Annali di studi religiosi
ISBN 979-12-218-2156-7
DOI 10.53136/979122182156730
pp. 539-558 (ottobre 2025)

L'INTELLIGENZA ARTIFICIALE ALLA PROVA DELL'EGUAGLIANZA L'EVOLUZIONE DEL DIRITTO COSTITUZIONALE TRA SCENARI ATTUALI E FUTURI

Marta Fasan

Abstract: In the context of the ongoing changes and evolution of the current era, this paper investigates the impact of Artificial Intelligence (AI) technologies on certain structural elements of society. It analyzes how Constitutional Law must adapt to face new and unprecedented challenges. Specifically, this examination focuses on the emergence of new protection needs within what is defined as an "algorithmic" society. The paper approaches this issue through the lens of the principle of equality and raises the question of which tools can be used to address the forms of discrimination that AI can create. In light of these considerations, this contribution aims to explore potential new interpretations of the principle of equality, with the goal of achieving an AI that is both human–centric and rights–based.

Nel contesto del continuo mutare ed evolversi dell'epoca attuale, il presente contributo indaga i cambiamenti dettati dall'avvento delle tecnologie di Intelligenza Artificiale (IA) su alcuni elementi strutturali della società e analizza come il diritto costituzionale sia chiamato ad affrontare nuove inedite sfide. Più nello specifico, il contributo esamina l'emergere di nuove esigenze di tutela all'interno dell'attuale modello di società, definibile "algoritmica", nella prospettiva dell'applicazione del principio di eguaglianza, interrogandosi su quali strumenti possano applicarsi alle forme di discriminazione create dall'intelligenza artificiale. Alla luce di ciò, il contributo prova a riflettere su quali possano essere le nuove declinazioni da dare al principio di eguaglianza nell'ottica di raggiungere l'obiettivo di un'IA *human–centered* e *rights–based*.

Keywords: Artificial Intelligence, Constitutional law, Bias, Principle of equality, Discrimination

Parole chiave: Intelligenza Artificiale, Diritto costituzionale, Bias, Eguaglianza, Discriminazione

1. L'intelligenza artificiale nella società contemporanea. Prime coordinate di una rivoluzione algoritmica

Nell'epoca attuale la società è chiamata ad affrontare numerose sfide e mutamenti, spesso caratterizzati da effetti di portata globale. L'insorgere di guerre e conflitti armati, l'aumento dei fenomeni migratori, le conseguenze del cambiamento climatico, il ritorno di eventi pandemici, l'acuirsi delle discriminazioni e la crescita dei divari economico–sociali, come nel caso di *Global North* e *Global South* (Arrighi 2001), sono tutti fenomeni che stanno mettendo fortemente alla prova la società nel suo insieme, portando cambiamenti significativi al suo interno anche in termini di dinamiche relazionali.

Tra questi avvenimenti, però, uno risulta particolarmente significativo alla luce dell'impatto che sta dimostrando di poter produrre sulla società contemporanea, tanto da poterlo considerare una tra le cifre caratterizzanti il periodo storico attuale: l'avvento dell'intelligenza artificiale (d'ora in poi, IA).

Lo sviluppo e l'uso delle nuove tecnologie di IA sono, infatti, fenomeni che negli ultimi anni hanno assunto un'importanza crescente nella quotidianità delle persone, investendo le sfere della vita privata, professionale e pubblica di ognuno di noi. Tale assunto è confermato dall'ampia varietà di ambiti e di scopi d'uso rispetto a cui trovano applicazione i sistemi di IA (Scherer 2016, p. 354). Così, e a titolo esemplificativo, questa tecnologia viene impiegata nella Pubblica Amministrazione; nel settore medico; nelle attività di amministrazione della giustizia; nelle azioni di polizia e di tutela della pubblica sicurezza; nel contesto bellico e militare; in ambito economico–finanziario; nel controllo delle frontiere e dei flussi migratori; nel campo dell'istruzione; nelle attività di ricerca; nelle politiche del lavoro; nel settore dell'informazione; nei meccanismi elettorali del circuito democratico e in ogni altro contesto in cui le capacità tecniche dell'IA trovano terreno fertile per una loro implementazione. Le ragioni di una diffusione su così larga scala (Sartor e Lagioia 2020), ancor più accentuata dalla presenza sul mercato di modelli di AI per finalità generali come ChatGPT, Grok e DeepSeek (Sarker 2024; Wangsa *et al.* 2024; Hong e Hu 2025), possono essere facilmente spiegate alla luce della tipologia

di funzioni tipicamente esercitate dai sistemi di IA. Se, infatti, si prendono in considerazione le definizioni di IA date dai recenti atti normativi adottati per regolare lo sviluppo e l'uso di questa tecnologia, secondo cui ciò che caratterizza questa tecnologia è la capacità, con un certo grado di autonomia e in base a obiettivi predefiniti, di trarre inferenze dai dati input per generare risultati quali previsioni, raccomandazioni e decisioni che possono influenzare l'ambiente circostante[1], risulta evidente come una tecnologia in grado di svolgere simili funzioni possa essere impiegata in un'ampia gamma di settori, anche molto differenti tra loro. Di fatto, ogni volta che sia richiesto lo svolgimento di un'attività decisionale, o preparatoria alla stessa, i sistemi di IA possono trovare applicazione in sostituzione o in supporto all'attività umana grazie alle capacità tecniche che sanno esprimere (Santosuosso e Sartor 2024).

Tale circostanza pone di fronte a una situazione finora inedita, o quantomeno fatta emergere in modo manifesto dalle specifiche funzionalità di questi sistemi tecnologici. L'IA, infatti, sta dimostrando la capacità di incidere in modo sempre più pervasivo sui meccanismi di funzionamento delle organizzazioni sociali umane e di plasmare, quindi, la società anche nei suoi elementi fondanti (Rodotà 2012). Ciò attribuisce alle tecnologie di IA la capacità, o meglio il potere, di definire le caratteristiche del complessivo assetto sociale in cui viene a essere inserita, tanto da poter prospettare l'emergere di un nuovo modello di società forgiato secondo la matrice algoritmica e digitale che contraddistingue le nuove tecnologie intelligenti (Bassini, Liguori e Pollicino 2018).

2. La società algoritmica e le nuove sfide per il diritto costituzionale

L'impiego diffuso dei sistemi di IA e l'incisività dei loro effetti sono, quindi, fattori che stanno contribuendo a modificare sostanzialmente le strutture organizzative sociali, dettando mutamenti anche radicali come la nascita di un nuovo modello di società. In particolare, negli ultimi anni si è affermata l'idea che si stia assistendo alla creazione di una nuova tipologia di società a opera delle tecnologie intelligenti, definita, per

(1) Questa la definizione di sistema di IA data all'art. 3, n. 1), Regolamento (UE) 1689/2024, meglio noto come AI Act.

l'appunto, "società algoritmica". Con questo termine si fa riferimento a una società "[…] organized around social and economic decision–making by algorithms, robots, and AI agents, who not only make the decisions but also, in some cases, carry them out" (Balkin 2017), e, cioè, a un modello sociale in cui i sistemi di IA non si limitano a supportare i processi decisionali, ma ne diventano i principali artefici, assumendo un ruolo primario anche nella loro implementazione. Due sono, infatti, le caratteristiche che contraddistinguono la società algoritmica: la presenza di piattaforme digitali multinazionali che si collocano in posizione intermedia tra gli Stati nazionali e i cittadini; e l'uso dei sistemi di IA quale strumento idoneo a governare le popolazioni (Balkin 2018).

Si è, quindi, in presenza di un'inedita tipologia di dinamiche organizzative e relazionali costruite intorno alla dimensione algoritmica ormai prevalente, in cui le decisioni tendono a essere sempre più determinate dall'uso dei sistemi di IA e dai risultati che questi producono (Fasan 2024).

Questo mutato scenario, in cui la società contemporanea si dimostra sempre più strutturalmente condizionata e condizionabile dall'impiego delle tecnologie intelligenti, porta con sé una conseguenza ulteriore, cioè l'emergere di nuove declinazioni del potere sia con riferimento al rapporto Stato–cittadino sia in relazione alla classica distinzione tra potere pubblico e potere privato (Graziani 2021). L'IA, infatti, costituisce espressione di potere, sia come strumento che ne consente l'esercizio, sia come soggetto dello stesso. Da un lato, i sistemi intelligenti esprimono il potere dei soggetti privati attivi nello sviluppo dei sistemi di IA e dei decisori pubblici che vi fanno ricorso per migliorare il governo della cosa pubblica; dall'altro lato, l'IA manifesta una capacità di influenzare, indirizzare e definire i comportamenti delle persone che aumenta con l'ampliarsi della sua autonomia d'azione, tanto da renderla, in alcune circostanze, un soggetto decisionale autonomo (Sulmicelli 2024).

Il riconoscimento nell'IA di una nuova manifestazione di potere sovrano, in grado di scardinare anche le più classiche categorizzazioni[2] (Simoncini e Cremona 2021), influisce anche sulla dimensione giuridica

(2) Si fa qui riferimento al venire meno della classica distinzione giuridica tra dimensione privata e dimensione pubblica, da sempre considerate due sfere d'azione e di regole tendenzialmente separate.

e, in particolare, del diritto costituzionale. Se il diritto costituzionale conserva la sua funzione ontologica "[…] di "misura" del potere e "fondamento" della sovranità" (Simoncini 2017), risulta evidente che le tecnologie di IA, e così anche i risultati e gli effetti che producono, costituiscono una nuova sfida a cui fare fronte. Nell'esercizio del potere di cui è manifestazione, l'IA mostra la capacità di incidere in modo significativo sulla tutela dei diritti fondamentali delle persone, richiamando, in questi termini, l'originale vocazione del costituzionalismo contemporaneo, e cioè assicurare la necessaria limitazione dei poteri in funzione di garantire protezione ai diritti e alle libertà delle persone (Fioravanti 2009; Casonato 2025).

Tuttavia, le nuove esigenze che nascono dalla mutata realtà non sempre riescono a trovare soluzione nelle categorie e negli strumenti giuridici tradizionali che il diritto costituzionale ha a disposizione.

Per questo motivo occorre riflettere su quali possano essere le strade da intraprendere, dal punto di vista giuridico, per garantire piena attuazione e tutela agli elementi che fondano il costituzionalismo anche nel contesto della nuova conformazione attribuibile alla società algoritmica (Celeste 2022).

3. L'intelligenza artificiale alla prova dell'eguaglianza: la manifestazione dei bias secondo una prospettiva socio–tecnologica

La comprensione del fenomeno in esame può essere agevolata analizzando l'impatto prodotto dai sistemi di IA su una delle categorie cardine del costituzionalismo contemporaneo: il principio di eguaglianza.

Questo principio, che tradizionalmente assume un ruolo essenziale nella garanzia e nella tutela dei diritti fondamentali di fronte all'esercizio del potere (Caretti 2002; Baer 2012), risulta particolarmente sensibile ai cambiamenti dettati dalla diffusione dei sistemi di IA, che, per l'appunto, ne mettono in discussione l'accezione classica e la concreta implementazione. Da questo punto di vista, osservare il modo in cui l'IA incide sulla tenuta del principio di eguaglianza, e sull'attuazione delle garanzie che ne costituiscono un'esternazione, offre una dimostrazione emblematica di quali sfide il diritto costituzionale sia

chiamato ad affrontare per continuare ad assicurare tutela alle persone, anche di fronte ai rischi prospettati dalle tecnologie di IA (Sulmicelli 2024; Fasan 2022).

In particolare, il principio di eguaglianza viene chiamato in causa dalla necessità di predisporre soluzioni ai bias presenti nei sistemi di IA e alle conseguenze discriminatorie che ne derivano.

Quale frutto dell'azione umana connotato dal punto di vista sociale (Orlikowski 1992), l'IA porta con sé i pregiudizi e gli squilibri di potere presenti nella società contemporanea. Questi, che per l'appunto vengono inquadrati nella comune categoria di bias (Barocas e Selbst 2016), possono innestarsi e manifestarsi, anche inconsciamente, in varie fasi del ciclo di funzionamento e di vita dei sistemi di IA. In particolare, tre sono le circostanze in cui è possibile esaminare con maggiore efficacia l'emergere di tale fenomeno.

Il primo momento in cui i bias possono insorgere è la creazione del dataset di addestramento e di validazione del sistema di IA. Nel contesto di una *datafication* sempre più significativa della società umana (Lycett 2013) che risulta essenziale per lo sviluppo di sistemi di IA sempre più avanzati, anche i pregiudizi e le situazioni di squilibrio tra gruppi di persone entrano a far parte del flusso di informazioni utilizzati dai sistemi di IA per il loro funzionamento, condizionando i risultati prodotti. In questi termini, il bias presente nel dataset viene classificato come un *pattern* rilevante dall'IA, con la conseguenza di sistematizzare il pregiudizio di cui è espressione e di replicarne il contenuto nell'elaborazione della decisione finale (Barocas e Selbst 2016). Tale eventualità, che interessa sia le forme di c.d. "model–based AI" (Traverso 2022) sia i sistemi di IA più avanzati, è concretamente dimostrata da un caso che, seppur risalente nel tempo, illustra chiaramente le conseguenze di queste forme di *historical bias*. Il caso, del 1988, riguarda lo sviluppo da parte della St. George's Hospital Medical School di un sistema algoritmico con funzioni di supporto nella selezione dei propri futuri studenti. L'impiego del sistema, che avrebbe dovuto aiutare il personale responsabile delle procedure di selezione eliminando informazioni considerate irrilevanti, viene, tuttavia, giudicato illegittimo da parte della Commission for Racial Equality. In particolare, la Commissione dichiarò la St. George's Hospital Medical School responsabile di aver praticato condotte discriminatorie nelle sue politiche

di ammissione nei confronti delle donne e degli appartenenti a gruppi etnici minoritari, anche in ragione della tecnologia usata (Commission for Racial Equality 1988). Infatti, l'algoritmo, basandosi sui criteri di selezione precedentemente adottati dalla *Medical School*, aveva continuato a riprodurre i pregiudizi che da sempre avevano portato all'esclusione dei candidati e delle candidate in ragione del genere e dell'etnia di appartenenza (Hacker 2018).

La seconda circostanza in cui i bias si manifestano è, invece, quella della realizzazione dell'algoritmo, o meglio della definizione dei criteri di inferenza impiegati dall'IA. Anche in questa circostanza, infatti, i pregiudizi e i preconcetti umani possono entrare a far parte dei meccanismi di funzionamento tecnologico e definire, sulla base di ciò, il peso da attribuire ai diversi valori che contribuiscono a creare il modello di analisi ed elaborazione delle informazioni implementato dall'IA. Anche in questo caso, quindi, il bias diventa un elemento strutturale delle operazioni svolte dell'IA, con il rischio, per effetto di un meccanismo di *feedback loop* (Barocas, Hardt e Narayanan 2023), di acuire e rafforzare pratiche pregiudizievoli e discriminatorie a danno di specifiche categorie di persone. Ciò, infatti, è quanto accaduto con il sistema SyRI, utilizzato dalla Pubblica Amministrazione olandese per profilare indentificare il rischio di commissione di una frode pubblica da parte delle persone richiedenti sussidi statali. Il sistema algoritmico, oltre a presentare significativi problemi in termini di dovute garanzie di trasparenza (van Bekkum e Zuiderveen Borgesius 2021; Avanzini 2022), veniva impiegato solo a fronte di richieste provenienti da abitanti di zone e quartieri considerati problematici. Nell'opinione delle autorità pubbliche olandesi, in questi luoghi sarebbe stata, infatti, più probabile la realizzazione di illeciti di natura fiscale e previdenziale in quanto riconducibili a indirizzi di persone già indagate per la commissione di questi illeciti (Falletti 2022). Perciò, il sistema SyRI, correlando i dati esaminati, di fatto aveva il potere di escludere le persone dall'accesso ai sussidi statali anche solo in ragione del loro status economico e geografico, di fatto discriminandole rispetto alle altre persone interessate a richiedere simili prestazioni sociali (Rachovitsa e Johann 2022).

Il terzo momento in cui i bias umani emergono in relazione al funzionamento dell'IA è quello che concerne la produzione dei risultati.

Possono verificarsi, infatti, situazioni in cui, pur avendo operato nelle fasi operative precedenti per rimuovere la presenza di eventuali elementi pregiudizievoli e parziali dal sistema, questo arriva a produrre un risultato in cui si configurano dei bias. Ciò implica l'adozione di decisioni pregiudizievoli, non tanto a causa della riproduzione di bias o di elementi discriminatori nel corso del processo decisionale, ma in ragione di altri fattori che incidono sui risultati prodotti dai sistemi di IA, andando a generare un impatto negativo sulle persone interessate (Sulmicelli 2025). Così, recentemente si è constatato come la scarsa efficacia dimostrata dai sistemi di IA sviluppati per il riconoscimento, la produzione, l'analisi e la traduzione del linguaggio dei segni sia causata principalmente dal mancato coinvolgimento della comunità non udente nella loro realizzazione. Tale carenza, infatti, limita gli sviluppatori nel possedere reale contezza dei problemi posti dall'IA, delle circostanze in cui la tecnologia potrebbe offrire valide soluzioni ai problemi esistenti e del modo in cui la lingua dei segni viene usata per comunicare dalle persone non udenti. La mancanza di consapevolezza e di informazioni nei termini descritti compromette l'utilità effettiva di questi potenziali strumenti e produce l'effetto di discriminare ulteriormente proprio coloro che dovrebbero averne beneficio (Bragg *et al*. 2019). Effetti simili si sono registrati anche con riferimento alle persone appartenenti alla comunità LGBTQ+. Per esempio, alcuni studi hanno dimostrato le conseguenze discriminatorie, in termini di risultati, derivanti dall'uso di ImageNet, uno dei *training set* di immagini più importanti nel settore dell'IA. Questo dataset si fonda su un sistema di classificazione delle immagini in cui le persone rappresentate vengono qualificate in "oggetti naturali" e poi ulteriormente distinte in "corpi maschili adulti" e "corpi femminili adulti". In ragione di ciò, e come sostenuto da parte della dottrina, i sistemi di IA contribuiscono a operare un processo di naturalizzazione del genere come costrutto biologico da intendersi solo in termini binari e a rendere le persone trans e di genere non binario invisibili, e quindi discriminabili, all'interno della realtà algoritmica (Crawford 2021; Keyes 2018). Effetti simili si verificano anche in contesti più specifici di applicazione dei sistemi di IA, come accade nelle attività di *content–moderation*. È questo il caso del sistema *Perspective*, impiegato per rilevare il livello di tossicità dei contenuti presenti sui social

network. Nello specifico, si è dimostrato come il sistema di IA classifichi i contenuti inseriti dalle drag queen e dalle persone LGBTQ+ con un livello di tossicità maggiore rispetto a quelli elaborati dai suprematisti bianchi (Oliva, Antonialli e Gomes 2021)[3]. Le ragioni di ciò sono riconducibili principalmente a due fattori. Il primo si sostanzia nella mancata presa in considerazione, in fase di sviluppo tecnologico, del significato identitario che assumono determinate parole e affermazioni per le persone LGBTQ+, denunciando un vuoto di rappresentatività nei dataset utilizzati per l'addestramento e la validazione dell'algoritmo. Il secondo fattore, invece, è riconducibile all'incapacità del sistema di IA di comprendere le differenze di contesto e di finalità con cui possono essere utilizzate le parole, e ciò, ancora una volta, per una mancata valorizzazione delle istanze che provengono da gruppi di persone considerati minoritari (Fosch–Villaronga e Poulsen 2022). Ciò che emerge dai casi esaminati è, dunque, la presenza di bias significativi nei risultati generati dall'IA che hanno ripercussioni discriminatorie nei confronti di persone già "vulnerabilizzate" dalle strutture di potere che permeano la società (Fineman 2019), e che possono incidere ulteriormente in modo negativo sull'esercizio dei diritti e delle libertà fondamentali riconosciute alle persone (Sulmicelli 2023)[4].

Alle tre circostanze descritte ne va poi aggiunta una quarta in cui nel ciclo di vita di un sistema di IA possono emergere bias dalle conseguenze discriminatorie. Questo momento, caratterizzato dalla trasversalità con cui può manifestarsi nelle diverse fasi di sviluppo e di applicazione dei sistemi intelligenti, si verifica nei casi in cui i pregiudizi riprodotti dall'IA non siano frutto dell'impiego di informazioni ascrivibili a categorie protette, ma bensì derivino dal ricorso a dati sostitutivi (c.d. *proxy data*). Nello specifico, si fa riferimento all'ipotesi in cui il sistema di IA, pur non avendo a disposizione informazioni e dati che permetterebbero di classificare le persone secondo criteri da considerarsi discriminatori alla luce

(3) In particolare, *Perspective* ha classificato con un livello di tossicità compreso tra il 16,68% e il 37,81% i contenuti generati dagli appartenenti alla comunità queer, mentre quelli elaborati da account rappresentativi di istanze nazionaliste e di suprematismo bianco sono stati inseriti in un range di tossicità che va dal 19,68% al 33,46%.

(4) In particolare, Sulmicelli evidenzia come le conseguenze delle discriminazioni descritte siano tali da compromettere l'esercizio anche di altri diritti e libertà fondamentali tutelate dall'ordinamento costituzionale, come, ed è questo il caso delle attività di content–moderation, l'esercizio della libertà di manifestazione del pensiero.

dell'ordinamento costituzionale vigente, arriva comunque a generare situazioni di discriminazione sulla base di informazioni di per sé apparentemente neutre (Prince e Schwarcz 2020). Ciò significa che anche qualora venga eliminata dal *dataset* di riferimento la presenza di dati potenzialmente discriminatori, le tecniche e le capacità di inferenza impiegate dai sistemi di IA potrebbero comunque risalire indirettamente a tali informazioni e perpetrare implicitamente le disparità di trattamento che ne sono conseguenza (Kleinberg *et al.* 2018). Questa nuova modalità di discriminazione, definita per l'appunto *proxy discrimination*, può essere esemplificata da quanto accaduto nel Regno Unito a seguito dell'implementazione del sistema DCP (*Direct Center Performance Model*). L'algoritmo in esame, sviluppato e utilizzato nel 2020 per elaborare una previsione standardizzata del rendimento scolastico degli studenti inglesi per indirizzarne l'accesso agli studi universitari[5], nelle procedure di valutazione delle prestazioni scolastiche ha dimostrato di classificare come meno capaci gli studenti provenienti da scuole pubbliche e a maggior numero di iscritti a parità di rendimento personale rispetto ai coetanei frequentanti scuole private (Kelly 2021). In questo modo, l'uso del sistema DCP e dei suoi risultati ha impedito l'accesso alle università più note e prestigiose a quegli studenti che, pur meritevoli dal punto di vista scolastico, appartenevano a nuclei familiari economicamente svantaggiati e/o a gruppi etnici minoritari, rendendoli destinatari di decisioni dal carattere discriminatorio (Wachter, Mittelstadt e Russell 2021).

4. L'intelligenza artificiale alla prova dell'eguaglianza: soluzioni e limiti secondo la prospettiva delle categorie giuridiche tradizionali

Le modalità descritte con cui possono insorgere elementi pregiudizievoli nelle diverse fasi di sviluppo e di applicazione dell'IA portano, dunque, a interrogarsi su quali siano i rimedi azionabili per contrastare

(5) Nello specifico, il governo britannico (in particolare l'*Office of Qualifications and Examinations Regulations*, Ofqual) ha scelto di utilizzare questo algoritmo per due ragioni: in primo luogo, per la necessità di avere uno strumento di valutazione che sostituisse i c.d. *A–Level examination*, sospesi nel 2020 a causa della pandemia da Covid–19; in secondo luogo, per avere un metodo di valutazione standardizzato ed evitare l'eccessiva generosità degli insegnanti nei giudizi espressi sui propri studenti.

e mitigare gli effetti discriminatori causati dall'uso delle tecnologie intelligenti, alla luce delle garanzie offerte dal principio di eguaglianza e non discriminazione secondo l'inquadramento offerto dal costituzionalismo contemporaneo. E a questo proposito, il dato da cui sembra più opportuno partire è l'individuazione della tipologia di discriminazione generata dai sistemi di IA.

Innanzitutto, va osservato come molte delle manifestazioni di bias che comportano l'insorgere di pratiche e di decisioni pregiudizievoli siano ascrivibili alle tradizionali categorie della discriminazione diretta e della discriminazione indiretta. Tali tipologie di disparità di trattamento, la prima caratterizzata dalla presenza di un intento discriminatorio e la seconda contraddistinta dal diverso impatto che possono generare decisioni e azioni apparentemente neutrali (Strazzari 2008), possono essere affrontate ricorrendo agli strumenti classici del diritto antidiscriminatorio, seppur ponendo la dovuta attenzione alla natura tecnologica delle discriminazioni in esame (Nardocci 2023). Sia che si tratti di discriminazioni dirette o indirette, le soluzioni giuridiche da implementare devono necessariamente prendere in considerazione non solo le operazioni tecniche di definizione e di preparazione del dataset da utilizzare, ma anche quelle caratteristiche di opacità e di potenziale autonomia operativa che contraddistinguono i sistemi di IA. L'incapacità di conoscere e comprendere i processi decisionali seguiti dall'IA, così come la sua capacità di modificare o di creare nuovi criteri inferenziali per l'elaborazione dei dati, possono, infatti, limitare l'efficacia degli strumenti tradizionali nell'individuazione sia dell'intento, sia dell'impatto discriminatorio espressi dal sistema, limitando fortemente le possibilità di agire in rimedio al danno subito (*ibidem*).

Da questo punto di vista, dunque, risultano particolarmente apprezzabili le soluzioni normative che consentono di implementare le garanzie offerte dal principio di eguaglianza tenendo conto di questi ulteriori elementi di complessità. In questo senso, due interventi normativi risultano interessanti, in prospettiva comparata, per le misure che hanno predisposto a tale scopo.

Il primo è senza dubbio il Regolamento (UE) 1689/2024 del Parlamento europeo e del Consiglio del 13 giugno 2024 che stabilisce regole armonizzate sull'intelligenza artificiale (noto anche come AI

Act). Il Regolamento europeo, infatti, prevede specifiche disposizioni in tema di data governance, trasparenza e sorveglianza umana che, almeno per i sistemi di IA classificati ad alto rischio (art. 6) possono porsi nella direzione di dare un'efficace implementazione alle misure tradizionali di contrasto alle discriminazioni. Infatti, la previsione di regole volte a garantire la completezza e la rappresentatività statistica e geografica dei dataset utilizzati anche in riferimento all'ambito di impiego prospettato per l'applicazione del sistema di IA (art. 10), unite ai requisiti di interpretabilità e di spiegabilità dello stesso (art. 13) e alla predisposizione di misure che assicurino il controllo della persona umana sulle operazioni svolte dall'IA (art. 14), sono strumenti che possono realizzare forme, da un lato, di mitigazione *ex ante* dei bias e delle conseguenti pratiche discriminatorie e, dall'altro lato, di rimedio *ex post* all'insorgere di disparità di trattamento determinate dall'uso dell'IA. In particolare, il requisito di spiegabilità può rappresentare una misura, anche di natura tecnica (Palmirani 2020), fondamentale nel fornire ai soggetti interessati possibili rimedi per contrastare le discriminazioni subite, come dimostra anche il riconoscimento del diritto alla spiegazione dei processi decisionali a opera dell'AI Act. Il Regolamento (UE), infatti, riconosce a coloro che siano destinatari di decisioni, basate sui risultati forniti da sistemi ad alto rischio, tali da incidere significativamente sulla tutela dei diritti fondamentali il diritto di ottenere spiegazioni chiare e significative sul ruolo avuto dall'IA nel processo decisionali e su quali siano gli elementi principali della decisione presa (art. 86).

Il secondo intervento, riconducibile all'ordinamento statunitense, è, invece, il *Colorado Act concerning consumer protections in interactions with artificial intelligence systems* del 2024. Tale atto normativo, che si pone appunto l'obiettivo esplicito di tutelare i consumatori anche dalle discriminazioni di matrice algoritmica, prevede a tale scopo la realizzazione di una valutazione di impatto prodotto dai sistemi di IA, rispetto al quale si stabilisce l'obbligo di analizzare i possibili rischi di discriminazione causati dall'IA (Nardocci 2024). L'introduzione di un simile strumento risulta, almeno dal punto di vista teorico, particolarmente apprezzabile in quanto consente di analizzare in concreto l'impatto che l'IA può produrre sulla tutela dell'eguaglianza, fornendo un ulteriore supporto alle misure tradizionali in questo senso predisposte dall'ordinamento statunitense. Anche

in questo caso, quindi, si assiste a un'implementazione delle garanzie di eguaglianza e non discriminazione che tiene conto della matrice tecnologica dei possibili effetti discriminatori, andando a rafforzare la tutela dei diritti fondamentali della persona.

Tuttavia, occorre anche osservare come altre tipologie di manifestazioni dei bias e le relative conseguenze discriminatorie risultino, invece, di più complessa soluzione dal punto di vista del diritto costituzionale e delle sue categorie tradizionali. Il riferimento è in questo caso alle forme di *proxy discrimination* e a quei risultati discriminanti che sono frutto del funzionamento specifico dei sistemi di IA. L'eventualità che si verifichino forme di disparità di trattamento a causa dell'impiego dell'IA ma senza che queste siano fondate sulle caratteristiche o sulle categorie tutelate dal principio di eguaglianza rischia, infatti, di lasciare le persone così discriminate prive di tutela e di rimedi rispetto al danno subito. In questi termini, l'IA, per le sue capacità di inferenza funzionali a individuare correlazioni anche nascoste tra le informazioni analizzate, potrebbe generare nuove caratteristiche discriminanti e persone discriminate che attualmente non trovano protezione alla luce delle categorie del costituzionalismo contemporaneo (Nardocci 2021). A ciò si aggiungano anche tutte le situazioni in cui l'IA opera intersezioni tra fattori considerati discriminanti nell'inquadramento fornito dal principio di eguaglianza e che, pur comportando un acuirsi della disparità di trattamento proprio a causa dell'unione di più fattori di vulnerabilità, non trovano rimedio alla luce degli strumenti giuridici tradizionali attualmente azionabili (Nardocci 2023).

In considerazione di tali ipotesi, occorre quindi chiedersi quali soluzioni possa offrire il diritto, nella dimensione costituzionale contemporanea, e se si renda necessario implementare nuovi approcci e strumenti per evitare ingiustificati e ingiustificabili vuoti di tutela per le persone destinatarie dei risultati e delle decisioni elaborate dall'IA.

5. Nuovi percorsi e prospettive per una *human–centered* e *right–based* IA

I limiti di tutela che presenta il principio di eguaglianza, rispetto alle situazioni e alle condotte discriminatorie descritte in chiusura del

precedente paragrafo, costituiscono il vero banco di prova per analizzare la capacità del diritto costituzionale contemporaneo di rispondere alle sfide poste dall'IA e dalle dinamiche di potere di cui è espressione. Le ipotesi considerate, infatti, permettono di riflettere sulla necessità che il diritto, e in particolare il diritto costituzionale, elabori nuovi strumenti e soluzioni per garantire una piena tutela della persona nella nuova società algoritmica e, di conseguenza, realizzare sistemi di IA che possano definirsi *human–centered* e *right–based* (Casonato 2023).

In questo processo di evoluzione giuridica, in cui il ricorso a un approccio metodologico interdisciplinare e comparato risulta essenziale per comprendere quali percorsi normativi e di tutela possano risultare più opportuni (Vedaschi 2024; Penasa 2024; Guerra 2018), risulta però già possibile svolgere alcune considerazioni circa le soluzioni implementabili per declinare algoritmicamente il principio di eguaglianza, tenuto conto dei limiti del suo approccio classificatorio tradizionale di fronte alle funzionalità dell'IA (Siegel 2004).

Una prima direzione auspicabile potrebbe essere l'aumento delle politiche di diversità e inclusione nel settore dell'IA. Questo approccio normativo, già presente in alcuni atti che si pongono l'obiettivo di regolare l'IA, garantirebbe la possibilità di minimizzare i risultati discriminatori prodotti dall'IA intervenendo in termini di rappresentatività anche dei *team* di persone esperte che sviluppano questi sistemi tecnologici. Come già osservato in dottrina soprattutto con riferimento alle discriminazioni etniche e di genere (Fosch–Villaronga e Poulsen 2022; D'Amico 2020), l'implementazione di tali politiche consentirebbe non solo di limitare l'impatto discriminatorio generato dall'IA, ma anche di impiegare l'IA per valorizzare gli elementi di differenza laddove ciò si renda necessario per tutelare i diritti delle persone. È questo il caso, per esempio, della medicina di genere, in cui l'impiego di sistemi di IA per identificare *pattern* diagnostici e di trattamento in prospettiva di genere risulterebbe fondamentale per promuovere un'eguaglianza effettiva delle persone nel godimento del diritto alla salute (Fosch–Villaronga e Poulsen 2022). Da questa prospettiva, l'IA potrebbe anche diventare strumento per l'implementazione di azioni positive volte ad assicurare l'applicazione del principio di eguaglianza anche nella sua accezione sostanziale (Stradella 2020). Sono da considerarsi, quindi, apprezzabili

le disposizioni in questo senso previste sia dell'AI Act, con riferimento all'elaborazione di politiche di diversità e inclusione per lo sviluppo dei sistemi di IA c.d. a rischio minimo (art. 95), sia dal *Projeto de Lei nº 2.338, de 2023* brasiliano, che riconosce nei principi di eguaglianza, diversità, inclusione e pluralismo gli elementi cardine per lo sviluppo e l'uso dell'IA (artt. 2 e 3).

Una seconda prospettiva da considerare nel tentativo di definire il processo di evoluzione del principio di eguaglianza potrebbe consistere nella valorizzazione dell'approccio intersezionale quale possibile chiave di lettura e di soluzione delle discriminazioni causate dall'IA. Gli strumenti e gli assunti teorici e metodologici di questa teoria critica (Crenshaw 1989; Marini 2021; Bello 2022) offrirebbero, infatti, validi rimedi ai rischi discriminatori prospettati dall'IA dal punto di vista dell'intersezione tra categorie e elementi che possono portare diseguaglianza nel trattamento delle persone anche rispetto alla combinazione di caratteristiche di per sé non discriminanti operata da questa tecnologia. Da questa prospettiva, risulta degna di nota la scelta operata dal legislatore canadese di utilizzare meccanismi di analisi intersezionale per valutare l'impatto prodotto dall'IA sui diritti e le libertà delle persone. Infatti, la *Directive on automated decision–making* del 2019 stabilisce l'impiego dello strumento di *Gender–based Analysis Plus* per identificare l'impatto prodotto dall'IA sui profili di genere e su altri fattori identitari e per predisporre le misure a tal fine necessarie (Appendix C). In questo modo, dunque, risulta possibile esaminare l'impatto discriminatorio generato dall'IA più ad ampio spettro, tenendo conto delle conseguenze subite dai soggetti discriminati secondo le modalità descritte.

Le considerazioni illustrate rappresentano, ovviamente, solo alcune tra le proposte di percorso che il diritto costituzionale potrebbe intraprendere per assicurare tutela ai diritti delle persone a fronte dei mutamenti e delle problematicità dettate dall'uso dei sistemi di IA. Tuttavia, l'attesa per un intervento giuridico che tenga conto delle istanze descritte non può essere ulteriormente tollerata alla luce di una diffusione dell'IA che sembra, per certi versi, realizzarsi in modo incontrollato. La realizzazione di un'IA realmente *human–centered* e *right–based*, anche dalla prospettiva dell'implementazione del principio di eguaglianza, dovrà rientrare tra i principali obiettivi dei prossimi interventi normativi

in materia per evitare che si crei un vuoto di tutela verso quelle persone che si trovano maggiormente esposte a fattori di vulnerabilizzazione nel contesto della società algoritmica.

Riferimenti bibliografici

Arrighi G. (2001) *Global Capitalism and the Persistence of the North–South Divide*, "Science & Society", 4: 469–476.

Avanzini G. (2022) *Intelligenza artificiale e nuovi modelli di vigilanza pubblica in Francia e Olanda*, "Giornale di diritto amministrativo", 3: 316–325.

Baer S. (2012) "Equality", in M. Rosenfeld e A. Sajó (a cura di), *The Oxford Handbook of Comparative Constitutional Law*, Oxford University Press, Oxford.

Balkin J.M. (2017) *The Three Laws of Robotics in the Age of Big Data*, "Ohio State Law Journal", 5: 1217–1241.

——. (2018) *Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation*, "U.C. Davis Law Review", 51: 1149–1210.

Barocas S. e A.D. Selbst (2016) *Big Data's Disparate Impact*, "California Law Review", 3: 671–732.

——., M. Hardt e A. Narayanan (2023) *Fairness and Machine Learning. Limitations and Opportunities*, MIT Press, Boston.

Bassini M., L. Liguori e O. Pollicino (2018) "Sistemi di intelligenza artificiale, responsabilità e accountability. Verso nuovi paradigmi?", in F. Pizzetti (a cura di), *Intelligenza artificiale, protezione dei dati personali e regolazione*, Giappichelli, Torino, 333–372.

Bello B.G. (2022) *Intersezionalità. Teorie e pratiche tra diritto e società*, FrancoAngeli, Milano.

Bragg D. *et al.* (2019) "Sign Language Recognition, Generation, and Translation: An Interdisciplinary Perspective", in J.P. Bigham, S. Azenkot e S. Kane (a cura di), *ASSETS'19. The 21st International ACM SIGACCESS Conference on Computers and Accessibility*, Association for Computing Machinery, New York, 16–31.

Caretti P. (2017) *I diritti fondamentali. Libertà e diritti sociali*, Giappichelli, Torino.

Casonato C. (2023) *Unlocking the Synergy: Artificial Intelligence and (old and new) Human Rights*, "BioLaw Journal – Rivista di BioDiritto", 3: 233–240.

——. (2025) *L'intelligenza artificiale tra pubblico e privato: una sfida per il costituzionalismo (e per i costituzionalisti)*, in "Diritto pubblico comparato ed europeo", 1: 5–13.

Celeste E. (2022) *Digital Constitutionalism. The Role of Internet Bills of Right*, Routledge, Londra.

Commission for Racial Equality (1988) *Medical School Admissions: Report of a formal investigation into St. George's Hospital Medical School*, *https://www.jstor.org/stable/pdf/community.28327674.pdf*, ultimo accesso, 20 aprile 2025.

Crawford K. (2021) *Né intelligente, né artificiale. Il lato oscuro dell'IA*, Il Mulino, Bologna.

Crenshaw K. (1989) *Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics*, "The University of Chicago Legal Forum", 140: 139–167.

D'Amico M. (2020) *Una parità ambigua. Costituzione e diritti delle donne*, Raffaello Cortina, Milano.

Dias Oliva T., D.M. Antonialli e A. Gomes (2021) *Fighting Hate Speech, Silencing Drag Queens? Artificial Intelligence in Content Moderation and Risks to LGBTQ Voices Online*, "Sexuality & Culture", 2: 700–732.

Falletti E. (2022) *Discriminazione algoritmica. Una prospettiva comparata*, Giappichelli, Torino.

Fasan M. (2022) *I principi costituzionali nella disciplina dell'Intelligenza Artificiale. Nuove prospettive interpretative*, "DPCE online", 1: 181—199.

——. (2024) *Intelligenza artificiale e costituzionalismo contemporaneo. Principi, diritti e modelli in prospettiva comparata*, Editoriale Scientifica, Napoli.

Fineman M.A. (2019) *Vulnerability and Social Justice*, "Valparaiso University Law Review", 53: 341–369.

Fioravanti M. (2009) *Costituzionalismo. Percorsi della storia e tendenze attuali*, Laterza, Roma–Bari.

Fosch–Villaronga E. e A. Poulsen (2022) "Diversity and Inclusion in Artificial Intelligence", in B. Custers e E. Fosch–Villaronga (a cura di), *Law and Artificial Intelligence. Regulating AI and Applying AI in Legal Practice*, Springer, Berlino,109–134.

Graziani C. (2021) *Intelligenza artificiale e fonti del diritto: verso un nuovo concetto di* soft law*? La rimozione dei contenuti terroristici* online *come* case–study, "DPCE online", n° speciale: 1473–1490.

Guerra G. (2018) *An Interdisciplinary Approach for Comparative Lawyers: Insights from the Fast–Moving Field of Law and Technology*, "German Law Journal", 3: 579–612.

Hacker P. (2018) *Teaching Fariness to Artificial Intelligence: Existing and Novel Strategies against Algorithmic Discrimination under EU Law*, "Common Market Law Review", 4: 1143–1186.

Hong T. e M. Hu (2025) *Opportunities, Challenges, and Regulatory Responses to China's AI Computing Power Development under DeepSeek's Changing Landscape*, "International Journal of Digital Law and Governance", 2: 1–23.

Kelly A. (2021) *A Tale of Two Algorithms: The Appeal and Repeal of Calculated Grades Systems in England and Ireland in 2020*, "British Educational Research Journal", 3: 725–741.

Keyes O. (2018) *The Misgendering Machines: Trans/HCI Implications of Automatic Gender Recognition*, "Proceedings of the ACM on Human–Computer Interaction", 2: 1–22.

Kleinberg J., J. Ludwig, S. Mullainathan e C.R. Sunstein (2018) *Discrimination in the Age of Algorithms*, "Journal of Legal Analysis", 10: 113–174.

Lycett M. (2013) *"Datafication": Making Sense of (Big) Data in a Complex World*, "European Journal of Information Systems", 4: 381–386.

Marini G. (2021) *Intersezionalità: genealogia di un metodo giuridico*, "Rivista Critica del Diritto Privato", 4: 473–502.

Nardocci C. (2021) *Intelligenza artificiale e discriminazioni*, "La Rivista del Gruppo di Pisa", 3: 9–60.

——. (2023) *Artificial Intelligence–based Discrimination: Theoretical and Normative Responses. Perspectives from Europe*, "DPCE online", 3: 2367–2393.

——. (2024), *Dalla "self–regulation" alla frammentata regolamentazione dei sistemi di intelligenza artificiale: uno sguardo alla diversa prospettiva statunitense*, "Diritto pubblico comparato ed europeo", 4: 859–954.

Orlikowski W.J. (1992) *The Duality of Technology: Rethinking the Conspet of Technology in Organizations*, "Organization Science", 3: 398–427.

Palmirani M. (2020) *Big Data e conoscenza*, "Rivista di filosofia del diritto", 1: 73–92.

Penasa S. (2024) *Diritto e tecnologia nella recente riflessione giuridica comparata: "etichette" concettuali, sistemi di produzione normativa e metodi della comparazione*, "Diritto pubblico comparato ed europeo", n° speciale: 951–978.

Prince A.E.R. e D. Schwarcz (2020) *Proxy Discrimination in the Age of Artificial Intelligence and Biga Data*, "Iowa Law Review", 3: 1257–1318.

Rachovitsa A. e N. Johann (2022) *The Human Rights Implications of the Use of AI in the Digital Welfare State: Lessons Learned from the Dutch SyRI Case*, "Human Rights Law Review", 2: 1–15.

Rodotà S. (2012) *Il diritto di avere diritti*, Laterza, Roma–Bari.

Santosuosso A. e G. Sartor (2024) *Decidere con l'IA. Intelligenze artificiali e naturali nel diritto*, Il Mulino, Bologna.

Sarker I.H. (2024) *LLM Potentially and Awareness: A Position Paper from the Perspective of Trustworthy and Responsible AI Modeling*, "Discover Artificial Intelligence", 4: 1–7.

Sartor G. e F. Lagioia (2020) "Le decisioni algoritmiche tra etica e diritto", in U. Ruffolo (a cura di), *Intelligenza artificiale. Il diritto, i diritti, l'etica*, Giuffré, Milano, 63–92.

Scherer M.U. (2016) *Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies*, "Harvard Journal of Law & Technology", 2: 354–400.

Siegel R.B. (2004) *Equality Talk: Antisubordination and Anticlassification Values in Constitutional Struggles over Brown*, "Harvard Law Review", 117: 1470–1547.

Simoncini A. (2017) "Sovranità e potere nell'era digitale", in T.E. Frosini, O. Pollicino, E. Apa e M. Bassini (a cura di), *Diritti e libertà in Internet*, Le Monnier Università, Milano–Firenze, 19–38.

——. e E. Cremona (2021) *L'AI fra pubblico e privato*, "DPCE online", 1: 253–271.

Stradella E. (2020) "Stereotipi e discriminazioni: dall'intelligenza umana all'intelligenza artificiale", in Aa.Vv. (a cura di), Liber Amicorum *per Pasquale Costanzo — Diritto costituzionale in trasformazione. Vol. I — Costituzionalismo, Reti e Intelligenza artificiale*, ConsultaOnline, Genova, 391–400.

Strazzari D. (2008) *Discriminazione razziale e diritto. Un'indagine comparata per un modello "europeo" dell'antidiscriminazione*, CEDAM, Padova.

Sulmicelli S. (2023) *Algorithmic Content Moderation and the LGBTQ+ Community's Freedom of Expression on Social Media: Insights From the EU Digital Service Act*, "BioLaw Journal – Rivista di BioDiritto", 2: 453–471.

——. (2024) "La transizione digitale nel prisma dell'intelligenza artificiale. Un'introduzione tra comparazione, interdisciplinarità e prospettive critiche", in S. Franca, A. Porcari e S. Sulmicelli (a cura di), *Le transizioni e il diritto. Atti delle giornate di studio 21–22 settembre 2023*, Università degli Studi di Trento, Trento, 395–416.

——. (2025) *Queer–Responsive Regulation for Artificiale Intelligence in Healthcare: A Comparative Study*, "University of New South Wales Law Journal", 4: (in corso di pubblicazione).

Traverso P. (2022) *Breve introduzione tecnica all'intelligenza artificiale*, "DPCE online", 1: 155–167.

van Bekkum M. e F. Zuiderveen Borgesius (2021) *Digital Welfare Fraud Detection and the Dutch Syri Judgment*, "European Journal of Social Security", 4: 323–340.

Vedaschi A. (2024) *Tecnologia,* counter–terrrorism *e diritti*, "Diritto pubblico comparato ed europeo", n° speciale: 979–1010.

Wachter S., B. Mittelstadt e C. Russell (2021) *Bias Preservation in Machine Learning: The Legality of Fairness Metrics under EU Non–Discrimination Law*, "West Virginia Law Review", 3: 735–790.

Wangsa K., S. Karim, E. Gide e M. Elkhodr (2024) *A Systematic Review and Comprehensive Analysis of Pioneering AI Chatbot Models from Education to Healthcare: ChatGPT, Bard, Llama, Ernie and Grok*, "Future Internet", 7: 1–23.