

27



# QUADERNI UNSI<sup>C</sup>

**Intelligenza artificiale e pregiudizi di genere**  
Implicazioni sociali e giuridiche

Giuseppe Mosa

# QUADERNI



## Intelligenza artificiale e pregiudizi di genere

Implicazioni sociali e giuridiche

Giuseppe Mosa\*

\*Avvocato - Consulente di imprese, organizzazioni sindacali ed enti pubblici

## Indice

<i>A mò di prefazione</i> .....	4
1. Introduzione .....	7
2. Bias cognitivi e discriminazione .....	10
3. Soluzioni per uno sviluppo inclusivo di genere nell'IA .....	16
4. Rimedi tecnici e giuridici.....	18
5. Conclusioni .....	24

*Un giorno le macchine riusciranno  
a risolvere tutti i problemi, ma mai  
nessuna di esse potrà porne uno.*

*- Albert Einstein*

## **A mò di prefazione**

Una delle più grandi ambizioni dell'uomo è stata sin dall'antichità quella di replicare sé stesso, la sua immagine, la sua mente e le sue capacità, al fine di creare un suo *clone* più efficiente e migliorare le proprie condizioni di vita: basti pensare al mito greco di Prometeo<sup>1</sup>, il titano progressista che, opponendosi alle volontà divine, rubò il fuoco agli Dei, allo scopo di offrire all'umanità un destino più favorevole, o a Paracelso, l'alchimista del XIV secolo che studiò un processo chimico-biologico al fine di creare un *Homunculus*, ovvero una forma di vita umana originata grazie all'alchimia.

Da sempre l'uomo ha immaginato di creare qualcosa sempre più simile a sé stesso o, ancora, di poter far prendere vita ad una massa inanimata.

---

<sup>1</sup> Prometeo era il Titano che rubò il fuoco agli dei dell'Olimpo e lo diede agli esseri umani, avviandoci su un percorso di gloria e disastro e incorrendo nell'ira gelosa di Zeus. Nel mondo moderno, soprattutto dall'inizio della rivoluzione industriale, è servito come simbolo di progresso e pericolo, un avatar sia del potere liberatorio della conoscenza che dei pericoli dello sfondamento tecnologico. Sul punto G. Anders, *L'obsolescence de l'homme. Sur l'âme à l'époque de la deuxième révolution industrielle*, Éditions Ivrea, Paris 2002.

Già nei miti greci la storia di Pigmalione e Galatea, in cui la donna prende vita da una statua creata dallo stesso Pigmalione<sup>2</sup>, riporta gli stessi desideri che si sono riscontrati nella creazione dell'intelligenza artificiale (IA).

L'intelligenza artificiale altro non è che il cercare di riprodurre la mente umana e le sue peculiarità, cercando di comprendere sempre più i dettagli e ciò che la contraddistingue nel suo funzionamento.

Ai giorni nostri, l'intelligenza artificiale fa parte della nostra quotidianità. La ritroviamo all'interno dei nostri *smartphone* e nelle *app* che sono al loro interno, nelle nostre case con *Alexa* che spegne le luci e riproduce la musica ad un nostro semplice comando. È presente sul luogo di lavoro, negli ospedali, nelle industrie e in tutti i posti a noi familiari.

Senza rendercene conto l'IA ha cominciato a far parte progressivamente ed ineluttabilmente della nostra vita. Nell'immaginario comune l'Intelligenza Artificiale è associata ai *robot* e a macchine che prendono vita, ma in realtà è molto più semplice.

Infatti, come già detto, è intorno a noi, "nascosta" dietro azioni e compiti svolti che si possono definire normali, perché possono essere ricondotti a quelli che potrebbe fare una mente umana.

La differenza sostanziale è che all'interno della mente umana vi è un **ragionamento**, dettato anche da emozioni e sentimenti, che nell'IA sono assenti<sup>3</sup>.

---

<sup>2</sup> Esempi di creature artificiali con sembianze umane, sono ricorrenti nella mitologia greca. Cadmo, il fondatore di Tebe, seppellì denti di drago che si trasformarono in soldati. Efesto, il dio del metallo, creò servi meccanici: damigelle dorate dotate di intelligenza e tavoli a tre gambe che si spostavano di propria volontà. Anche il poeta latino Ovidio aveva narrato nelle *Metamorfosi* (8 d.C.) la trasformazione di un corpo inanimato in una creatura viva.

Per creare un apprendimento l'IA necessita di **algoritmi** che vengono addestrati attraverso un numero molto alto di dati c.d. **dataset**, che la "formerà" per poter svolgere al meglio le proprie azioni.

Nella definizione e messa a punto delle diverse fasi di acquisizione di tali dati ed estrazione dell'informazione, il pregiudizio tipico del sentire umano si insinua e finisce per creare macchine che, similmente a quando capita alle persone, manifestano comportamenti discriminatori.

Con la diffusione pervasiva dell'IA, questo fenomeno diventa uno strumento formidabile per alimentare e rinforzare stereotipi e pregiudizi.

All'interno della comunità scientifica si sta sviluppando un vivace dibattito su questo tema che potrebbe e dovrebbe portare ad aumentare la consapevolezza di consumatori e cittadini e alla definizione di pratiche e strumenti per **mitigare gli effetti delle discriminazioni** a opera delle macchine.

Dal canto suo il giurista ha il compito di prevenire questi problemi al fine di elaborare gli strumenti normativi e giurisprudenziali per "regolare" in modo corretto queste trasformazioni epocali.

Infatti, al contrario di quanto sostenuto dai più, riteniamo che sia sciocco pensare che leggi e regolamenti possano e debbano correre appresso all'incessante evolversi dell'IA: ciò perché le norme non riguardano il RITMO ma la DIREZIONE dell'innovazione, atteso che il loro fine ultimo è quello di guidare il corretto sviluppo di una società<sup>4</sup>.

---

<sup>3</sup> A. Fichera, *Tecnofila e tecnocrazia ?, riflessioni per i "prossimi terrestri"*, Edizioni Sindacali, 2024, Roma, p. 7 ss.

<sup>4</sup> L. Floridi, *Etica dell'intelligenza artificiale*, Raffaello Cortina Editore, Milano, 2023, p. 86 ss.

# 1. Introduzione

L'inarrestabile avanzamento dell'**intelligenza artificiale (IA)** ha catapultato gli **algoritmi** decisionali al centro della nostra vita quotidiana, dall'uso di assistenti vocali ai sistemi di riconoscimento facciale e motori di ricerca<sup>5</sup>.

Questo sviluppo ha portato innegabili vantaggi economici e sociali, ma è giunto il momento di considerare il suo **costo etico**, in particolare per quanto riguarda i pregiudizi e le discriminazioni di genere che possono emergere nei confronti di gruppi specifici.

Già priorità dell'agenda dell'Unione dal 2018, l'Intelligenza Artificiale ha manifestato l'esigenza di essere regolata da un approccio coordinato a livello europeo.

---

<sup>5</sup> Per algoritmo si intende la procedura di calcolo connotato da un numero finito di regole che conduce al risultato dopo un numero finito di operazioni (applicazioni delle regole).

Il concetto di algoritmo risale già ai Babilonesi, il quali, quasi quattro millenni fa, riescono ad elaborare delle tavolette che, in maniera sbalorditiva, *"non sono propriamente delle soluzioni a problemi individuali specifici, ma si tratta di procedure generali per risolvere una intera categoria di problemi"* (Donald Knuth).

Gli algoritmi sono pertanto istruzioni matematiche, procedure e insieme di istruzioni passo-passo utilizzate per eseguire classi di attività e/o risolvere classi di problemi. Vi sono varie categorie di algoritmi informatici, alcuni dei quali implementati per consentire l'apprendimento automatico delle macchine (*machine learning*).

Sulla scorta degli algoritmi di *machine learning* è possibile istruire ed addestrare altri algoritmi su cosa generare: come accade per un bimbo a cui la maestra insegna le lettere dell'alfabeto attraverso l'abbecedario, così l'intelligenza artificiale impara da un set di dati a produrre un output predefinito (*c.d. machine learning supervisionato dall'uomo*).

Nei sistemi più evoluti, l'intelligenza artificiale impara a identificare processi e schemi complessi senza la guida attenta di una mente umana (*c.d. machine learning non supervisionato*): come il bimbo che ormai ha imparato le lettere, le parole, e la grammatica, ed è così in grado di svolgere dei propri ragionamenti producendo parole e frasi che non sono *output* predefiniti.

Oggi, infatti, si parla di *deep learning* quale evoluzione del *machine learning* realizzata mediante **reti neurali artificiali**, organizzate su più livelli, che elaborano le informazioni in modo più evoluto, senza necessità di dati strutturati né dell'intervento umano.

Si tratta della svolta epocale dell'intelligenza artificiale generativa che non solo apprende, ma genera autonomamente dei contenuti alla luce di quello che ha appreso.

Gli algoritmi, sfruttando enormi volumi di macro e micro-dati, possono produrre modelli per coadiuvare ma anche per influenzare le persone e le decisioni che riguardano le persone, in ogni ambito della vita.

Il suo crescente utilizzo richiede un'attenzione adeguata dovuta alle conseguenti implicazioni etiche, quali pregiudizi e discriminazioni di genere nei confronti di determinati gruppi che, così come esistenti all'interno della società, possono essere integrati negli algoritmi, intenzionalmente o no.

Nonostante il settore dell'IA stia vivendo una rapida crescita in Europa, emerge

un dato allarmante: solo il 16% degli operatori in questo campo sono donne<sup>6</sup>.



Questa disparità è in parte attribuibile, come evidenziato dal **rapporto UNESCO 2023** alla persistente sotto rappresentanza delle donne in ambito tecnologico e, più in generale, nelle **discipline STEM (Scienza, Tecnologia,**

**Ingegneria e Matematica)**, caratterizzate da un lungo e consolidato divario di genere<sup>7</sup>.

La soluzione a questo problema richiede un duplice approccio: a) approfondire la cultura informatica e b) promuovere un ambiente in cui le donne possano sentirsi apprezzate e supportate, riducendo così il rischio di pregiudizi di genere nell'IA e, di conseguenza, il notevole *turnover* e abbandono del settore.

Un **rapporto dell'Organizzazione Internazionale del Lavoro** ha poi evidenziato che le donne sono ad un rischio più elevato di *turn-over* da parte

---

<sup>6</sup> Ci sono nel mondo 234 milioni di utenti donne in meno rispetto agli uomini, secondo una statistica del 2021, che generano dati che poi verranno utilizzati nei processi di AI. C'è inoltre un altro aspetto, ancora più visibile: l'algoritmo lo scrive una persona o un pool. Nella maggior parte dei casi sono uomini e non donne a scriverli, in qualsiasi posto nel mondo; nel campo specifico dell'AI, le donne rappresentano in Europa il 16% del personale impegnato.

<sup>7</sup> Negli ambiti della cultura e delle scienze persiste il gender gap: solo il 33% dei premi cinematografici è assegnato a donne e solo 1/3 delle donne è laureato nelle discipline STEM (*Science Technology Engineering and Mathematics*) a livello mondiale.



dell'IA rispetto agli uomini, con il 79% delle donne lavoratrici occupate in settori suscettibili all'automazione.

Questo perché, a causa di stereotipi di genere, sono state storicamente associate a ruoli come segretarie e assistenti sanitari, considerati tradizionalmente femminili. Questi settori, insieme ad altri come l'assistenza sociale e insegnamento, sono più suscettibili all'automazione a causa della presenza di compiti ripetitivi e routinari.

Questo rischio è significativamente più basso per l'occupazione maschile, evidenziando una disparità che richiede un approccio strategico e inclusivo.

Secondo il **rapporto Future of Jobs** del **World Economic Forum** si prevede, da un lato, la creazione di nuovi 69 milioni di posti di lavoro e, dall'altro, la contestuale cancellazione di oltre 83 milioni di posti, con una diminuzione del 2% dell'attuale occupazione. Tra queste cancellazioni, la situazione peggiore è quella dell'impiegato di inserimento dati, seguito dalle posizioni di segretario amministrativo e di contabilità.

## 2. Bias cognitivi e discriminazione

Gli algoritmi di apprendimento automatico imparano da grandi set di dati. Cosa accade se tali dati contengono pregiudizi storici o sociali? L'algoritmo li apprenderà e li replicherà.

Ad esempio, se un **sistema di AI per la selezione del personale** viene addestrato su dati storici in cui le posizioni di leadership sono prevalentemente occupate da uomini, potrebbe sviluppare una preferenza per i candidati di genere maschile.

Programmatori e sviluppatori possono, involontariamente, immettere i propri pregiudizi nella progettazione e nella realizzazione degli algoritmi.

Nella selezione degli elementi da prescegliere e nell'individuazione di determinati caratteri, in luogo di altri, si pongono le basi per la **struttura cognitiva** che poi troverà traduzione concreta nell'algoritmo. Tali elementi possono afferire a qualsiasi aspetto del pensiero, della società, della vita, della natura.

Già in sede di *trial*, alcuni algoritmi si rivelano sconsiderati: vi sono applicazioni a cui ponendo determinati quesiti, tendono a rispondere tralasciando inopinatamente alcuni aspetti critici della realtà, come l'idoneità all'accesso a determinati servizi di assistenza, o la parità di trattamento tra i generi o etnie.

I meccanismi di automatizzazione nelle procedure decisionali certamente costituiscono un settore molto sensibile, per la ricaduta sulla realtà, sortita dagli effetti della loro applicazione. Pertanto, se da un lato, le decisioni

automatizzate rappresentano una delle aree più promettenti dell'intelligenza artificiale, dall'altro sono tra le più problematiche ed inesplorate.

Quando gli algoritmi vengono utilizzati per prendere decisioni che influenzano aspetti critici della vita delle persone, come l'idoneità al credito o le decisioni giudiziarie, oppure le parità di trattamento, le questioni di equità, trasparenza e giustizia diventano di primaria importanza.

Siamo di fronte al problema dei *bias*.

Cosa è un *bias*?

**È una distorsione cognitiva, un costrutto derivante da percezioni errate, automatismi mentali che generano credenze, che inducono a veloci valutazioni e decisioni, che, in ultima analisi, conducono a formare un pensiero che contiene pregiudizi e stereotipi.**

Si tratta travisamenti della realtà che finiscono per influenzare, anche significativamente, i nostri processi mentali, da quello valutativo a quello decisionale.

Tali distorsioni possono essere presenti nei processi di intelligenza artificiale.

In sede di elaborazione e implementazione degli algoritmi, tra gli *input* inseriti, possono esservi anche elementi distorti che danno luogo a percezioni stereotipate, e per questo non obiettive, della realtà, conducendo a generare produzioni inficiate<sup>8</sup>.

---

<sup>8</sup> M.D., Taramundi, "Le sfide della discriminazione algoritmica" in "GenIUS", rivista di studi giuridici sull'orientamento sessuale e l'identità di genere, online first, 24 ottobre 2022.

In definitiva, nei sistemi *biased* si riscontra il concreto rischio di discriminare sistematicamente e ingiustamente gli individui o alcuni gruppi sociali: così i pregiudizi sociali vengono perpetuati con gravi conseguenze.

I *bias* (umani, statistici, computazionali) possono essere presenti in tutte le fasi di vita dell'algoritmo, introdotti a partire dalla selezione dei dati e delle sorgenti, nell'elaborazione o nel raggruppamento dei dati, mutuati da altri sistemi, possono non essere presenti nel dato di partenza ma generati nel processo di apprendimento tramite connessioni complesse e comunque andare ad incidere sull'output finale, inficiandone, in qualche misura, la congruità o l'imparzialità.

In sostanza, algoritmi viziati conducono a decisioni non corrette, inadeguate, in ultima analisi non oggettivamente rispondenti all'effettiva configurazione della realtà sottoposta.

Un algoritmo che debba gestire ed eseguire procedure di reclutamento di una determinata tipologia di lavoratori, mediante analisi di requisiti, titoli, riserve e preferenze, con la conseguente elaborazione di una graduatoria, nel caso in cui contenga uno o più *bias*, può condurre ad un esito aberrante: formulare una graduatoria iniqua e discriminatoria, ponendo ai livelli superiori chi andava collocato, secondo la corretta applicazione di regole predefinite e universali, ai gradini sottostanti, o viceversa. O ancora, la scelta preselettiva di determinati gusti e modi di pensare, introitati negli algoritmi, può influenzare il sistema decisionale degli utenti potenziali di una determinata campagna promozionale di un prodotto commerciale, o quello elettorale di una linea politica.

Ovviamente i *bias* non vanno considerati come dei meri errori di implementazione dell'algoritmo, ma come lo specchio dei pregiudizi insiti in una società, o talvolta, peggio, come lo strumento, nascosto e insidioso, finalizzato a perseguire e conseguire forme, più o meno velate, di influenza in danno ai destinatari.

**Fondamentale appare il ricorso a dati quanto più rappresentativi e di alta qualità per addestrare il sistema: necessario per garantire un'intelligenza artificiale quanto più giusta, equa e democratica.**

I sistemi di intelligenza artificiale dovrebbero rispecchiare la nostra società in tutta la sua diversità, e non soltanto alcuni aspetti della stessa, altrimenti di *default* si finirebbe per perpetuare la tendenza a pregiudizi e discriminazioni.

Nei casi di carenza del ricorso a dati sufficientemente rappresentativi si parla di ***representation gap***<sup>9</sup>.

I sistemi ad apprendimento automatico, come detto, imparano estraendo informazioni da grandi quantità di dati, se i dati sono espressione di pregiudizi, non sono correttamente rappresentativi della realtà e della varietà umana, ma solo di una sua parte, oppure se riflettono modelli storici discriminatori, permangono, e si amplificano, gli stereotipi e i pregiudizi, inclusi quelli di genere con i conseguenti effetti discriminatori.

---

<sup>9</sup> The representation gap refers to the fact that a significant proportion of workers in non-union enterprises would like to be represented either by a trade union or by some other form of representative body. It is a product of the decline of trade unionism and collective bargaining and the continued aspiration of many employees to participate in the government of industry through representative institutions. The existence of the representation gap suggests that for many workers management communication and attempts at employee involvement are not an adequate substitute for independent representation at work.

Il comportamento della macchina, in pratica, diventa iniquo, come se la macchina, avendo conosciuto solo una parte di un certo contesto, ne avesse elaborato una conoscenza ristretta e limitata.

A volte tali *bias* risultano chiari e percepibili, altre volte più sfumati e insidiosi in quanto difficile misurarne la portata effettiva e l'espressione.

Esaminiamo più da vicino come questo fenomeno si manifesti e le sue **implicazioni sociali e giuridiche**, in rapporto allo stereotipo probabilmente più diffuso, quello di genere.

Il *bias* di genere è un comportamento che mostra un favoritismo verso un genere rispetto a un altro. Si tratta di una forma di pregiudizio più o meno inconscio o implicito, che si verifica quando un individuo attribuisce, in maniera apodittica, determinati atteggiamenti e stereotipi a un certo genere di persona o gruppo di persone.

La presenza di un tale tipo di *bias* nella struttura di AI produce sistematicamente e ingiustificatamente risultati meno favorevoli, ingiusti o dannosi per un genere rispetto ad un altro.

Il genere è una tipizzazione sociale, culturale e psicologica. È un costrutto sociale, a differenza del sesso che caratterizza e definisce l'aspetto biologico-anatomico di una persona.

Certi concetti evolvono nel tempo, e così l'identità di genere, con il superamento di una concezione non più pienamente binaria.

L'identità di genere riguarda il modo in cui ci si identifica con il proprio sesso ed in cui si intende essere percepiti dagli altri, può corrispondere al sesso biologico o discostarsene ed è un concetto diverso dall'orientamento sessuale.

Partita come differenze tra maschi e femmine, il discorso sul genere si è socialmente evoluto sino a considerare l'esistenza di una molteplicità di identità di genere che va oltre la dualità (si pensi alle persone non binarie come gli *a-gender* e i *transgender*).

### 3. Soluzioni per uno sviluppo inclusivo di genere nell'IA

Questo scenario sottolinea l'importanza di **individuare le modalità con cui queste tecnologie interagiscono con i ruoli sociali femminili**, trattando delle possibili soluzioni per arginare il problema e promuovere quindi uno sviluppo inclusivo di genere attraverso il rilevamento e la correzione dei pregiudizi negli algoritmi.

Soluzione chiave sarà l'implementazione di controlli regolari avanzati per garantire l'assenza di distorsioni di genere e in parallelo **assicurare la rappresentanza di donne e gruppi sottorappresentati all'interno del settore<sup>10</sup>**.

È necessario, perciò, dotare le donne delle competenze digitali necessarie e includerle nello sviluppo delle tecnologie di intelligenza artificiale, promuovendo nuovi posti di lavoro, una migliore selezione e, soprattutto, l'alfabetizzazione digitale, da realizzare con programmi specifici sviluppati in istituti scolastici, organizzazioni e aziende e incoraggiando gli uomini a sostenere l'uguaglianza di genere ed essere alleati attivi, in un impegno collettivo cruciale per **plasmare un futuro in cui l'IA rifletta la diversità di genere**.

La mancanza di rappresentanza femminile in tali settori ha conseguenze rilevanti nell'ambito dell'utilizzo dell'IA, poiché **i dati utilizzati per**

---

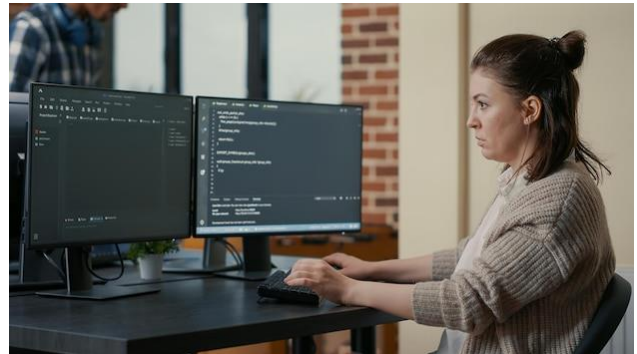
<sup>10</sup> All'inizio di questo anno, il governo del Regno Unito, nel Rapporto di aggiornamento sulla Gran Bretagna inclusiva, ha riferito di aver utilizzato gli stessi strumenti basati sull'IA per elaborare una guida utile ai datori nella misurazione dei divari retributivi etnici nella forza lavoro. Il fine ultimo è quello di garantire che i benefici di questa tecnologia siano distribuiti equamente, con la consapevolezza delle possibili discriminazioni sulle minoranze etniche.

L'art. 72 dell'iniziativa prevede lo sviluppo di una regolamentazione dell'IA, attuata con il Libro Bianco, un documento che definisce un approccio orientato ai risultati e teso al rafforzamento della fiducia da parte degli utenti.



**addestrare gli algoritmi riflettono spesso i pregiudizi culturali e di genere radicati nella società.**

Aumentando la percentuale di donne nei ruoli di sviluppo e utilizzo di questi sistemi, ci sono maggiori possibilità di correzione di questi pregiudizi.



Prova tangibile la troviamo con

**Amazon**, che nel 2018 ha sollevato un monito cruciale sull'impiego dell'IA nella selezione del personale abbandonando un sistema basato sull'intelligenza artificiale perché dimostrava di preferire candidati maschi nelle posizioni tecniche<sup>11</sup>.

Addestrato su dati provenienti da *curriculum* inviati in un decennio, il sistema favoriva inconsciamente i candidati di sesso maschile a causa della prevalenza di questi nell'insieme dei documenti assunti come base di apprendimento, per motivazioni dovute a disuguaglianze di genere nella formazione tecnica e nella cultura aziendale, spesso discriminatoria, insieme alla presenza di **barriere di accesso per le donne nel campo tecnico**.

Al fine di consentire una selezione imparziale ed equa dei candidati sarebbe stato necessario agire verso un'adeguata progettazione del sistema, che poteva essere utilizzato con risultati significativi nell'analisi di grandi volumi di candidati o per la ricerca di posizioni con competenze particolari.

---

<sup>11</sup> L'algoritmo usato da *Amazon* esaminava e classificava i dipendenti tramite un set di dati di addestramento che presentava i profili dei dipendenti assunti in passato. Sebbene il genere o l'etnia non facessero parte dell'insieme di variabili inserite nel modello, l'algoritmo autonomamente correlava (*proxy discrimination*) queste caratteristiche alle prestazioni dei candidati. In conclusione, i dati finivano per riflettere le discriminazioni sociali presenti nel contesto sociale e amplificarle su un grande numero di individui.

## 4. Rimedi tecnici e giuridici

I rischi generati dai *bias* cognitivi di cui sopra possono essere in qualche misura prevenuti o limitati, mediante **azioni sia di natura tecnica che di tutela legale**<sup>12</sup>.

In primis, su può agire sull'algoritmo stesso, ricorrendo al concetto, sopra illustrato, di dati rappresentativi: addestrare gli algoritmi su un **set di dati il più diversificato e rappresentativo possibile**, monitorando costantemente gli output prodotti al fine di intercettarne i *bias* e rettificarne gli algoritmi.

Nel processo di selezione dei dati, è senza dubbio importante includere, sia nella fase di creazione che in quella di revisione dell'algoritmo, non unicamente tecnici informatici, ma una collegialità di esperti di più domini del sapere, con la finalità di prevenire la creazione involontaria di discriminazioni.

Dal punto di vista eminentemente giuridico, il diritto antidiscriminatorio risulta molto articolato e trova le sue fonti sia in norme di diritto nazionale che europee<sup>13</sup>.

In primo luogo, va annoverata la **Carta dei diritti fondamentali dell'Unione Europea**, nota anche come Carta di Nizza, che, all'art. 21, rubricato "Non discriminazione", così recita:

*"1. È vietata qualsiasi forma di discriminazione fondata, in particolare, sul sesso, la razza, il colore della pelle o l'origine etnica o sociale, le caratteristiche*

---

<sup>12</sup> G. Mosa, "Intelligenza artificiale: profili etici e giuridici", in "I Quaderni di Mosa", n. 5/2023, Fondolavoro, Roma, 2023.

<sup>13</sup> S. Tommasi, "Algoritmi e nuove forme di discriminazione: uno sguardo al diritto europeo" in Revista de Direito Brasileira, Florianopolis, SC v. 27, n.10, p. 112-129, Set/Dez, 2020.

*genetiche, la lingua, la religione o le convinzioni personali, le opinioni politiche o di qualsiasi altra natura, l'appartenenza ad una minoranza nazionale, il patrimonio, la nascita, la disabilità, l'età o l'orientamento sessuale.*

*2. Nell'ambito d'applicazione dei trattati e fatte salve disposizioni specifiche in essi contenute, è vietata qualsiasi discriminazione in base alla nazionalità."*

L'art. 23 della citata Carta, sancisce al primo comma, *"la parità tra uomini e donne deve essere assicurata in tutti i campi, compreso in materia di occupazione, di lavoro e di retribuzione"*.

L'art. 14 **Convenzione Europea dei Diritti dell'Uomo (CEDU)** sancisce che il godimento dei diritti e delle libertà riconosciuti nella Convenzione *"deve essere assicurato senza nessuna discriminazione, in particolare quelle fondate sul sesso, la razza, il colore, la lingua, la religione, le opinioni politiche o quelle di altro genere, l'origine nazionale o sociale, l'appartenenza a una minoranza nazionale, la ricchezza, la nascita od ogni altra condizione"*.

Vi è un riferimento alla disciplina antidiscriminatoria, connessa al genere, in diverse direttive europee:

1) la Direttiva 79/7/CEE del Consiglio, del 19 dicembre 1978, relativa alla graduale attuazione del principio di parità di trattamento tra gli uomini e le donne in materia di sicurezza sociale;

2) la Direttiva 2006/54/CE del Parlamento europeo e del Consiglio, del 5 luglio 2006, riguardante l'attuazione del principio delle pari opportunità e della parità di trattamento fra uomini e donne in materia di occupazione e impiego;

3) la Direttiva 2010/41/UE del Parlamento europeo e del Consiglio, del 7 luglio 2010 sull'applicazione del principio della parità di trattamento fra gli uomini e le donne che esercitano un'attività autonoma.

In Italia vige il Decreto legislativo, 11 aprile 2006, n. 198 (**c.d. Codice delle Pari Opportunità**), novellato dal Decreto legislativo n. 5/2010.

La Legge n. 162/2021 ha istituito, per i datori di lavoro che attuano politiche *“per ridurre il divario di genere in relazione alle opportunità di crescita in azienda, alla parità salariale a parità di mansioni, alle politiche di gestione delle differenze di genere e alla tutela della maternità”*, una specifica **certificazione della parità di genere** (art. 46 bis d.lgs. 198/2006).

Di recente, lo scorso 2 febbraio 2024, è stato concordato all'unanimità il progetto di **Regolamento dei Paesi UE sull'Artificial Intelligence (c.d. AI Act)**, in vista dell'approvazione definitiva avvenuta il 24 aprile 2024.

**Il nuovo set di regole, il primo al mondo in materia, è stato adottato all'unanimità**, a segnalare la compattezza dell'Unione davanti alle sfide poste dall'IA. Un esito che per il relatore al Parlamento europeo dell'AI Act, Brando Benifei, *“certifica la qualità del lavoro svolto su un tema di grande sensibilità politica e di mercato, che rende l'Ue il capofila globale nella regolamentazione dell'IA”*.

La *ratio iuris* del dell'innovativo impianto normativo si coglie nell'obiettivo precipuo di minimizzare *“il rischio di discriminazione algoritmica”*. Esso, infatti, è formulato in adesione ai suindicati principi di cui alla Carta dei diritti fondamentali dell'Unione europea (art. 21) ed in coerenza a tutte le fonti del diritto antidiscriminatorio prodotto dall'UE.

Prevede una serie di regole da applicare ai sistemi di AI per prevenire i rischi sistemici e garantire "diversità, non discriminazione ed equità", individuando quattro livelli di rischio (**inaccettabile, alto, e limitato, minimo o nullo**)<sup>14</sup>.



1. *Approccio risk-based. Fonte: Rielaborazione Intellera Consulting su bozza AI Act.*

<sup>14</sup> **Rischio inaccettabile:** Sono quelli che contraddicono i valori e i principi fondamentali dell'UE, come il rispetto della dignità umana, della democrazia e dello stato di diritto. Questi sistemi sono vietati o (nel caso della sorveglianza biometrica in tempo reale per motivi di sicurezza) soggetti a severe restrizioni. Ad esempio, sono proibiti i sistemi di IA che manipolano il comportamento umano in modo da eludere la volontà degli utenti, o che consentono lo "scoring sociale" da parte delle autorità pubbliche.

**Rischio elevato:** I sistemi di IA che possono avere un impatto "sistemico", ossia significativo sui diritti fondamentali o sulla sicurezza delle persone. Questi sistemi sono soggetti a rigorosi obblighi e requisiti prima di poter essere immessi sul mercato o utilizzati. Ad esempio, rientrano in questa categoria i sistemi di IA utilizzati per la selezione e il reclutamento del personale, per l'ammissione all'istruzione, per l'erogazione di servizi sociali essenziali, come la Sanità, per la sorveglianza biometrica a distanza (non in tempo reale), per le applicazioni giudiziarie e di polizia, o per la gestione della sicurezza critica delle infrastrutture. Sono ad alto rischio in automatico i sistemi di intelligenza artificiale generativa dotati di potenza di calcolo superiore a  $10^{25}$  Flop.

**Rischio limitato:** I sistemi di IA che presentano che possono influenzare i diritti o le volontà degli utenti, ma in misura minore rispetto ai sistemi ad alto rischio. Questi sistemi sono soggetti a requisiti di trasparenza, che consentono agli utenti di essere consapevoli del fatto che interagiscono con un sistema di IA e di comprenderne le caratteristiche e le limitazioni. Ad esempio, rientrano in questa categoria i sistemi di IA utilizzati per generare o manipolare contenuti audiovisivi (come i deepfake), o per fornire suggerimenti personalizzati (come le chatbot). C'è il diritto a sapere di stare parlando con un bot (invece di un umano) e che quell'immagine è creata o artefatta dall'IA.

**Rischio minimo o nullo:** I sistemi di IA che non hanno alcun impatto diretto sui diritti fondamentali o sulla sicurezza delle persone, e che offrono ampi margini di scelta e controllo agli utenti. Questi sistemi sono liberi da qualsiasi obbligo normativo, al fine di incoraggiare l'innovazione e l'esperimentazione. Ad esempio, rientrano in questa categoria i sistemi di IA utilizzati per scopi ludici (come i videogiochi) o per scopi puramente estetici (come i filtri fotografici).

I sistemi di IA saranno sviluppati e utilizzati in modo da includere soggetti diversi e promuovere la parità di accesso, l'uguaglianza di genere e la diversità culturale, evitando nel contempo effetti discriminatori e pregiudizi ingiusti vietati dal diritto dell'Unione Europea. Sarà garantito un accesso ai dati di alta qualità e rappresentatività. Per disporre di set di dati di addestramento, convalida e prova di elevata qualità verranno attuate adeguate pratiche di *governance*, integrità e gestione dei dati.

Nell'approccio ai sistemi di IA a rischio limitato, i set di dati dovranno essere sufficientemente pertinenti, rappresentativi, adeguatamente verificati in termini di errori e il più possibile completi alla luce della finalità prevista del sistema.

Dovranno inoltre possedere le proprietà statistiche adeguate, anche per quanto concerne persone o gruppi di persone nei cui confronti il sistema è destinato a essere usato, prestando particolare attenzione all'attenuazione di possibili distorsioni nei set di dati, che potrebbero comportare lesioni ai diritti fondamentali o esiti discriminatori per le persone interessate dal sistema stesso.

I sistemi di IA ad alto rischio dovranno essere formati e testati con un set di dati pertinenti, rappresentativi, verificati, inclusivi e completi per ridurre al minimo il rischio di pregiudizi e garantire che questi possano essere affrontati attraverso individuazione, correzione e altre misure di mitigazione.

Le misure riguarderanno in particolare un esame atto a valutare le possibili distorsioni che possono incidere sulla salute e sulla sicurezza delle persone, avere un impatto negativo sui diritti fondamentali o comportare discriminazioni

vietate dal diritto dell'Unione, in particolare quando i dati di output influenzano gli input per operazioni future (circuiti di *feedback*, *feedback loops*), nonché misure adeguate per individuare, prevenire e attenuare le possibili distorsioni.

**Con le prefate strategia di mitigazione delle distorsioni e delle discriminazioni, ci si avvia pertanto verso l'imparzialità e la trasparenza dell'intelligenza artificiale.**

Naturalmente, è evidente che l'applicazione di queste norme e delle misure ivi codificate va vista in proiezione futura.

Allo stato attuale, ci si affida alla normativa vigente inerente agli obblighi di trasparenza ed alle facoltà di rinuncia al trattamento svolto da questi sistemi, nonché al diritto a richiedere l'intervento umano dietro al trattamento di questi dati.

Tali disposizioni sono costituite dall'art. 22 del Regolamento 2016/679 UE (c.d. GDPR) del Parlamento europeo e del Consiglio, del 27 aprile 2016, di disciplina generale sulla protezione dei dati, e da quelle in tema di trasparenza di cui dal Decreto legislativo n. 104/2022 (c.d. Decreto Trasparenza), che sanciscono precisi limiti ed obblighi ove vengano utilizzati "sistemi decisionali o di monitoraggio automatizzati" nei confronti dei lavoratori.

## 5. Conclusioni

In sintesi, l'utilizzo dei sistemi di intelligenza artificiale nei processi umani (selezione del personale, creazione di offerte personalizzate, giurimetria e giustizia predittiva, accesso a determinati servizi, diagnostica medico-sanitaria, e innumerevoli altri) può condurre a miglioramenti significativi dell'efficacia, dell'efficienza e dell'accuratezza.

Tuttavia, è importante apprestare i giusti rimedi alla prevenzione e alla gestione dei rischi di discriminazione e *bias* cognitivi che possono essere associati all'utilizzo dei suindicati algoritmi.

Solo un'equilibrata **sintesi di cultura e diritto** può garantire l'utilizzo più equo e responsabile dell'intelligenza artificiale.

Valori come trasparenza e equità, unitamente ad un impianto normativo quanto più sovranazionale, esteso e condiviso, che imponga obblighi ben specifici in capo a sviluppatori e utilizzatori dei prefati sistemi tecnologici, sembrano essere gli ingredienti essenziali di una ricetta non semplice.

La via verso forme di intelligenza artificiale equa e responsabile, lastricata di questioni etiche, sociali, giuridiche e tecnologiche, ci prospetta infatti una lunga successione di sfide inedite ed articolate.

La meta è un avvenire in cui le fatidiche "*fantasie di possibilità*" di Wells si tramutino in una realtà in cui i nuovi strumenti di IA restino tali e come tali siano mezzi di progresso per (e accessibili a) tutti, ed operino al servizio di diritti, libertà e valori fondamentali della persona e dell'umanità intera, sì da



favorire contesti ove parità, giustizia, *privacy* ed inclusione prevalgano sempre su abusi, discriminazioni, distorsioni.

In definitiva, il fine ultimo è e deve restare quello di prevenire il rischio di un'unica dominante prospettiva sul mondo, favorendo, per converso, una visione, scevra da modelli stereotipati e aberrazioni varie, che rappresenti le culture, le strutture valoriali e le identità tutte che liberamente muovono il mondo<sup>15</sup>.

---

<sup>15</sup> P. Pellegrino, *Miti, Favole, Fiabe, Modelli alternativi di comunicazione giuridica tra prosa e poesia*. Lezioni di narratologia del diritto. Editore ARACNE, Roma, 2013