OLTRE L'UGUAGLIANZA *BY DESIGN*: GENERE E INTELLIGENZA ARTIFICIALE*

di Federico Savastano**

Sommario. 1. Sistemi di IA e processi decisionali automatizzati nelle discriminazioni di genere. – 2. *Equal Treatment by design*: eliminare i *bias* alla radice. – 3. I rischi di un'uguaglianza di genere standardizzata. – 4. Conflitto e controllo per una tutela dinamica nell'era dell'intelligenza artificiale.

1. Sistemi di IA e processi decisionali automatizzati nelle discriminazioni di genere.

La sempre più pervasiva diffusione di sistemi di intelligenza artificiale incide anche sul diritto antidiscriminatorio e, pertanto, sull'uguaglianza di genere. Tale ambito si pone infatti come paradigma d'analisi ideale per svelare come l'astrattezza del dato possa occultare asimmetrie di potere profonde, offrendo un osservatorio privilegiato sulla necessità di un controllo giuridico che sappia andare oltre la mera conformità tecnica del sistema. Le dinamiche algoritmiche intervengono infatti su dati che riflettono diseguaglianze storicamente sedimentate e possono contribuire alla loro riproduzione in una forma mediata e – per questo – «legittimata» dalla tecnica. La dottrina ha individuato accuratamente le modalità in cui, ad esempio, la moderazione automatizzata dei contenuti online, basata su sistemi di intelligenza artificiale, incorpora e amplifica i pregiudizi di genere già presenti nella società, producendo effetti discriminatori sistematici nei testi, nelle immagini e nei sistemi di raccomandazione[1].

Queste dinamiche riportano al centro del dibattito il principio di uguaglianza sostanziale ex art. 3, Cost., che rischia di subire delle violazioni non immediatamente evidenti, ma pericolosamente concrete.

L'output dei sistemi di intelligenza artificiale, del resto, è l'esito di modelli predittivi in cui la percezione della dimensione sociale è necessariamente attenuata, producendo così una situazione per cui gli esiti decisionali sono riconducibili alla tecnica e non sono il prodotto di relazioni sociali canalizzate nel circuito della responsabilità democratica, né sono tantomeno la risposta a necessità ed esigenze sociali che quel circuito è portato ad interpretare e tradurre in decisione.

Il rischio è che l'automazione non sia in grado di intercettare le diseguaglianze di fatto che emergono da processi complessi e che invece sia più portata ad interpretare l'uguaglianza come mera assenza di differenziazioni esplicite, in una logica esclusivamente formale[2].

---

* *Sottoposto a referaggio*.

** Ricercatore di Diritto costituzionale – Università di Roma La Sapienza.

[1] G. Vasino, *Gender bias e intelligenza artificiale: la disuguaglianza nell'algoritmo*, in F. Savastano (a cura di), *Genere e diritti. Sfide, rappresentanza e discriminazioni nella società contemporanea*, Roma, 2025, 269-282.

[2] Sull'evoluzione del concetto di discriminazione alla luce dell'avvento dell'automazione, vedi E. Falletti, *Discriminazione algoritmica. Una prospettiva comparata*, Torino, 2022, 17-64.

I sistemi algoritmici funzionano su dati e modelli che riflettono diseguaglianze sociali già esistenti, in particolare quelle di genere[3]: quando decidono, dunque, riproducono e stabilizzano queste diseguaglianze[4].

La risposta normativa più immediata del diritto antidiscriminatorio punta ad anticipare l'uguaglianza nella tecnica: il concetto di fondo è quello della *equality by design*, che consiste nell'imporre al fornitore l'introduzione di meccanismi che garantiscano l'uguaglianza nella fase di sviluppo dei modelli[5]. Questo approccio è basato sul fatto che intervenire in una fase strutturale è senz'altro più efficace e realistico rispetto ad un intervento *ex post* su discriminazioni perpetrate e ormai difficili da correggere.

Una strategia necessaria e condivisibile, dunque, ma non del tutto sufficiente: inserire criteri di uguaglianza nel codice, infatti, aiuta ma rischia di trasformare l'uguaglianza in qualcosa di tecnico, misurabile e standardizzato, laddove l'art. 3, Cost. configura invece questo principio come strettamente legato a situazioni di conflitto, che lo rendono oggetto di costante evoluzione[6].

L'introduzione di sistemi di garanzia egualitaria nei modelli di IA rischia pertanto di indebolire il conflitto antidiscriminatorio, che diventa difficile da attivare e può determinare uno svuotamento del principio di uguaglianza della sua forza critica. Come contestare, del resto, qualcosa che è stato certificato come «equo» già in partenza?

Il caso delle diseguaglianze di genere si pone come emblematico sotto questo profilo, dato che esse sono spesso frutto di discriminazioni storiche, strutturali e relazionali che nessun modello tecnico può catturare fino in fondo senza incorrere nel rischio di eccessive semplificazioni. I sistemi di intelligenza artificiale tendono infatti a riprodurre e amplificare stereotipi e asimmetrie già presenti nei dati e nei contesti sociali in cui vengono sviluppati e utilizzati: dataset sbilanciati possono produrre risultati peggiori per le donne; procedure automatizzate di selezione, valutazione o classificazione possono determinare forme indirette di discriminazione; sistemi biometrici e di riconoscimento facciale mostrano margini di errore più elevati nei confronti delle donne, soprattutto se appartenenti a gruppi vulnerabili o minoritari; la stessa riduzione del genere a categorie fisse e binarie rischia di incorporare concezioni essenzialistiche e semplificate dell'identità personale[7].

L'uguaglianza di genere, che la nostra Costituzione garantisce a partire dall'art. 3 Cost., può essere pertanto «anticipata» nei codici algoritmici ma non deve intendersi come completamente formalizzabile, laddove una tutela adeguata richiede che essa resti controllabile, contestabile e, soprattutto, correggibile nel tempo. Si tratta, concretamente, di comprendere le modalità in cui un principio costituzionale aperto viene incorporato all'interno di processi tecnologicamente strutturati, di interrogarsi sulla relazione tra razionalità tecnica e valutazione giuridica, nonché sulla capacità del diritto costituzionale di

[3] Per una disamina specifica, v. C. Nardocci, *Intelligenza artificiale e discriminazioni*, in *Rivista Gruppo di Pisa*, 3, 2021, 9-60; per una prospettiva più ampia, sempre C. Nardocci, *Algoritmi, eguaglianza, discriminazione. Le sfide dell'intelligenza artificiale*, Torino, 2025.

[4] Sottolinea E. Stradella, *AI, tecnologie innovative e produzione normativa: potenzialità e rischi*, in *DPCE Online*, 3, 2020, 3350, che «nella prospettiva giuridica il bias rappresenta lo stereotipo pronto a trasformarsi in scelta discriminatoria, e per questo il principale bersaglio del principio di non discriminazione»

[5] In relazione ai *bias* algoritmici di genere, l'espressione *«Equality by Design»* (o *«Gender Equality by Design»*) è stata introdotta e resa celebre principalmente da I. Bohnet, *What Works: Gender Equality by Design*, Harvard, 2018, in cui l'Autrice ha applicato i principi dell'economia comportamentale per dimostrare che è più efficace «progettare» sistemi equi piuttosto che tentare di cambiare i pregiudizi individuali.

[6] Vedi la tesi di fondo del saggio di E. Stradella, *Stereotipi e discriminazioni: dall'intelligenza umana all'intelligenza artificiale*, in AA. VV., *Liber Amicorum per Pasquale Costanzo*, I, Genova, 2020, 391-400.

[7] Un'utile rassegna è rintracciabile nel report A. Balayn, S. Gürses, *Beyond Debiasing. Regulating AI and its inequalities*, Delft University of Technology, 2021, 37-48.

mantenere la propria funzione critica in un contesto caratterizzato da opacità, complessità e standardizzazione delle decisioni.

Le diseguaglianze di genere derivanti dalle disfunzioni dell'intelligenza artificiale non sono imputabili a una legge sbagliata, a una decisione palesemente o addirittura consapevolmente discriminatoria adottata da un soggetto individuabile, ma sono il prodotto del funzionamento di «sistemi». Tali «sistemi», come accennato, sono costruiti e funzionano sulla base di dati che riflettono una società già diseguale e trasformano quelle diseguaglianze in criteri tecnici sui quali si fondano e sulla base dei quali producono risultati che presentano come «oggettivi». In sostanza, un prodotto della tecnica, una rielaborazione e trasformazione di discriminazioni radicate più o meno visibili[8] che vengono processate e assumono un nuovo carattere di inattaccabile «equità». La tecnica, in questo senso, trasforma le diseguaglianze da prodotto a elemento strutturale, e le normalizza interiorizzandole nella forma del «dato», ammantandole di «neutralità».

Questa neutralità è solo apparente, dato che l'esito algoritmico è la risultante di un processo cumulativo che tiene conto e somma tante piccole scelte e tante piccole decisioni, interpretazioni e informazioni che incorporano il dato storico, ossia un passato diseguale da cui l'algoritmo ha imparato, che l'algoritmo ha riprodotto e che proietterà nel futuro[9].

In questo modo, la produzione di discriminazioni di genere diventa sistemica ed esponenziale[10].

L'art. 3 Cost., nella sua duplice dimensione, offre invece una chiave di lettura più ampia. Il suo scopo è infatti quello di spingere ad intervenire su disuguaglianze di fatto, indipendentemente dalle cause che le hanno prodotte o dall'intenzionalità con cui si sono generate: esso guarda agli effetti, alle condizioni materiali e agli svantaggi strutturali; guarda dunque non tanto alla presunta «neutralità» con cui una decisione viene presa, quanto piuttosto alla situazione che si determina come conseguenza della stessa.

In questo senso, la neutralità – o meglio, l'«apparente» neutralità – della tecnica passa in secondo piano rispetto alle condizioni di fatto da tutelare.

Questo passaggio può apparire banale ma, declinato sul tema dell'intelligenza artificiale, non lo è affatto. Le decisioni algoritmiche infatti si presentano come necessarie, inevitabili e come risultato di complesse e inattaccabili analisi statistiche di una enorme mole di dati, riducendo, in qualche modo, lo spazio del giudizio critico e la stessa considerazione dell'ipotesi di alternative possibili ad un «modello» tecnico[11].

Ma le decisioni restano decisioni e da esse scaturiscono situazioni che devono essere controllate, ai fini della verifica delle eventuali discriminazioni e della tutela dei soggetti

---

[8] Spiega metodi e output F. Fedorczyk, *Addressing AI-driven gender discrimination: the role of the forthcoming EU AI Act and Corporate Social Responsibility*, in *Rivista di Diritto comparati*, 3, 2023, 258-265.

[9] In questo senso, esemplare è l'ordinanza sul *Caso Deliveroo*, con cui il Tribunale di Bologna ha sanzionato la discriminatorietà indiretta dell'algoritmo usato dalla nota azienda di consegne a domicilio per gestire i rapporti con i lavoratori (Trib. Bologna, Sez. Lavoro, ord. 31 dicembre 2020, n. r.g. 2949/2019). Più approfonditamente, S. Borelli, M. Ranieri, *La discriminazione nel lavoro autonomo. Riflessioni a partire dall'algoritmo Frank*, in *Labour Law Issues*, 7, 1, 2021, 20-47.

[10] Lo spiega C. Nardocci, *Discrimination Revised. How AI Is Reshaping Anti-Discrimination Law*, in *Rivista di diritti comparati*, Special Issue VIII, 2026, 168-176, applicando il concetto di *proxy discrimination*, secondo cui anche se il programmatore elimina il dato «genere» (per essere neutrale), il sistema può ricostruirlo attraverso altri dati (es. abitudini di acquisto, luoghi frequentati, linguaggio usato)

[11] A.C. Amato Mangiameli, *Intelligenza artificiale, big data e nuovi diritti*, in *Rivista italiana di informatica e diritto*, 1, 2022, 96, evidenzia come *«nell'era della crescita esponenziale di dati, il paradigma logico-deduttivo cede il passo all'approccio statistico»*. In senso critico, v. L. Di Majo, *Le incognite sull'utilizzo dell'intelligenza artificiale nel processo costituzionale in via incidentale*, in *Rivista Gruppo di Pisa*, 2, 2023, 35, il quale evidenzia come l'introduzione di modelli algoritmici nei processi decisionali giurisdizionali ponga in tensione il principio del libero convincimento del giudice e la stessa struttura del contraddittorio, rischiando di incidere sulla natura del giudizio.

coinvolti. In questo quadro assume rilievo anche il principio della *human oversight*, che attraversa la regolazione europea dei sistemi di intelligenza artificiale (art. 14, *AI Act*): anche quando un sistema algoritmico è impiegato nei processi decisionali, la decisione giuridicamente rilevante resta, in linea di principio, imputabile a un soggetto umano, chiamato ad esercitare una funzione di controllo e di eventuale revisione dell'esito automatizzato[12].
Proprio per questo, le decisioni restano comunque soggette a scrutinio di ragionevolezza e proporzionalità, in quanto portatrici di scelte rilevanti sul piano costituzionale. L'art. 3, Cost. deve dunque porsi come parametro di valutazione delle strutture decisionali stesse, non limitandosi più ad imporre una valutazione sulla decisione ma ampliando la propria portata, spingendo ad una valutazione del percorso che ha portato a quella decisione: non è più un problema di decisione sbagliata quanto dell'errore del sistema che ha prodotto quella decisione sbagliata.
L'intelligenza artificiale richiede pertanto un'interpretazione dell'uguaglianza costituzionale come strumento per interrogare sistemi complessi che producono disuguaglianze senza dichiararle.

2. Equal Treatment by design: eliminare i bias alla radice. Uno dei problemi principali del rapporto tra intelligenza artificiale e diritto è costituito dal tempo. Il diritto arriva infatti spesso «dopo» che si sono prodotte delle discriminazioni algoritmiche e, quindi, rischia di arrivare «tardi». Una volta prodotte, peraltro, tali discriminazioni sono difficili da correggere perché sono il frutto di decisioni opache[13], perché producono effetti cumulati e che continuano a cumularsi nel tempo e, non da ultimo, perché difficilmente sono riconducibili a un singolo atto o a un singolo responsabile. Non può pertanto applicarsi un modello classico di intervento in base al quale, una volta scoperta una forma di discriminazione, si interviene ad eliminarla con un singolo atto specifico e mirato.
Bisogna intervenire prima e altrove: non guardare solo all'output ma al modo in cui una decisione si costruisce. Se è un «sistema» a produrre decisioni discriminatorie, allora è quel «sistema» l'oggetto di intervento, e c'è bisogno di intervenire «prima» che lo stesso funzioni. È in ragione di queste considerazioni che si afferma il concetto di *equality by design*, che sposta l'attenzione dalla decisione alla produzione della stessa, dal «merito» alle «modalità», con l'obiettivo di correggere la «cosa» concentrandosi sul «come» essa è stata determinata. Ed è su questo concetto che si basa l'intervento normativo a livello europeo[14]. Il diritto dell'Unione ha provato ad agire sulla prevenzione delle discriminazioni, e quindi sul momento precedente

[12] Vedi M. Fasan, *I principi costituzionali nella disciplina dell'Intelligenza Artificiale. Nuove prospettive interpretative*, in DPCE Online, 1, 2022, 181-199; I.P. Di Ciommo, *La prospettiva del controllo nell'era dell'Intelligenza Artificiale: alcune osservazioni sul modello Human In The Loop*, in *Federalismi.it*, 3, 2023, 68-90.
[13] F. Faini, *Intelligenza artificiale e regolazione giuridica: il ruolo del diritto nel rapporto tra uomo e macchina*, in *federalismi.it*, 2, 2023, 5, parla di «congenita opacità che caratterizza gli algoritmi, che si declina in un'opacità strutturale, derivante dal funzionamento degli stessi e dal fatto che resta non comprensibile persino ai programmatori l'iter logico seguito dalla macchina per giungere al risultato partendo dai dati a disposizione». Il tema dell'opacità dell'algoritmo è comunque *ab origine* presentato in F. Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information*, Harvard, 2015.
[14] L'approccio *by design* emerge in modo implicito nella disciplina europea dell'intelligenza artificiale: l'*AI Act* non utilizza questa espressione, ma impone una serie di obblighi che incidono direttamente sulla fase di progettazione dei sistemi, tra cui la gestione dei rischi (art. 9), i requisiti relativi alla qualità dei dati utilizzati per l'addestramento (art. 10), la documentazione tecnica (art. 11) e la supervisione umana (art. 14). In particolare, l'attenzione alla qualità, alla rappresentatività e all'assenza di distorsioni nei dataset utilizzati per i sistemi ad alto rischio mira proprio a prevenire la riproduzione di *bias* discriminatori, inclusi quelli di genere. In questo senso, l'idea di *equality by design* può essere letta come una chiave interpretativa della logica preventiva che caratterizza la regolazione europea, orientata a intervenire sulla progettazione dei sistemi.

il loro manifestarsi: l'*AI Act*[15] – come del resto il *Digital Services Act*[16] e il *GDPR*[17] – disegna un quadro di responsabilità differenziate: se al fornitore spettano gli obblighi di qualità dei dati e conformità tecnica nella fase di addestramento (ex art. 10 AI Act), sul *deployer* (o utilizzatore) grava la responsabilità di monitorare il sistema durante la fase di utilizzo, garantendo che il contesto d'impiego non generi distorsioni impreviste [18]. Peraltro, proprio l'art. 10 dell'*AI Act* eleva la *governance* dei dati a requisito di conformità ineludibile. Non si tratta solo di selezionare *dataset* ampi, ma di garantire che essi siano adeguati e rappresentativi dell'ambiente in cui il sistema opererà; un set di dati parziale, infatti, restituirebbe una visione distorta della realtà, paragonabile all'analisi di una singola parte di un organismo complesso che preclude la comprensione dell'insieme[19]. In questo senso, l'art. 10, par. 5, introduce la c.d. *debiasing exception*, una clausola di salvaguardia che autorizza, in via eccezionale e per motivi di interesse pubblico, il trattamento di categorie particolari di dati personali[20].

L'approccio della regolazione europea guarda dunque alla struttura dei processi, perché percepisce che le diseguaglianze possono essere insite nei modelli, nelle metriche e nei dati che vengono selezionati e utilizzati: se il problema è nella progettazione, allora è questa che deve passare il vaglio del principio di uguaglianza.

Equality by design prevede una scelta a priori di cosa conta come equo, così da poter determinare dati e obiettivi tenendo conto in partenza delle asimmetrie e costruire modelli che siano strutturalmente orientati all'uguaglianza. In questo senso, il principio di uguaglianza diventa una sorta di criterio tecnico di costruzione del processo decisionale algoritmico; un criterio di conformazione preventiva del potere decisionale, destinato a orientare le scelte relative alla selezione dei dati, alla definizione degli obiettivi e alla costruzione dei modelli. La dottrina, non senza qualche iniziale riluttanza[21], ha accolto questa svolta rendendosi conto della difficoltà di interventi *ex post*, sia in termini di dimostrazione della discriminazione, sia di ricostruzione della *ratio decidendi* (opacità), sia per l'impossibilità di individuare l'attore responsabile.

L'uguaglianza *by design*, al netto delle perplessità che parte della dottrina ha mostrato di

---

[15] Reg. (UE) 2024/1689.

[16] Reg. (UE) 2022/2065.

[17] Reg. (UE) 2016/679. «Sia il Regolamento sulla protezione dei dati personali sia il Digital Services Act condividono l'approccio basato sul rischio, quale strumento finalizzato ad incentivare una maggiore assunzione di responsabilità da parte dei soggetti pubblici e privati operanti sul mercato digitale». Così I.P. Di Ciommo, op. cit., 83; vedi anche G. De Gregorio, P. Dunn, O. Pollicino, *A partire dalla Strategia per il Mercato Unico Digitale, l'approccio basato sul rischio è stato applicato in tutte le principali normative UE: analizziamo le analogie e le differenze tra GDPR, DSA e AI Act*, in *Agenda Digitale*, 15 settembre 2022.

[18] G. Vasino, op. cit., 279.

[19] Secondo il GDPR, il trattamento di dati sensibili (origine razziale, orientamento sessuale, genere, ecc.) è generalmente vietato. Tuttavia, per eliminare i *bias* da un'intelligenza artificiale, è tecnicamente necessario «vedere» quei dati per capire se il sistema sta discriminando. Cfr. A. Palma Ortigosa, *Data and Data Governance (Article 10)*, in A. Huergo Lora, G.M. Díaz González (a cura di), *The EU Regulation on Artificial Intelligence: A Commentary*, Wolters Kluwer, Milan, 2025, pp. 197-200. L'autore collega questa eccezione al *"Substantial Public Interest"* (Considerando 70 dell'*AI Act*). Sostiene che la lotta alle discriminazioni algoritmiche è un valore sociale così alto da giustificare una compressione temporanea e protetta del diritto alla riservatezza dei dati sensibili.

[20] Tale deroga è subordinata a rigorose garanzie, tra cui l'obbligo di pseudonimizzazione, limitazioni tecniche al riutilizzo e il dovere di cancellazione dei dati una volta raggiunta la finalità di correzione.

[21] Secondo O. Pollicino, *Regolazione e innovazione tecnologica nell''ordinamento della rete"*, in *Rivista AIC*, 2, 2025, 124, «non è un caso che il *Digital Markets Act* abbia di fatto sfatato il tabù dell'intervento solo *ex post* del diritto della concorrenza, prevedendo un quasi eretico (per alcuni), ma assai efficace (per quasi tutti), intervento della disciplina rilevante *ex ante*, proprio a causa delle nuove sfide che l'accrescimento del potere di tali piattaforme sta facendo emergere»

mantenere[22], non è però concetto di facile applicazione. Per intervenire sul codice, infatti, è necessario tradurre l'uguaglianza in criteri tecnici, operare scelte su cosa misurare e procedere a conseguenti standardizzazioni. In altre parole, l'integrazione dell'eguaglianza nella progettazione algoritmica presuppone la possibilità di tradurre un principio costituzionale aperto e relazionale in criteri tecnici formalizzati, idonei a guidare la costruzione dei sistemi. Una tale operazione comporta una selezione preventiva delle dimensioni rilevanti dell'eguaglianza, che tende a privilegiare aspetti misurabili e standardizzabili. Ne deriva una rappresentazione dell'eguaglianza come insieme di requisiti procedurali o statistici, suscettibili di verifica attraverso strumenti di audit e valutazione.

Un principio aperto, relazionale e frutto di un'evoluzione storica basata sul conflitto sociale va tradotto in qualcosa di misurabile, verificabile e certificabile, con il concreto rischio che si trasformi in una sorta di *checklist*.

Le fonti europee alle quali si è fatto riferimento rafforzano questa preoccupazione. Esse insistono su audit, documentazione e compliance, determinando un'idea di uguaglianza «tecnicamente» dimostrabile e postulabile «a priori»: una coincidenza pressoché piena tra uguaglianza e conformità procedurale, in altri termini.

L'approccio sistemico di queste fonti è ciò che dovrebbe garantire un *equal treatment by design* anche in relazione alle discriminazioni di genere, disinnescando le dinamiche dell'algoritmo e prevenendone i *bias*[23].

In questo senso, le problematiche connesse alle discriminazioni di genere costituiscono una vera e propria cartina al tornasole per la misurazione della validità dell'approccio *by design*. Il genere, infatti, non è una variabile semplice: è un concetto discusso e divisivo, in costante evoluzione, che coinvolge fattori storici, dinamiche di potere e implicazioni culturali. Le diseguaglianze di genere si radicano in dinamiche sociali storicamente consolidate, che difficilmente possono essere esaurite in variabili tecniche o in modelli statistici.

La traduzione dell'eguaglianza in parametri di progettazione tende a operare entro rappresentazioni semplificate del contesto sociale, assumendo come dati elementi che sono invece il prodotto di rapporti di potere e di processi di esclusione. Anche quando l'obiettivo è correttivo, la formalizzazione ex ante rischia di circoscrivere l'eguaglianza entro confini che riflettono una lettura parziale delle diseguaglianze di genere. Il rischio è dunque quello di un'eccessiva semplificazione; la considerazione che emerge è invece che l'uguaglianza *by design* funziona solo su una parte delle disuguaglianze di genere, e dovrebbe di essere affiancata da strumenti che garantiscano la possibilità di interrogare e rivedere le scelte incorporate nei sistemi algoritmici.

Se il *design* non basta, serve anche altro, ossia il controllo giuridico nel tempo: la prevenzione strutturale incide sulle condizioni di produzione della decisione, mentre il controllo giuridico consente di valutare gli effetti concreti delle decisioni nel tempo. In questa prospettiva, la progettazione algoritmica non esaurisce ogni forma di tutela dell'eguaglianza, ma ne rappresenta un «livello», destinato a operare in relazione con altri strumenti di garanzia.

L'intervento strutturale, pertanto, è una risposta necessaria alle trasformazioni che l'intelligenza artificiale propone e produce, ma l'eccessiva formalizzazione dei principi costituzionali che esso comporta apre una riflessione sulla tensione tra razionalità tecnica e principi costituzionali che coinvolge la natura stessa del funzionamento del principio di uguaglianza.

---

[22] A. Oddenino, *Intelligenza artificiale e tutela dei diritti fondamentali: alcune notazioni critiche sulla recente proposta di regolamento della IA con particolare riferimento all'approccio basato sul rischio e al pericolo di discriminazione algoritmica*, in A. Pajno, F. Donati, A. Perrucci (a cura di), *Intelligenza artificiale e diritto: una rivoluzione?*, Vol. I, Bologna, 2022, 196, richiamato in I.P. Di Ciommo, op. cit., 84.

[23] G. Vasino, op. cit., 279.

3. I rischi di un'uguaglianza di genere standardizzata. Estendere il principio di uguaglianza alla fase di progettazione algoritmica è dunque utile perché incide sulle strutture che producono discriminazioni, ma si corre il rischio di trasformare l'uguaglianza da principio giuridico a standard progettuale, in ragione dell'esigenza di definirla *a priori*, renderla misurabile e tradurla in parametro tecnico.

Tradurre l'uguaglianza in codice, infatti, significa scegliere quali sono le dimensioni che contano, lasciando fuori ciò che non è misurabile e fissando una versione dell'uguaglianza che rischia di non seguire le evoluzioni sociali.

L'uguaglianza di genere, in particolare, mostra i limiti dell'uguaglianza *by design*. Può valere la pena, in questo senso, proporre alcuni esempi per contestualizzare meglio il ragionamento.

Pensiamo ad un algoritmo di selezione del personale[24], progettato in modo che eviti forme di discriminazione di genere. Esso premierebbe persone che hanno avuto una carriera rapida, ad esempio, tenendo in considerazione la continuità di incarichi ad alta visibilità, gestione di grandi team e grandi budget e progressioni lineari. Sono tutti criteri formalmente neutri che però premierebbero gli uomini in quanto le donne potrebbero scontare, ad esempio, periodi di maternità o interruzioni lavorative. Ma anche ipotizzando di neutralizzare i periodi di maternità e di ponderare le interruzioni lavorative per cura familiare – come evidentemente farebbe un algoritmo costruito sulla base dell'uguaglianza *by design* – ci sarebbe il rischio di una preferenza nei confronti degli uomini dovuta al fatto che le donne hanno avuto finora spesso carriere più frammentate e ruoli meno visibili. L'algoritmo registrerebbe questa realtà trasformandola in modello e «normalizzando» di fatto un passato diseguale.

Pensiamo, in alternativa, ad un altro ambito economicamente strategico quale quello del credito, dove il genere, per definizione, non è o non dovrebbe essere una variabile. Ciò nonostante, le donne rischiano comunque di ottenere condizioni peggiori o di essere escluse dal credito perché i modelli valutano la solvibilità sulla base di dati storici legati alla continuità occupazionale, alla capacità di accumulo patrimoniale o alla stabilità del reddito. Si tratta di fattori in cui le donne sono mediamente più deboli in ragione dei gap salariali e degli altri problemi noti relativi all'occupazione femminile, alle possibilità e alle modalità di carriera. Si può tentare di neutralizzare alcuni *bias*, ma non si potrà invertire il peso del dato storico, rischiando di trasformare una disuguaglianza perpetrata in un parametro tecnico[25].

È opportuno ribadire come, anche negli esempi svolti[26], il metodo *by design* funziona, dal

---

[24] A. Fabris, N. Baranowska, M.J. Dennis, D. Graus, P. Hacker, J. Saldivar, F. Zuiderveen Borgesius, A.J. Biega, *Fairness and Bias in Algorithmic Hiring: a Multidisciplinary Survey*, in *ACM Transactions on Intelligent Systems and Technology*, January, 2025, 1-18; S. Zhang, P. Kuhn, *Measuring Bias in Job Recommender Systems: Auditing the Algorithms*, in *IZA Discussion Papers,* 17245, August, 2024, 1-67.

[25] Z. Song, S. Rehman, C. Ping, Y. Zhoue, P. Washingtonfand, R. Verschueren, *Do FinTech algorithms reduce gender inequality in banksloans? A quantitative study from the USA*, in *Journal of Applied Economics*, 27(1), 2024, 1-18; C. Hurlin, C. Pérignon, S. Saurin, *The Fairness of Credit Scoring Models*, in *HEC Paris Research Paper*, FIN-2021-1411, 2024, 1-69.

[26] Interessanti anche gli studi condotti in merito alla valutazione accademica (D.W. Aksnes, L.P. Nygaard, R.B. Reiling, *A matter of time? How absence from work affects gender gaps in research productivity*, in *Higher Education*, 2025; M. Thelwall, K. Kousha, E. Stuart, M. Makita, M. Abdoli, P. Wilson, J. Levitt, *Do bibliometrics introduce gender, institutional or interdisciplinary biases into research evaluations?*, in *Elsevier Research Policy*, 52(8), 2023) e all'accesso alle prestazioni di welfare (L.M. Haitsma, *Addressing discrimination in algorithmic profiling: Examining risk governance in Dutch public social security agencies*, in *European Journal of Social Security*, 27(2), 2025, 191-212).
Gli esempi scelti riguardano ambiti diversi – talvolta pubblici, talvolta privati – nei quali vengono impiegati sistemi decisionali automatizzati. Questa scelta è motivata dal fatto che l'intelligenza artificiale incide infatti tanto sull'esercizio delle funzioni pubbliche quanto su attività economiche private che producono effetti rilevanti sull'accesso a diritti, opportunità e risorse. In entrambi i contesti, i sistemi algoritmici possono incidere in modo significativo sulle condizioni materiali di vita: dall'accesso al lavoro nel settore pubblico, alla possibilità

momento in cui elimina i *bias* più evidenti e senz'altro evita sbilanciamenti grossolani.
Ma il sistema sarà comunque costruito su concetti – come «successo», «merito» o «affidabilità» – ai quali va dato un contenuto, che spesso riflette assetti sociali storicamente diseguali. Determinati assetti ne escono stabilizzati, trasformando scelte contingenti in esiti apparentemente naturali[27].
Pur intervenendo sulla struttura, quindi, qualcosa può restare fuori e, a quel punto, è difficile da recuperare o sanare, perché la formalizzazione dell'uguaglianza di genere produrrebbe un sistema chiuso nel modello che è stato costruito e nel quale rimarrebbe incorporata[28]. Una sistemazione che, proprio perché volta a certificare *ex ante*, renderebbe difficile una contestazione *ex post*, chiudendo potenzialmente quel conflitto che è parte essenziale dell'evoluzione dei principi costituzionale e dell'avanzamento della tutela dei diritti.
Il conflitto rischia di indebolirsi, spostandosi dalla rilevazione e soluzione del caso concreto (l'effetto) all'analisi delle metriche e del modello generativo: un conflitto meno sociale e più tecnico, più specialistico e, soprattutto, meno accessibile alla giurisdizione ordinaria.
Questo anche in ragione della dimensione probabilistica dei sistemi algoritmici, in forza dei quali le decisioni sono assunte sulla base di soglie di rischio e modelli predittivi che sono strutturati in modo da accettare anche una componente fisiologica di errore[29].
Il principio di uguaglianza guarda invece al singolo svantaggio, senza tener conto di medie o calcoli, e senza poter accettare che un soggetto resti in una condizione di discriminazione per ragioni «statistiche» o per situazioni talmente peculiari da sfuggire alla predizione standardizzata. L'art. 3, Cost. è un parametro che si applica a condizioni materiali di disuguaglianza e che presuppone la possibilità di rivedere le decisioni alla luce di ciò che generano, indipendentemente dal fatto che siano frutto di situazioni previste o non previste. In altre parole, l'art. 3, Cost. ha una dimensione dinamica che è caratteristica principale del principio di uguaglianza sostanziale.
Pertanto, può concludersi che i modelli di *equality by design* che anticipano, incorporano e formalizzano il principio di uguaglianza possono essere solo parzialmente la soluzione alle discriminazioni di genere potenzialmente perpetrate dall'intelligenza artificiale, dato che il nucleo costituzionale dell'uguaglianza sostanziale richiede un costante controllo giuridico anche sugli effetti, la possibilità costante di revisione e, soprattutto, spazio per la

di ottenere credito nel settore finanziario. L'analisi congiunta consente di evidenziare come le dinamiche di discriminazione algoritmica possano manifestarsi trasversalmente rispetto alla distinzione tra pubblico e privato, pur richiedendo strumenti giuridici di controllo in parte differenti. Ciò che rileva, ai presenti fini, è la dimostrazione del permanere di discriminazioni ineliminate e difficilmente eliminabili.

[27] Spiega A. Papa, *Intelligenza Artificiale e decisioni pubbliche tra tecnica, politica e tutela dei diritti*, in *federalismi,it*, 22, 2022, 109: «in presenza di sistemi di IA basati sull'utilizzo dei dati, appare indispensabile garantire che questi ultimi, nella loro forma di data set, non nascondano discriminazioni. Come è noto, queste ultime rappresentano, nell'approccio tecnologico, un errore di valutazione, un elemento (anche nella sua possibile forma del preconcetto) che rischia di minare la correttezza e l'affidabilità dei risultati di un'analisi ma non ne inficia la validità complessiva; nella prospettiva giuridica esse rappresentano, invece, un elemento di grande criticità soprattutto qualora siano la conseguenza di stereotipiche sono in sé suscettibili di produrre una scelta discriminatoria, ossia l'esatto contrario dell'obiettivo stabilito».

[28] Come sostiene E. Stradella, *Stereotipi e discriminazioni*, cit., 396: «D'altra parte, le questioni relative alla qualità dei dati, e alla presenza di *bias*, paiono assumere sempre maggiore rilevanza, anche nella riflessione dei tecnologi, ragione per cui non sembra irrealistico ritenere che nei prossimi anni sarà possibile progettare sistemi antidiscriminatori *by default* e *by design*, in grado di superare le criticità delineate per lo meno rispetto alla presenza di *bias* all'interno dei sistemi».

[29] Lo chiarisce E. Falletti, *L'Artificial Intelligence Act Proposal e la regolamentazione degli algoritmi predittivi: luci e ombre*, in *Rivista interdisciplinare sul diritto delle amministrazioni pubbliche*, 4, 2023, 180: «l'utilizzo dei modelli predittivi si basa sulla loro efficienza nelle decisioni seriali, pertanto il loro margine di errore, che può presentare importanti conseguenze discriminatorie, viene considerato, malgrado ciò, accettabile».

contestazione ed il conflitto. Uno spazio che non sta «nel» codice algoritmico, ma è «oltre» il codice.
L'intelligenza artificiale diventa così un banco di prova per la capacità del diritto costituzionale di mantenere aperto lo spazio del giudizio sulle diseguaglianze di fatto, riaffermando la centralità dell'eguaglianza sostanziale come criterio di valutazione continua delle condizioni di esercizio dei diritti fondamentali.

4. Conflitto e controllo per una tutela dinamica nell'era dell'intelligenza artificiale.

Ciò che si tenta di sostenere, in sintesi, è che il diritto costituzionale deve continuare ad operare come ha sempre fatto anche nell'era dell'intelligenza artificiale, ma su oggetti diversi. Questo significa che deve continuare a controllare gli effetti delle decisioni sulle condizioni di vita delle persone, al di là della correttezza delle procedure che le hanno prodotte. Nel caso dell'intelligenza artificiale, al di là del fatto che un modello sia stato ben progettato e costruito, bisogna considerare che un sistema di selezione automatica può essere conforme alle regole – e quindi ai requisiti del fornitore – usare variabili neutrali e superare tutti gli audit, ma se, a causa delle modalità di impiego adottate dal *deployer*, finisce per escludere sistematicamente le donne, rafforzare forme di segregazione occupazionale e peggiorare le opportunità, allora è il momento in cui deve intervenire il controllo giuridico[30].
Tale controllo opera su diversi piani. Sul piano giurisdizionale, il giudice non può valutare l'algoritmo, ma la decisione che incide su una persona o su un gruppo, applicando i tradizionali criteri di ragionevolezza e proporzionalità, in coerenza con quanto previsto dall'art. 3, Cost.[31]: anche una decisione tecnicamente corretta può essere costituzionalmente illegittima nei suoi effetti.
Questo tipo di controllo serve a verificare le condizioni materiali di esercizio dei diritti, ossia comprendere se persistono sistematici svantaggi di genere o distorsioni *gender based* nell'accesso a diritti o opportunità. Continuare, in altre parole, a controllare gli effetti indiretti, gli ostacoli strutturali e le discriminazioni di fatto che persistono nonostante l'utilizza di sistemi tecnici orientati.
Un controllo che serve anche a dare continuità al monitoraggio di queste situazioni: gli interventi sulla struttura tecnica, infatti, fotografano un momento, mentre molte disuguaglianze di genere emergono col tempo in relazione alle evoluzioni sociali, si accumulano e diventano progressivamente strutturali. Il controllo giurisdizionale, in questo senso, serve a intervenire laddove si consolidino discriminazioni di genere non visibili al momento della strutturazione algoritmica, ovvero da essa indirettamente causate.
Gli strumenti di questo controllo sono quelli classici del diritto costituzionale, solo da applicare a oggetti nuovi: sindacato sulle decisioni automatizzate; possibilità di contestare decisioni individuali o sistemiche; obbligo di revisione e correzione di sistemi che producono discriminazioni; eventualmente, forme di responsabilità giuridica per effetti lesivi persistenti.
Questo non significa che il giudice debba entrare nel funzionamento tecnico del sistema: non riscrive il codice, non sceglie le metriche, non ottimizza il modello né tantomeno spiega come deve essere tecnicamente impostato; non si sostituisce, dunque, né al progettista né al legislatore, che mantengono queste prerogative. Il controllo giurisdizionale infatti non è un controllo tecnico, ma giuridico: il giudice deve solo valutare se una decisione, qualunque sia

[30] Evidenzia C. Nardocci, *Discrimination revised*, cit., 181, che la discriminazione algoritmica non è solo pregiudizio umano tradotto in codice (*human-driven*), ma una forma separata di trattamento irragionevole che nasce dall'interazione tra agenzia umana e agenzia automatica.
[31] L. Di Majo, op. cit., 48-55, sottolinea come il ricorso all'intelligenza artificiale rischi di incidere non solo sugli strumenti decisionali, ma sulla stessa configurazione del ruolo del giudice, ponendo in discussione il paradigma del libero convincimento.

il mezzo che l'ha prodotta, comporti delle discriminazioni di genere nei suoi effetti.
La discrezionalità tecnica, del resto, è sempre stata sindacabile laddove ha prodotto effetti irragionevoli o sproporzionati e la decisione algoritmica non fa eccezione[32].
Come già osservato, le diseguaglianze di genere conseguente l'utilizzo di codici non dipendono da una singola scelta tecnica, ma emergono come effetti sistemici che si manifestano nel corso del tempo. Al giudice non si chiede di intervenire sulla struttura dei codici, correggendone le parti che tecnicamente creano squilibri di genere, ma solo di riconoscere che certi effetti non sono costituzionalmente tollerabili; non deve ingerire nel funzionamento tecnico, ma rimanere sul piano tradizionale di valutazione dell'effetto delle decisioni sulle condizioni materiali di esercizio dei diritti, secondo i parametri costituzionalmente consolidati di ragionevolezza e proporzionalità.
Nell'integrare gli aspetti virtuosi degli interventi sulla progettazione degli algoritmi affinché essi siano in grado di neutralizzare i *bias* di genere, dunque, resta fermo il principio dell'irrinunciabilità e del ruolo fondamentale del controllo giurisdizionale sugli effetti, sia per limitare e correggere le distorsioni dei sistemi, sia, soprattutto, per mantenere sempre aperto e vivo il conflitto antidiscriminatorio.

Abstract. L'articolo esamina il rapporto tra intelligenza artificiale e uguaglianza di genere alla luce dell'art. 3 della Costituzione, concentrandosi sui limiti costituzionali della formalizzazione tecnica dell'uguaglianza. I sistemi di intelligenza artificiale operano su dati che riflettono diseguaglianze storicamente sedimentate, contribuendo alla loro stabilizzazione attraverso processi decisionali automatizzati. Il contributo ricostruisce l'evoluzione del dibattito normativo e dottrinale verso modelli di prevenzione della discriminazione fondati sull'*equal treatment by design*, mettendone in luce le implicazioni per l'eguaglianza sostanziale. Alla luce della giurisprudenza costituzionale, l'eguaglianza viene interpretata come principio dinamico e relazionale, orientato alla valutazione degli effetti concreti delle decisioni sulle condizioni materiali di vita delle persone. L'analisi evidenzia come la traduzione dell'eguaglianza in parametri tecnici *ex ante* incida sulla sua funzione critica, rendendo necessario il mantenimento di spazi di controllo, contestazione e revisione delle decisioni automatizzate. L'intelligenza artificiale emerge così come un terreno decisivo per la capacità del diritto costituzionale di preservare la portata trasformativa dell'eguaglianza sostanziale.

Abstract. This article examines the relationship between artificial intelligence and gender equality through the lens of Article 3 of the Italian Constitution, focusing on the constitutional limits of the technical formalization of equality. Artificial intelligence systems operate on data shaped by historically entrenched inequalities, contributing to their stabilization through automated decision-making processes. The analysis traces the evolution of legal and scholarly debates toward preventive models of discrimination, centered on *equal treatment by design*, and explores their implications for substantive equality. Drawing on constitutional case law, equality is framed as a dynamic and relational principle oriented toward assessing the concrete effects of decisions on individuals' material conditions. The article highlights how translating equality into ex ante technical parameters affects its critical function, underscoring the importance of maintaining mechanisms of control, contestation,

[32] Cons. St., sez. VI, sent. n. 2270 del 2019, par. 8.2 «l'utilizzo di procedure robotizzate non può, tuttavia, essere motivo di elusione dei principi che conformano il nostro ordinamento e che regolano lo svolgersi dell'attività amministrativa». V. G. Botto, *Decisione algoritmica, discrezionalità e sindacato del giudice amministrativo*, in *Federalismi.it*, 17, 2024, 1-24.

and revision of automated decisions. Artificial intelligence thus emerges as a crucial testing ground for the ability of constitutional law to preserve the transformative dimension of substantive equality in the context of automated decision-making.

Parole chiave. Intelligenza artificiale – Uguaglianza sostanziale – Discriminazione di genere – Decisione algoritmica – Gender bias – *equality by design*.

Key words. Artificial intelligence – Substantive equality – Gender discrimination – Algorithmic decision-making – Gender bias – *Equality by design*.