedura valutativa e decisionale e dei dati selezionati come rilevanti». Ciò, anche in conformità al diritto di difesa del cittadino, al quale non può essere preclusa la conoscenza delle modalità con cui è stata assunta una decisione destinata a ripercuotersi sulla sua sfera giuridica. Con le successive sentenze n. 8472 del 13 dicembre 2019 e n. 881 del 4 febbraio 2020 la VI sezione del Consiglio di Stato ha compiuto un ulteriore passo in avanti osservando che non vi sono ragioni di principio «per limitare l'utilizzo all'attività amministrativa vincolata piuttosto che discrezionale, entrambe espressione di attività autoritativa svolta nel perseguimento del pubblico interesse».

Inoltre, in entrambe le pronunce viene statuito che non può assumere rilievo la riservatezza³⁶ delle imprese produttrici dei meccanismi informatici utilizzati in quanto gli stessi, ponendosi al servizio del potere autoritativo, ne accettano le relative conseguenze in termini di necessaria trasparenza.

Il Collegio ha evidenziato, peraltro, che, in virtù di quanto previsto dal diritto nazionale ed europeo, emergono sostanzialmente tre principi da tenere in debita considerazione affinché l'utilizzo degli strumenti informatici avvenga in maniera idonea: il principio di conoscibilità e di comprensibilità dell'algoritmo; il principio della non esclusività della

³⁵ Sull'ammissibilità dell'atto automatizzato vincolato in dottrina si registra ormai da tempo un consenso pressoché unanime, vedi tra tutti L. Viola, *L'intelligenza artificiale nel procedimento e nel processo amministrativo: lo stato dell'arte*, in *federalismi.it*, 21, 2018, e G. Duni, *L'utilizzabilità delle tecniche elettroniche nell'emanazione degli atti e nei procedimenti amministrativi. Spunto per una teoria dell'atto emanato nella forma elettronica*, in *Rivista amministrativa della Repubblica Italiana*, CXXIX, 1978, 407 ss.

³⁶ Sulle questioni relative alla rilevanza del “codice sorgente” ai fini difensivi ed alle esigenze di riservatezza e di sicurezza informatica vedi anche TAR Lazio Roma, Sez. III Bis, 30 giugno 2020, n. 7370 e TAR Lazio Roma, Sez. III Bis, 1° luglio 2020 n. 7526.

decisione algoritmica: nel processo decisionale, infatti, deve esserci un'interazione con l'essere umano prima di produrre un risultato conformemente al modello cosiddetto HITLM (*human in the loop*)³⁷; il principio di non discriminazione algoritmica.

È stata proprio la giurisprudenza, in mancanza di un documento normativo che facesse esplicito riferimento alla possibilità di utilizzare l'Intelligenza Artificiale nel settore pubblico, ad aver individuato alcune restrizioni con riguardo alla possibilità di fare ricorso alla decisione amministrativa algoritmica.

Solo con il Nuovo Codice degli appalti - d.lgs. 36/2023 è stata finalmente positivizzata l'interpretazione giurisprudenziale sopra menzionata disponendo espressamente che le decisioni assunte mediante automazione devono rispettare i principi di conoscibilità e comprensibilità; non esclusività e "umanità" della decisione algoritmica; non discriminazione algoritmica (art.30).

In altri Stati europei, invero, il riferimento normativo alla decisione algoritmica amministrativa è stato introdotto tempo addietro.

In Spagna l'art. 30 della ley n. 11/2007 rubricato *Actuación administrativa automatizada* dispone che in caso di azione automatizzata, l'organismo o gli organismi competenti devono essere preventivamente istituiti, a seconda dei casi, per la definizione delle specifiche, la programmazione, la manutenzione, la supervisione e il controllo della qualità e, se del caso, l'*audit* del sistema di informazione e del suo codice sorgente. All'art. 41 della ley 40/2015 viene invece fornita una definizione dell'azione amministrativa automatica che consiste in qualsiasi atto o azione compiuta interamente per via elettronica da una Pubblica Amministrazione nell'ambito di un procedimento amministrativo e in cui un dipendente pubblico non sia intervenuto direttamente.

In Francia l'automazione della decisione amministrativa implica, secondo quanto previsto dal 2016 nel *Code des relations entre la public et l'administration*, che la Pubblica Amministrazione deve previamente informarne l'interessato e fornirgli contestualmente delle indicazioni sul trattamento dei propri dati.

Il legislatore tedesco³⁸, sempre nel 2016, ha previsto la possibilità di assegnare tutti gli atti della procedura amministrativa ad un *software* con l'introduzione dell'art. 35a nella legge generale sulle procedure amministrative (*Verwaltungsverfahrensgesetz*), il quale, nello specifico, dispone che un atto amministrativo può essere emesso interamente con mezzi automatizzati se ciò è consentito dalla legge e non vi è discrezionalità o margine di giudizio.

In prospettiva extraeuropea, l'Argentina, con il decreto legislativo 733/2018 del *Ministero de Modernización*, ha fatto per la prima volta riferimento all'utilizzo di decisioni automatizzate nel settore pubblico al fine «di costruire un governo più aperto e collaborativo che si adatti alla vita sempre più digitale e mobile dei cittadini nella società dell'informazione»³⁹.

³⁷ Si rimanda a M.L. Jones, *The right to a human in the loop: Political constructions of computer automation and personhood*, in *Social Studies of Sciences*, 47(2), 2021.

³⁸ E. Buoso, *Fully Automated Administrative Acts in the German Legal System*, in *European Review of Digital Administration & Law*, 1, 2020, 113 ss.

³⁹ Per un confronto fra Italia e Argentina sul tema dell'amministrazione digitale: D. U. Galetta - J. G. Corvalán, *Intelligenza Artificiale per una Pubblica Amministrazione 4.0? Potenzialità, rischi e sfide della rivoluzione tecnologica in atto*, in *federalismi.it*, 3, 2019.

3. Gli strumenti forniti dal GDPR

Pur non esistendo (fino a poco tempo fa) una normativa organica in materia, un richiamo alle decisioni automatizzate era già presente nella direttiva europea 95/46 del Parlamento Europeo e del Consiglio, relativa alla tutela delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati. In particolare, l'art. 15 della normativa dispone che «Gli stati membri riconoscono a qualsiasi persona il diritto di non essere sottoposta ad una decisione che produca effetti giuridici o abbia effetti significativi nei suoi confronti fondata esclusivamente su un trattamento automatizzato di dati destinati a valutare taluni aspetti della sua personalità, quali il rendimento professionale, il credito, l'affidabilità, il comportamento.»

Tuttavia, la ratio di tale disposizione è stata fortemente messa in discussione dalla interpretazione riduttiva che ne hanno dato le legislazioni nazionali nonché da un progressivo mutamento del significato attribuito al termine “decisione” a seguito del raffinarsi di tecniche di costruzione dei profili⁴⁰.

Il regolamento sulla protezione dei dati personali 2016/679, senza alcuna pretesa di completezza⁴¹, contiene alcune disposizioni relative all'utilizzo di specifiche tecniche di intelligenza artificiale.

L'art. 22 al par. 1 dispone in via generale che «L'interessato ha il diritto di non essere sottoposto a una decisione basata unicamente⁴² sul trattamento automatizzato, compresa la profilazione, che produca effetti giuridici che lo riguardano o che incida in modo analogo significativamente sulla sua persona».

Tuttavia, il secondo paragrafo prevede alcune eccezioni al predetto divieto: a) qualora il trattamento sia necessario per la conclusione o l'esecuzione di un contratto; b) se vi sia una disposizione del diritto dell'UE o di uno stato membro che lo autorizzi; c) in presenza del consenso dell'interessato.

Il terzo paragrafo precisa che nelle ipotesi di cui alle lettere a) e c) devono essere applicate misure appropriate per tutelare i diritti, le libertà e gli interessi legittimi dell'interessato, quali il diritto a richiedere l'intervento umano, il diritto di esprimere la propria opinione e di contestare la decisione.

Da ultimo, il par. 4 dispone che le decisioni di cui al par. 2 non possono basarsi sulle categorie particolari di dati personali di cui all'art. 9, par. 1, a meno che non vi sia il

⁴⁰ S. Rodotà, *Il diritto di avere diritti*, Bari, 2012, 328. Per la definizione di profilazione si veda art. 4, par. 4 del regolamento 2016/679: «qualsiasi forma di trattamento automatizzato di dati personali consistente nell'utilizzo di tali dati personali per valutare determinati aspetti personali relativi a una persona fisica, in particolare per analizzare o prevedere aspetti riguardanti il rendimento professionale, la situazione economica, la salute, le preferenze personali, gli interessi, l'affidabilità, il comportamento, l'ubicazione o gli spostamenti di detta persona fisica»

⁴¹ Così E. Mantovani, *Intelligenza artificiale e discriminazione: quali prospettive? Il modello inglese del data trust*, in *La Rivista Gruppo di Pisa*, 3, 2021, 373.

⁴² Come sottolinea E. Pellicchia, *Profilazione e decisioni automatizzate al tempo della black box society: qualità dei dati e leggibilità dell'algoritmo nella cornice della responsible research and innovation*, in *Le Nuove Leggi Civili Commentate*, V, 2018, 1224-1225 non vi sarebbe unanimità nell'interpretazione dell'avverbio «unicamente». Infatti, parte della dottrina ritiene che la tutela di cui all'art. 22 non possa estendersi a tutte le decisioni che comprendono l'ausilio dell'intervento umano (anche minimo), mentre altri sostengono che, ai fini di tale esonero, l'intervento umano deve essere significativo.

consenso dell'interessato (art. 9, par. 2, lett. a) o che il trattamento non sia necessario per motivi di interesse pubblico (art. 9, par. 2, lett. a).

Inoltre, pur avendo un'efficacia meramente interpretativa, il considerando 71 specifica che il titolare del trattamento deve utilizzare procedure appropriate per la profilazione e assicurare misure tecniche e organizzative adeguate al fine di garantire che siano «rettificati i fattori che comportano inesattezze dei dati e sia minimizzato il rischio di errori» e di impedire «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 che comportano misure aventi tali effetti».

La Corte di giustizia⁴³, pronunciandosi su un ricorso in materia di affidabilità creditizia, ha fornito alcuni criteri interpretativi al fine di precisare quali siano, di fatto, le decisioni automatizzate a cui può essere applicato l'art. 22 del regolamento (UE) 2016/679.

La Corte ha sottolineato che l'applicabilità dell'art. 22 è soggetta a tre condizioni cumulative: deve sussistere una decisione; la decisione deve essere «basata unicamente sul trattamento automatizzato, compresa la profilazione» e deve produrre «effetti giuridici (riguardanti l'interessato)» o incidere «in modo analogo significativamente sulla persona».

La definizione di decisione non è contenuta nel GDPR, tuttavia, secondo il ragionamento della Corte, dalla formulazione dell'art. 22 e del considerando 71 si può dedurre che il termine decisione «rinvia non solo ad atti che producono effetti giuridici riguardanti il soggetto di cui trattasi, ma anche ad atti che incidono significativamente su di esso in modo analogo⁴⁴». In altre parole, il Giudice Europeo individua come elemento dirimente, per classificare una decisione come tale, l'incidenza della valutazione sulla sfera personale degli interessati.

Sotto un ulteriore profilo, va rilevato che il considerando 63 del GDPR prevede il diritto dell'interessato ad ottenere informazioni sulla «logica cui risponde qualsiasi trattamento automatizzato dei dati e, almeno quando è basato sulla profilazione, alle possibili conseguenze di tale trattamento».

Ancora, il considerando 71 fa riferimento ad un diritto dell'interessato ad ottenere una spiegazione circa i processi decisionali automatizzati previsti dall'algoritmo.

Come noto, tali disposizioni non hanno un contenuto precettivo. Pertanto, si è cercato di ricavare un diritto alla spiegazione dell'interessato a partire dal combinato disposto degli artt. 22; 13, par. 2, lett. f); 14, par. 2, lett. g); 15, par. 1, lett. h) del GDPR.

Infatti, è ormai pacifico in dottrina, che l'interessato, a seguito di una decisione automatizzata assunta nei suoi confronti (di cui era stato previamente informato ex art. 22) può verificare ex post se sia stata effettivamente assunta, mediante l'esercizio del proprio diritto di accesso ex art 15 GDPR⁴⁵.

⁴³ CGUE, C-634/21, *OQ v Land Hessen* (2023).

⁴⁴ La Corte nel caso di specie ha precisato che la nozione può ricomprendere certamente «il risultato del calcolo della solvibilità di una persona sotto forma di tasso di probabilità relativo alla capacità di tale persona di onorare impegni di pagamento in futuro».

⁴⁵ G. Gaudio, *L'algorithmic management e il problema della opacità algoritmica nel diritto oggi vigente e nella Proposta di Direttiva sul miglioramento delle condizioni dei lavoratori tramite piattaforma*, in *Lavoro Diritti Europa*, 1, 2021.

Tuttavia, il termine *meaningful* (nel testo inglese del GDPR all'art. 15) deve essere interpretato non come la completa spiegazione del modello matematico sotteso al funzionamento dell'algoritmo, ma come un mero chiarimento che renda comprensibile all'interessato gli effetti che discendono dalla propria scelta⁴⁶.

Sull'interpretazione dell'art. 15 è opportuno menzionare le conclusioni dell'avvocato generale della Corte di giustizia europea Jean Richard De La Tour nella causa C-203/22 in cui il giudice del rinvio ha richiesto precisazioni su cosa vada inteso per «informazioni significative sulla logica utilizzata» nell'ambito di un processo decisionale automatizzato ai sensi dell'art. 15, par. 1, lett. h), del GDPR; se dette informazioni comprendano l'algoritmo utilizzato a tal fine e in che misura e con quale grado di concretezza si possa esigere dal titolare del trattamento che comunichi informazioni sufficienti per consentire all'interessato di verificare l'esattezza di dette informazioni e la loro coerenza con la decisione inerente al *rating* di cui trattasi.

In secondo luogo, ha chiesto in che misura la protezione dei segreti commerciali possa influire sull'obbligo del titolare del trattamento di fornire informazioni significative sulla logica sottesa nell'ambito di una decisione automatizzata e quali siano i meccanismi che possono consentire di risolvere tale eventuale conflitto.

L'Avvocato Generale ritiene che l'art. 15, par. 1, lett. h), del GDPR non possa essere interpretato «nel senso che fa gravare sul titolare del trattamento un obbligo di divulgare all'interessato informazioni che, in ragione del loro carattere tecnico, presentano un livello di complessità tale da non poter essere comprese dalle persone che non dispongono di una competenza tecnica particolare».

Aggiunge in seguito: «Si potrebbe certamente sostenere, in nome di una lettura estensiva dell'obbligo di trasparenza, che il controllo della modalità con cui i dati personali sono trattati da un algoritmo impone che quest'ultimo sia rivelato all'interessato. Tuttavia, ritengo che la ragion d'essere di tale obbligo sia quella di consentire a tale persona di comprendere le informazioni che le sono comunicate affinché quest'ultima possa far valere i diritti di cui gode a norma del RGPD. In tale ottica, spiegazioni accessibili che non presuppongono una particolare competenza tecnica sono certamente più “significative” di una formula matematica complessa.»

Sul secondo profilo, l'Avvocato Generale osserva che, qualora le informazioni che devono essere fornite all'interessato in virtù di quanto previsto dall'art. 15, par. 1, lett. h) possano comportare una lesione dei diritti e delle libertà altrui, segnatamente perché contengono dati personali di terzi tutelati da detto regolamento o un segreto commerciale, dette informazioni «devono essere comunicate all'autorità di controllo o all'organo giurisdizionale competenti affinché questi ultimi possano ponderare, con piena cognizione di causa e nel rispetto del principio di proporzionalità e della riservatezza di dette informazioni, gli interessi in gioco e stabilire la portata del diritto di accesso che deve essere riconosciuto a tale persona».

⁴⁶ E. Palmerini, *Decisioni algoritmiche e diritto dei dati*, in *giudicedonna.it*, 1-2, 2023,13; v. anche A.D. Selbst - J. Powles, *Meaningful information and the right to explanation*, in *International Data Privacy Law*, 7(4), 2014, 236.

4. AI Act: un argine alla discriminazione algoritmica?

Negli ultimi anni le istituzioni europee hanno mostrato una crescente attenzione al tema dell'intelligenza artificiale enfatizzando la necessità di una *Governance* integrata al fine di implementarne lo sviluppo in armonia con il quadro dei principi e dei diritti fondamentali condivisi dall'UE.

Nel febbraio 2017 il Parlamento Europeo ha approvato una risoluzione⁴⁷ invitando gli Stati membri a disciplinare in maniera omogenea gli aspetti civilistici della robotica ed ha elencato una serie di principi etici che dovrebbero permeare il quadro giuridico dell'Unione in quest'ambito, fra i quali: sicurezza, salute, libertà, vita privata, integrità, dignità, autodeterminazione, non discriminazione, protezione dei dati personali, nonché tutti i principi previsti dall'art. 2 del TUE e dalla Carta dei diritti fondamentali dell'UE.

Nel 2018 la Commissione Europea con le comunicazioni nn. 237 e 795 ha dichiarato di voler perseguire tre obiettivi principali in materia di intelligenza artificiale: dare impulso alla capacità tecnologica e industriale dell'Unione Europea e all'adozione dell'IA in tutti i settori economici; prepararsi ai cambiamenti socioeconomici ed assicurare un quadro etico e giuridico adeguato a tali fini.

Ancora, la Commissione Europea a giugno 2018 ha nominato un gruppo di esperti di alto livello che hanno redatto il documento «Orientamenti etici per un'intelligenza artificiale affidabile» pubblicato l'8 aprile 2019 in cui sono stati declinati sette requisiti per un'intelligenza artificiale sicura: intervento e sorveglianza umani; robustezza tecnica e sicurezza; riservatezza e *governance* dei dati; trasparenza; diversità, non discriminazione ed equità; benessere sociale e ambientale ed *accountability*.

Ha fatto seguito il Libro Bianco sull'intelligenza artificiale del 19 febbraio 2020⁴⁸ con lo scopo di definire «le opzioni strategiche» da seguire per «promuovere l'adozione dell'IA» ed «affrontare i rischi associati a determinati utilizzi di questa tecnologia». Nel documento la Commissione si interrogava sul possibile adeguamento del quadro legislativo vigente nell'UE alle applicazioni di intelligenza artificiale, ad esempio, della direttiva 2000/43/CE sull'uguaglianza razziale; della direttiva 2000/78/CE sulla parità di trattamento in materia di occupazione e di condizioni di lavoro, delle direttive 2004/113/CE; 2006/54/CE sulla parità di trattamento tra uomini e donne per quanto riguarda l'accesso a beni e servizi in materia di occupazione.

Pertanto, dal testo emerge chiaramente, ancor prima dell'adozione di nuove regole, la necessità di una valutazione delle suddette normative in tema di diritti fondamentali al fine di appurare se le stesse possano dirsi applicabili anche alle problematiche derivanti dai sistemi di IA, atteso che l'elasticità che contraddistingue gli ordinamenti giuridici permette, tendenzialmente, di includervi i mutamenti della società.⁴⁹

I numerosi atti di impulso e di *soft law* in materia di intelligenza artificiale hanno con-

⁴⁷ Parlamento europeo, Risoluzione del 16 febbraio 2017 recante raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica (2015/2103(INL)).

⁴⁸ Commissione europea, Libro Bianco sull'intelligenza artificiale – Un approccio europeo all'eccellenza e alla fiducia, COM (2020) 65, 19 febbraio 2020.

⁴⁹ G. Finocchiaro, *Intelligenza Artificiale: quali regole?*, cit., 18.

dotto, a seguito di un lungo processo di negoziazione fra gli Stati membri, alla formulazione di una regolamentazione organica in materia, l'*Artificial Intelligence Act*⁵⁰, che è stata approvata da una maggioranza di 523 voti favorevoli, contro 46 voti contrari e 49 astensioni.

Va ritenuta opportuna la scelta di affidare la regolazione di tale ambito ad un regolamento, atto di portata europea e di applicazione diretta, così da evitare la frammentazione del mercato unico e l'incertezza giuridica che sarebbe derivata da una regolamentazione esclusivamente nazionale⁵¹.

La normativa segue un approccio basato sul rischio individuando quattro differenti intensità: rischio inaccettabile, rischio elevato, rischio limitato, rischio minimo o nullo. Le decisioni assunte dai sistemi di intelligenza artificiale in settori particolarmente sensibili, oggetto della seguente trattazione, possono ricondursi alle classi dei sistemi di IA «a rischio inaccettabile» o «ad alto rischio».

Infatti, l'art. 5, lett. c), fra le pratiche vietate prevede l'immissione sul mercato, la messa in servizio o l'uso di sistemi per la valutazione o la classificazione di persone o di gruppi sulla base del loro comportamento sociale o delle loro caratteristiche personali. Sono, altresì, vietati, ai sensi dell'art.5, lett. d), quei sistemi che misurano la probabilità che un individuo commetta un reato sulla base dei suoi tratti somatici o sulle caratteristiche della sua personalità⁵².

Peraltro, l'allegato III del regolamento elenca i settori in cui si sviluppano i sistemi di IA ad alto rischio, fra i quali: istruzione e formazione professionale; occupazione, gestione dei lavoratori e accesso al lavoro autonomo; accesso a servizi privati essenziali e a prestazioni e servizi pubblici essenziali e fruizione degli stessi; attività di contrasto; migrazione, asilo e gestione delle frontiere; amministrazione della giustizia e processi democratici.

Tuttavia, come precisa l'art. 6, par. 3, «un sistema di IA non rientra nella categoria ad alto rischio se non presenta un rischio significativo di danno per la salute, la sicurezza o i diritti fondamentali delle persone fisiche».

L'utilizzo di un sistema di questa tipologia implica la previa identificazione dei rischi ragionevolmente prevedibili e l'adozione di misure necessarie ad evitarli. In particolare, deve essere effettuato un controllo dei dati utilizzati onde evitare distorsioni o discriminazioni vietate dal diritto dell'Unione Europea. Inoltre, ogni sistema ad alto rischio ha l'obbligo di detenere una documentazione tecnica per dimostrare la conformità ai requisiti previsti dal regolamento e deve registrare l'attività effettuata al fine di garantirne la tracciabilità.

A seguire, l'art. 13 dispone che tali sistemi devono essere progettati in modo tale da consentire ai *deployer*⁵³ di poter comprendere l'*output* garantendo una «trasparenza ade-

⁵⁰ Regolamento (UE) 2024/1689.

⁵¹ In questi termini già la 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)).

⁵² La disposizione prevede un'eccezione: «tale divieto non si applica ai sistemi di IA utilizzati a sostegno della valutazione umana del coinvolgimento di una persona in un'attività criminosa, che si basa già su fatti oggettivi e verificabili direttamente connessi a un'attività criminosa».

⁵³ Per la definizione di *deployer* si rinvia all'art 3. n. 4, dell'*AI Act*: «“*deployer*”: persona fisica o giuridica,

guata». Il regolamento individua puntualmente le informazioni che il sistema ad alto rischio deve fornire: l'identità e il contatto del fornitore, la finalità del sistema, il livello di accuratezza che il sistema può garantire nonché gli eventuali rischi prevedibili per la salute, per la sicurezza o per i diritti fondamentali, le informazioni circa le caratteristiche tecniche del sistema al fine di comprenderne opportunamente l'*output*, eventuali informazioni circa le prestazioni del sistema con riguardo a determinate persone o gruppi di persone, le misure di sorveglianza umana.

L'art.14 dell'AI Act dispone che i sistemi ad alto rischio devono essere progettati in modo da garantirne la supervisione da parte di persone fisiche: il fine della sorveglianza umana è quello di prevenire e ridurre al minimo i rischi per la salute, la sicurezza e i diritti fondamentali.

La normativa precisa che le persone fisiche a cui è affidata la sorveglianza devono essere ben consapevoli della gravosa influenza esercitata dal sistema di intelligenza artificiale sulla propria decisione e della tendenziale propensione a fare affidamento sull'*output* suggerito dalla macchina.

Sul ruolo ausiliare della persona fisica rispetto ai sistemi di intelligenza artificiale la dottrina è divisa: vi è chi sostiene che quando verranno introdotte soluzioni alle problematiche dei sistemi di apprendimento automatico ovvero quando non si correrà più il rischio di creare *biases* o meccanismi decisionali opacizzati le decisioni potranno essere del tutto affidate ai sistemi di intelligenza artificiale che, molto più dell'essere umano, sono connotati da coerenza e imparzialità; altri ritengono, invece, che i sistemi di intelligenza artificiale impiegati in ambiti particolarmente sensibili dovranno in ogni caso mantenere una componente umana⁵⁴.

Nel regolamento si specifica che le persone fisiche alle quali è affidata la sorveglianza umana devono interpretare l'*output* del sistema tenendo conto di tutti gli strumenti a loro disposizione potendo eventualmente scegliere di non usare il sistema di intelligenza artificiale, di ignorarne la decisione o di intervenire sul funzionamento arrestandone la procedura.

L'art. 15, peraltro, prevede che tali applicazioni di intelligenza artificiale devono conformarsi ad un adeguato livello di accuratezza, robustezza e cybersicurezza.

Inoltre, gli artt. 13, 14, 15 stabiliscono che i sistemi ad alto rischio devono essere «progettati» sin dal principio in modo tale da garantire la piena conformità alle suddette disposizioni e, in adesione al principio cd. *Ethics by design*, all'art. 27 è previsto che, prima di utilizzare un sistema di IA ad alto rischio, i *deployer* hanno l'onere di effettuare una valutazione dell'impatto sui diritti fondamentali che l'uso di tale sistema potrà produrre. Ciò a dimostrazione che, in linea con quanto era già stato stabilito dall'art. 25 GDPR (rubricato *Data protection by design and by default*), anche nel Regolamento sull'intelligenza artificiale si predispone una «nuova collocazione delle regole rispetto al fenomeno da regolare» al fine di includere già nella fase della progettazione i valori e i principi fon-

autorità pubblica, agenzia o altro organismo che utilizza un sistema di IA sotto la propria autorità, tranne nel caso in cui il sistema di IA sia utilizzato nel corso di un'attività personale non professionale».

⁵⁴ C. Casonato - B. Marchetti, *Prime osservazioni sulla proposta di regolamento dell'unione europea in materia di intelligenza artificiale*, in *Rivista di BioDiritto*, 3, 2021, 18.

damentali dell'Unione europea⁵⁵.

5. Qualche considerazione conclusiva

Come è stato osservato da autorevole dottrina⁵⁶, il regolamento si limita ad individuare delle soluzioni «formali» richiamando i principi generali condivisi dagli stati membri dell'UE e senza, di fatto, introdurre «nuovi efficaci e rapidi strumenti di tutela contro la discriminazione».

Tuttavia, proprio per la natura sfuggibile della materia oggetto della normativa, che quasi risulta «refrattaria alla giuridificazione»⁵⁷ e sembra «atteggiarsi a ordine spontaneo»⁵⁸, possono ragionevolmente comprendersi le difficoltà sottese alla definizione di un quadro regolatorio specifico.

Per quel che qui maggiormente interessa, va rilevato che l'AI Act non fa menzione delle numerose direttive europee concernenti il diritto antidiscriminatorio. Pare, dunque, che la normativa, almeno per quanto concerne le discriminazioni algoritmiche, non discenda da una consona valutazione (e da una conseguente possibile integrazione) degli strumenti di contrasto già a disposizione, come, peraltro, aveva suggerito anche la Commissione Europea nel Libro Bianco sull'Intelligenza Artificiale.

In ogni caso, per fare un bilancio complessivo si devono attendere gli interventi legislativi e amministrativi degli stati membri in attuazione del Regolamento.

Preme rilevare che l'ordinamento italiano con riguardo alle procedure decisionali automatizzate non prevede disposizioni che regolano appositamente il fenomeno se non, relativamente all'azione amministrativa, quanto contenuto nel Nuovo Codice degli Appalti.

Mentre, come è emerso nel corso della trattazione, altri stati europei e non europei già da molto tempo hanno introdotto delle normative di settore sulle decisioni algoritmiche.

Dunque, in Italia, è stata fondamentale l'interpretazione in via giurisprudenziale che, nel suo ruolo di supplente e in conformità ai principi generali nazionali e sovranazionali (che sono stati ribaditi anche nell'AI Act), ha fissato alcuni limiti in materia di decisione automatizzata, e, come visto, ricorre nelle varie pronunce il diritto dell'interessato alla conoscibilità e alla comprensibilità della logica sottesa alla procedura decisionale.

D'altra parte, però, è pur vero che il titolare del trattamento potrebbe non disvelare la logica sottesa all'algoritmo celandosi dietro alla sua estrema complessità. In tal senso, oltre all'interpretazione fornita dall'Avvocato Generale della Corte di giustizia con riguardo all'art. 15 GDPR, si direziona anche l'art. 13 dell'AI Act che dispone che, rispetto al sistema di IA ad alto rischio utilizzato, venga garantita una trasparenza «adeguata» e non completa.

Premesso ciò, ci si domanda quale rilevanza possa avere l'omissione di tale informazio-

⁵⁵ G. Lo Sapia, *La black box: l'esplicabilità delle scelte algoritmiche quale garanzia di buona amministrazione*, cit., 127.

⁵⁶ G. Finocchiaro, *Intelligenza artificiale: quali regole?*, cit., 123ss.

⁵⁷ A. Celotto, *Algoritmi e algoritica: quali regole per l'intelligenza artificiale?*, in *ConsultaOnline*, 27 marzo 2020, 9.

⁵⁸ *Ibid.*

ne sull'esercizio di difesa dell'interessato che nella procedura di decisione automatizzata -il cui meccanismo è a lui sconosciuto- potrebbe essere stato esposto ad una discriminazione algoritmica, considerando, oltretutto, che generalmente nemmeno il giudice ha le competenze per comprendere se la profilazione di un sistema di intelligenza artificiale possa nascondere nella incomprensibilità del suo funzionamento eventuali limiti o errori di impostazione.