Intelligenza artificiale, trattamenti discriminatori e danno

Irene Ambrosi[1]

Sommario: *1. L'impatto della intelligenza artificiale sul fenomeno discriminatorio.- 2. La questione della regolazione dei sistemi di intelligenza artificiale in rapporto al fenomeno discriminatorio.- 3. L'accesso alla tutela e l'accertamento del fenomeno discriminatorio derivante da sistemi di IA.-4. Casistica. -5. La necessità della supervisione umana dei sistemi di IA.*

1. L'impatto della intelligenza artificiale sul fenomeno discriminatorio

L'impatto della AI sul fenomeno discriminatorio è descritto come "*realtà on life"* proprio per sottolineare come la realtà virtuale si ponga, ormai, in modo parallelo rispetto a quella reale e sia destinata a mutare anche la fenomenologia della discriminazione[2].

Per effetto della tecnologia, il fenomeno della discriminazione si è arricchito di un relazione inedita, non più mediata dalla persona umana, ma direttamente influenzata e provocata da un ente non umano: la macchina che, da mezzo, addirittura, può assumere ruolo di soggetto agente, recidendo la relazione causale tra la condotta umana e l'evento, con il rischio di sacrificare nel nome della tecnologia i diritti di autodeterminazione, la dignità delle persone [3] e primo, fra tutti, il principio di eguaglianza e di non discriminazione.

Il fattore primario algoritmico con cui la macchina viene predisposta al funzionamento è concepito proprio per distinguere tra caso e caso ed è evidente che non ogni distinzione discrimina, ma che nel procedimento utilizzato dalla macchina si può celare una distinzione irragionevole.

Il procedimento di ricostruzione delle sequenze funzionali del *come* la macchina sia giunta ad un risultato problematico è assai complesso e due fattori rendono difficile la ricostruzione: il primo attiene al difetto di conoscenza del processo di funzionamento interno, il secondo alla difficoltà di individuare la causa che ha determinato la differenziazione e

---

[1] Rielaborazione della relazione tenuta al Corso di formazione dal titolo "*Come cambia la responsabilità civile al tempo dell'intelligenza artificiale*", svoltosi presso la Scuola Superiore della Magistratura, in Roma, Fontana di Trevi, 2-4 marzo 2026, nell'ambito del ciclo di seminari dedicati a "*Le sfide dell'intelligenza artificiale al diritto, tra neopositivistica fiducia nella tecnica e rimeditazione delle categorie giuridiche*".

[2] L. Floridi, *La quarta rivoluzione, come l'infosfera sta trasformando il mondo*, Milano, 2017.

[3] T. Groppi, *Alle frontiere dello stato costituzionale: innovazione tecnologica e intelligenza artificiale*, in Giurecost., 2020, 675.

tale possibilità di discriminazione viene descritta come basata o derivata dalla intelligenza artificiale (*AI based discrimination*) [4].

La descrizione e la scansione temporale dei passaggi che presiedono alla programmazione delle tecniche di IA è particolarmente importante[5].

L'analisi del funzionamento dei sistemi di IA può essere compiuta attraverso l'esame delle varie fasi in cui il procedimento di differenziazione si scandisce e che, schematicamente, possono così riassumersi: relazione che corre tra la *caratteristica* che il sistema ricerca e la *categoria* che le viene associata; la *raccolta* e la *selezione dei dati;* la selezione della *caratteristica inserita nel modello di ricerca*; la scelta del *"proxy"* ovvero *dell'elemento sul quale si incardina il processo di decisione della macchina*. Il proxy è l'elemento che la macchina ricerca e sul quale agisce per compiere la distinzione e per raggiungere il risultato desiderato[6].

Pertanto, se guardiamo al risultato del sistema di IA, esso può essere costruito in funzione sia della *utilità* sia della *necessità* (quale strumento predittivo: di sorveglianza, polizia e giustizia), al fine di evidenziare caratteristiche individuali e fattori sociali e proprio tale evidenziazione può determinare possibili criticità differenziatrici in una prospettiva antidiscriminatoria.

Il sistema potrebbe essere integrato da meccanismi di differenziazione che sviluppino principi di eticità e di empatia affinché la nuova complessità sociotecnica e gli elementi fondanti il processo decisionale consentano – e questo è tema essenziale - il mantenimento di un orizzonte tecnologico "a misura d'uomo" [7].

In questo specifico ambito, l'UE impone una serie di cautele etiche[8] per assicurare che la costruzione e l'utilizzo dei *robot* sia idonea a

---

[4] C. Nardocci, *Algoritmi, Eguaglianza, discriminazione, le sfide dell'intelligenza artificiale,* Torino, 2025, 6.

[5] Così le classifica la letteratura specialistica: 1) la individuazione e selezione delle "target variables" 2) e delle "class labels" (la TV è la caratteristica che il sistema ricerca, e la C.L è la categoria che le viene associata); 3) la raccolta e la selezione dei dati (data training); 4) la selezione della caratteristica inserita nel modello ("feature selection"); 5) la scelta del "proxy" (dell'elemento sul quale si incardina il processo di decisione della macchina).

[6] P. Zuddas, *, Intelligenza artificiale e discriminazioni, in AA.VV., Liber amicorum per Pasquale Costanzo,* in Consulta Online, 2020

[7] A. Celotto, *Come regolare gli algoritmi. Il difficile bilanciamento fra scienza, etica e diritto*, Analisi Giuridica dell'Economia, 2019, 57.

[8] Reg. 2024/UE. cfr. considerando (44) "Sussistono serie preoccupazioni in merito alla base scientifica dei sistemi di IA volti a identificare o inferire emozioni, in particolare perché l'espressione delle emozioni varia notevolmente in base alle culture e alle situazioni e persino in relazione a una stessa persona. Tra le principali carenze di tali sistemi figurano la limitata affidabilità, la mancanza di specificità e la limitata generalizzabilità. Pertanto, i sistemi di IA che identificano o inferiscono emozioni o intenzioni di persone fisiche sulla base dei loro dati biometrici possono portare a risultati discriminatori e possono essere invasivi dei diritti e delle libertà delle persone interessate. Considerando lo squilibrio di potere nel contesto del lavoro o dell'istruzione, combinato con la natura

preservare la dignità, l'autodeterminazione, la tutela della *privacy* delle persone.

Tali sistemi sono vietati nel contesto lavorativo e della istruzione stante lo squilibrio di potere esistente in quei contesti, nonché nella gestione dei gruppi vulnerabili (bambini, anziani e disabili), potendosi determinare un attaccamento emotivo tra l'uomo e il robot con impatti emotivi e fisici non prevedibili e rischiosi, a prescindere se gli strumenti siano o meno ad Alto rischio [9].

La questione delle questioni è, dunque, quella di fissare le regole con cui programmare gli algoritmi e con cui orientare le forme di intelligenza artificiale, con le quali guidare i sillogismi, posto che il processo decisionale di un algoritmo è squisitamente statistico in quanto la macchina è capace di selezionare una scelta solo in base ai dati che il programmatore umano le ha fornito[10].

Ai fini dei principi di eguaglianza e non discriminazione la classificazione che più interessa dei sistemi di IA è quella riferita alla bipartizione tra sistema simbolico e *sub* simbolico.

Con il primo tipo di IA, si intende quella che agisce in perfetta corrispondenza tra i dati immessi in fase di programmazione e dati richiesti in uscita.

Con il secondo tipo, si intende, viceversa, un sistema di IA che si allontana dalla programmazione iniziale, con due effetti in concreto molto problematici ai fini delle discriminazioni che *in tesi* possono conseguire al loro funzionamento (immissione dei dati e aggiornamenti) e alle

---

invasiva di tali sistemi, questi ultimi potrebbero determinare un trattamento pregiudizievole o sfavorevole di talune persone fisiche o di interi gruppi di persone fisiche. È pertanto opportuno vietare l'immissione sul mercato, la messa in servizio o l'uso di sistemi di IA destinati a essere utilizzati per rilevare lo stato emotivo delle persone in situazioni relative al luogo di lavoro e all'istruzione. Tale divieto non dovrebbe riguardare i sistemi di IA immessi sul mercato esclusivamente per motivi medici o di sicurezza, come i sistemi destinati all'uso terapeutico.

[9] Reg. 2024/UE. Cfr considerando (132) "Alcuni sistemi di IA destinati all'interazione con persone fisiche o alla generazione di contenuti possono comportare rischi specifici di impersonificazione o inganno, a prescindere dal fatto che siano considerati ad alto rischio o no. L'uso di tali sistemi dovrebbe pertanto essere, in determinate circostanze, soggetto a specifici obblighi di trasparenza, fatti salvi i requisiti e gli obblighi per i sistemi di IA ad alto rischio, e soggetto ad eccezioni mirate per tenere conto delle particolari esigenze delle attività di contrasto. Le persone fisiche dovrebbero in particolare ricevere una notifica nel momento in cui interagiscono con un sistema di IA, a meno che tale interazione non risulti evidente dal punto di vista di una persona fisica ragionevolmente informata, attenta e avveduta, tenendo conto delle circostanze e del contesto di utilizzo. Nell'attuare tale obbligo, le caratteristiche delle persone fisiche appartenenti a gruppi vulnerabili a causa della loro età o disabilità dovrebbero essere prese in considerazione nella misura in cui il sistema di IA sia destinato a interagire anche con tali gruppi. È inoltre opportuno che le persone fisiche ricevano una notifica quando sono esposte a sistemi di IA che, nel trattamento dei loro dati biometrici, possono identificare o inferire le emozioni o intenzioni di tali persone o assegnarle a categorie specifiche. Tali categorie specifiche possono riguardare aspetti quali il sesso, l'età, il colore dei capelli, il colore degli occhi, i tatuaggi, i tratti personali, l'origine etnica, le preferenze e gli interessi personali. Tali informazioni e notifiche dovrebbero essere fornite in formati accessibili alle persone con disabilità".

[10] A. Celotto, *ibidem*, 49.

conseguenti responsabilità: accade che la macchina (*machine learning, deep learning*) perviene ad un risultato, senza che il programmatore lo abbia previsto (*Black box*).

È interessante rilevare, altresì, come posto in luce dalla più attenta dottrina, che in ordine agli effetti della differenziazione, le stesse vittime colpite dalla discriminazione derivata dalla IA smarriscono la consapevolezza di essere state discriminate o il senso di aver subito un trattamento irragionevolmente differenziato, presupposto del diritto di accesso agli strumenti di tutela non discriminatoria.

In tale ottica, un primo aspetto da evidenziare è quello legato alle difficoltà conoscitive in cui può trovarsi l'interprete sia in ordine alla conoscenza soggettiva (sono davvero una vittima? Sono parte di un certo gruppo che viene discriminato?) sia oggettiva (il trattamento subito è effettivamente discriminatorio?) essendo difficile conoscere nel dettaglio un sistema di IA (caratterizzato da opacità, "*non luogo"* algoritmico [11]) governato dalla logica della concorrenza e del mercato e non da quella della solidarietà e dell'eguaglianza.

L'assenza del legame, in termini di capacità di azione, tra condotta umana e discriminazione, identifica, inoltre, una caratteristica aggiuntiva della tipologia discriminatoria in esame, cioè il suo essere non più intuitiva e, in ultima analisi, prevedibile, dal legislatore, e sanzionabile, dal giudice[12].

Giova richiamare alcuni casi emblematici che hanno evidenziato la necessità di normare l'algoritmo al fine di garantirne il carattere oggettivo, trasparente, conoscibile e non discriminatorio.

Il primo, deciso dalla Corte suprema dello Stato del Wisconsin, State v. Eric L. Loomis, in data 13 luglio 2016; un cittadino di quello Stato, Eric Loomis, afroamericano, è stato condannato a sei anni di carcere, per aver partecipato ad una sparatoria, anche sulla base del punteggio "alto" assegnato da COMPAS (sistema di calcolo del rischio di recidiva sulla base di una serie di dati forniti al sistema) progettato dalla società Northponte Inc.

La difesa ha sostenuto che l'uso di un algoritmo proprietario (oggetto di segreto industriale) viola il diritto all'equo processo, poiché né l'imputato né il giudice potevano verificare il funzionamento del *software*.

La *ONG ProPublica* ha rivelato gli effetti discriminatori dell'algoritmo utilizzato da COMPAS che prevedeva una probabilità di recidiva della popolazione nera doppia rispetto a quella della popolazione

[11] Definisce "lo spazio telematico come un non - luogo", N. IRTI, *Norma e luoghi. Problemi di geo-diritto*, Roma-Bari, 2001, 65.
[12] C. Nardocci, *Intelligenza artificiale e discriminazioni*, Riv. Gruppo di Pisa 2021, 32.

bianca nei due anni successivi alla condanna, mentre considerava il rischio di recidiva tra i bianchi molto meno probabile[13].

Va notato che la Corte Suprema dello Stato di Wisconsin negò accesso collettivo al codice sorgente, ma, riconoscendone i rischi, relegò l'uso dell'algoritmo come supporto tecnico che non avrebbe mutato la decisione giudiziaria e non, come fattore esclusivo della decisione dell'algoritmo, che fu escluso dalla Corte[14].

Al primo caso può essere affiancato quello deciso dalla Corte Suprema canadese, nel giudizio Ewert vs Canada, 13 giugno 2018, ove la discriminazione algoritmica riguardava un uomo, detenuto in un centro di detenzione federale ed appartenente alla comunità indigena dei Métis, che lamentava il carattere discriminatorio degli strumenti informatici utilizzati dal Correctional Service of Canada (CSC) allo scopo di stabilire il livello di pericolosità sociale dei detenuti[15].

Il terzo, deciso dalla Corte costituzionale tedesca in data 16 febbraio 2023, la quale ha dichiarato incostituzionale l'utilizzo di programmi predittivi di polizia (nella specie, Palantir) negli Stati di Amburgo e Assia ove la polizia, elaborando i dati disponibili inerenti al passato della persona oggetto di trattamento, calcolavano la probabilità che tale persona si comporti in un certo modo (eventualmente reiterando il reato).

Di sicuro rilievo, l'affermazione della Corte tedesca secondo cui le tecnologie di IA utilizzate per prevenire la commissione di reati si pongono in contrasto con il principio costituzionale di eguaglianza e non discriminazione e con il diritto di autodeterminazione legato all'omesso consenso di coloro che risultino sottoposti a tali strumenti. La Corte afferma che il *principio di limitazione delle finalità* costituisce requisito necessario e sufficiente per legittimare l'ulteriore utilizzo dei dati, se l'autorità autorizzata alla raccolta dei dati li utilizzi all'interno e nell'ambito delle stesse competenze che le sono attribuite e per la prevenzione degli stessi reati.

Interessante è quanto ritenuto dalla Corte tedesca secondo cui, non solo è illegittima la raccolta dei dati e dei rischi nel contesto delle tecnologie di polizia predittiva, ma anche dei rischi di discriminazione associati alle modalità di utilizzo dei dati personali raccolti, specie laddove le autorità ne dispongano in grande quantità rendendo agevoli

---

[13] Cfr., *Carta etica europea sull'utilizzo dell'intelligenza artificiale nei sistemi giudiziari e negli ambiti connessi*, 2018, punto 137.

[14] S. Carrer, *Se l'amicus curiae è un algoritmo: il chiacchierato caso Loomis alla Corte Suprema del Wisconsin*, in Giurisprudenza Penale *Web*, 2019, 1 ss.

[15] L. Giacomelli, *Big brother is «gendering» you. Il diritto antidiscriminatorio alla prova dell'intelligenza artificiale: quale tutela per il corpo digitale?* in BioLaw Journal, 2019, 269 ss.;

correlazioni, che nell'ottica differenziatrice del sistema di IA, divengono anche discriminatorie.

Nel bilanciamento tra esigenza di pubblica sicurezza e principi costituzionali fondamentali della persona, la Corte cost. tedesca ha lanciato un monito al legislatore al fine di regolare l'impiego di sistemi di IA sulla base di linee guida che ne precisino i requisiti e criteri d'uso.

Il quarto caso, infine, seppure non strettamente attinente al problema della discriminazione derivante da algoritmi, ma relativo ad un conflitto di attribuzione tra la competenza statale trasversale in materia di "tutela della concorrenza" e le competenze regionali (in particolare quella residuale sul "trasporto pubblico locale"), è stato affrontato dalla Corte costituzionale italiana con la recente sentenza n. 163 del 2025, ove, decidendo su un conflitto di attribuzione sollevato dalla Regione Calabria, ha ritenuto necessario il ricorso ai principi di proporzionalità e ragionevolezza, ritenendo non spettante allo Stato (Ministero Infrastrutture e Trasporti) imporre l'uso forzato di applicazioni informatiche (nella fattispecie, il foglio di servizio elettronico (FDSE) per il servizio di noleggio con conducente (NCC). Sottolinea la dottrina che potrebbe soccorrere anche la giurisprudenza costituzionale in tema di "automatismi legislativi" ove la disciplina fissi criteri rigidi, senza consentire all'interprete un'applicazione ragionevole[16]

2. *La questione della regolazione dei sistemi di intelligenza artificiale in rapporto al fenomeno discriminatorio*

Nella piena consapevolezza delle possibili e rischiose implicazioni dei sistemi di AI sul versante del principio di eguaglianza e della non discriminazione, un primo problema logico affrontato a livello globale è stato, di sicuro, quello regolatorio, ambito in cui gli Stati della Unione europea hanno atteso le determinazioni decise dalle istituzioni europee.

Le determinazioni vigenti che, qui si richiamano, sono principalmente tre.

Va anzi tutto ricordata la *Carta Etica sull'uso dell'intelligenza artificiale nei sistemi giudiziari,* adottata dal Consiglio di Europa nel dicembre 2018, che dedica l'intero paragrafo 2 al principio di non discriminazione e ai rischi di interferenze tra la costruzione dei modelli di intelligenza artificiale e la tutela della non discriminazione tra individui e gruppi[17].

---

[16] C. Nardocci, *Algoritmi*, *op cit.*, 393, nota 44, ove il riferimento è fatto alla giurisprudenza costituzionale al caso dell'adozione (tra l'altro) in merito alla rigida fissità di una differenza di età tra l'adottante e l'adottato.

[17] *European Ethical Charter on the use of Artificial Intelligence in Judicial Systems and their environment*, 2018.

Al fine di attenuare i rischi ed assicurare che le metodologie non riproducano e non aggravino le discriminazioni esistenti e che non conducano ad analisi o usi deterministici, la Carta dispone nel par. 1 il "*Principio del rispetto dei diritti fondamentali*" ovvero l'esercizio di una particolare vigilanza sia nella fase dell'elaborazione che in quella dell'utilizzo della IA, specialmente quando il trattamento si basa direttamente o indirettamente su dati "sensibili" (l'origine razziale o etnica, le condizioni socio-economiche, le opinioni politiche, la fede religiosa o filosofica, l'appartenenza a un sindacato, i dati genetici, i dati biometrici, i dati sanitari o i dati relativi alla vita sessuale o all'orientamento sessuale), con correttivi volti a neutralizzare tali rischi e sensibilizzare gli enti pubblici e privati. Vanno ricordati anche gli altri tre paragrafi della Carta dedicati al principio di qualità e sicurezza (3 par.), quello di trasparenza, imparzialità , equità (4 par.) e quello del controllo da parte dell'utilizzatore.

In una prospettiva costituzionale l'UE si è posta il problema di evitare un impatto negativo del sistema di IA sui diritti fondamentali protetti dalla Carta e dopo un triennio di lavori (con un *iter* di adozione altalenante proprio per le difficoltà di definizione dell'oggetto da regolare, a partire dalla prima versione adottata nel 2021), in data 13 giugno 2024, è stato approvato il Reg. (UE) 2024/1689 in vigore dal successivo agosto [18].

Nel par. 2 si legge: "Data la capacità di tali metodologie di trattamento di rivelare le discriminazioni esistenti, mediante il raggruppamento o la classificazione di dati relativi a persone o a gruppi di persone, gli attori pubblici e privati devono assicurare che le metodologie non riproducano e non aggravino tali discriminazioni e che non conducano ad analisi o usi deterministici. Deve essere esercitata una particolare vigilanza sia nella fase dell'elaborazione che in quella dell'utilizzo, specialmente quando il trattamento si basa direttamente o indirettamente su dati "sensibili". Essi possono comprendere l'origine razziale o etnica, le condizioni socioeconomiche, le opinioni politiche, la fede religiosa o filosofica, l'appartenenza a un sindacato, i dati genetici, i dati biometrici, i dati sanitari o i dati relativi alla vita sessuale o all'orientamento sessuale. Quando è individuata una di queste discriminazioni, devono essere previste le misure correttive al fine di limitare o, se possibile, neutralizzare tali rischi e sensibilizzare gli attori. Dovrebbe tuttavia essere incoraggiato l'utilizzo dell'apprendimento automatico e delle analisi scientifiche multidisciplinari, al fine di contrastare tali discriminazioni".

[18] A corredo della proposta di regolamento, la Commissione europea aveva proposto due direttive: una prima «sulla responsabilità civile da IA» presentata nel 2022/0303 UE, -volta a consentire a coloro i quali hanno subito danni causati dalla tecnologia di IA di accedere al risarcimento come se avessero subito danni in qualsiasi altra circostanza. Veniva prevista una «presunzione di esistenza del nesso di causalità tra la colpa del convenuto e l'output prodotto da un sistema di IA o la mancata produzione di un output da parte di tale sistema» e l'accesso agli elementi di prova di imprese o fornitori, quando si tratta di IA ad alto rischio -, è stata ritirata nel febbraio 2025 da parte della Commissione europea con il comunicato "nessun accordo prevedibile"; una seconda, presentata nel 2024/2853 UE «Responsabilità per danno da prodotti difettosi», volta a ridisegnare il quadro in materia (abrogando la Direttiva 85/374/Cee), con l'intento di ampliare l'ambito di applicazione alla sfera delle tecnologie digitali (comprende software, Intelligenza Artificiale e prodotti o sistemi basati sull'IA destinati all'attività commerciale) al fine di dettare specifiche regole di protezione per le evenienze che da esse possono essere generate, includendo nella definizione di prodotto: l'elettricità, i file per la fabbricazione digitale e il software (art. 4) e in quella di difetto anche gli effetti conseguenti alla «eventuale capacità di continuare ad imparare o acquisire nuove funzionalità dopo la sua diffusione sul mercato» (art. 6).

Il Regolamento è considerato come la prima regolamentazione dell'IA "a livello globale"[19] nel definire che cosa si intenda per sistema di IA e nello stabilire regole normative comuni sull'intelligenza artificiale in ambito europeo, volte alla protezione della persona al cospetto dell'innovazione tecnologica in una prospettiva etica[20].

La versione vigente, all'art. 3, n. 1, definisce «sistema di IA»: "*un sistema automatizzato progettato per funzionare con livelli di autonomia variabili e che può presentare adattabilità dopo la diffusione e che, per obiettivi espliciti o impliciti, deduce dall'input che riceve come generare output, quali previsioni, contenuti, raccomandazioni o decisioni che possono influenzare ambienti fisici o virtuali*".

Si è scelta una definizione di IA idonea a comprendere in sè i sistemi che possiedono maggiore autonomia rispetto alla componente umana originaria e una capacità di funzionamento – di assumere decisioni e influenzare l'ambiente esterno con tecniche *learning reasoning* o *modeling* –, senza un legame diretto causale immediato con la condotta umana[21].

Condivisibilmente il legislatore europeo mette in guardia dal fatto che l'IA presenta, accanto a molti utilizzi benefici, la possibilità di essere utilizzata impropriamente e di fornire strumenti nuovi e potenti per pratiche di manipolazione, sfruttamento e controllo sociale [22].

A ciò come contraltare sta la necessità della supervisione umana cd. Human oversight art. 14 reg IA Act "La sorveglianza umana mira a prevenire o ridurre al minimo i rischi per la salute, la sicurezza o i diritti fondamentali che possono emergere quando un sistema di IA ad alto rischio è utilizzato conformemente alla sua finalità prevista o in

(da recepire entro il 9 dicembre 2026).

[19] I. Carnat, *Intelligenza artificiale e responsabilità civile*, Enc. del diritto. I Tematici, VII, Volume la *Responsabilità civile* diretto da C. Scognamiglio, Milano, 2024, 655.

[20] Sino alla sua adozione, non si era affrontato il tema della discriminazione derivante da uso di sistemi di IA, si faceva riferimento al Regolamento (Ue) 2016/679 del Parlamento Europeo e del Consiglio del 27 aprile 2016, relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (regolamento generale sulla protezione dei dati GDPR).

[21] C. Nardocci, op. cit., 19.

[22] Nel *preambolo* (28) del Reg. si mette in guardia dal fatto che «l'IA presenta, accanto a molti utilizzi benefici, la possibilità di essere utilizzata impropriamente e di fornire strumenti nuovi e potenti per pratiche di manipolazione, sfruttamento e controllo sociale. Tali pratiche sono particolarmente dannose e abusive e dovrebbero essere vietate poiché sono contrarie ai valori dell'Unione relativi al rispetto della dignità umana, alla libertà, all'uguaglianza, alla democrazia e allo Stato di diritto e ai diritti fondamentali sanciti dalla Carta, compresi il diritto alla non discriminazione, alla protezione dei dati e alla vita privata e i diritti dei minori»; inoltre, si mette in guardia in merito ai sistemi di IA che permettono ad attori pubblici o privati di attribuire « un punteggio sociale alle persone fisiche possono portare a risultati discriminatori e all'esclusione di determinati gruppi. Possono inoltre ledere il diritto alla dignità e alla non discriminazione e i valori di uguaglianza e giustizia. Tali sistemi di IA valutano o classificano le persone fisiche o i gruppi di persone fisiche sulla base di vari punti di dati riguardanti il loro comportamento sociale in molteplici contesti o di caratteristiche personali o della personalità note, inferite o previste nell'arco di determinati periodi di tempo».

condizioni di uso improprio ragionevolmente prevedibile, in particolare qualora tali rischi persistano nonostante l'applicazione di altri requisiti di cui alla presente sezione".

La disciplina dettata dal Regolamento, sebbene enunci un approccio regolatore *forte* al tema della IA, sembrerebbe contraddirne l'enunciato, introducendo, viceversa, con il criterio del rischio, un approccio più debole.

In effetti, per un verso, pone l'obbligo di valutare l'impatto che l'uso di tale sistema può produrre sui diritti fondamentali, per l'altro, non dice "come" adempiere a tale obbligo[23].

Va pure notato che, seppure il termine armonizzazione compaia 56 volte nel testo del Reg. , tuttavia, la disciplina affida la regolazione agli Stati membri e si badi, non soltanto alle Autorità pubbliche statali, ma anche a quelle private, con rischio di una frammentazione delle regole e una rilevante disomogeneità di regole nel contesto unionale. Ciò si desume dall'esame dell'art. 27, norma rubricata "*Valutazione d'impatto sui diritti fondamentali per i sistemi di IA ad alto rischio*"[24].

La scelta effettuata dal Regolamento è quella della dicotomia tra (misurabilità del) rischio vs (non quantificabilità della) violazione del diritto[25].

Si identifica il grado o livello di lesività, quantificando in termini statistici la probabilità di violazione di un diritto fondamentale.

---

[23] Nel considerando (48) del Reg. si individua la misura dell'impatto negativo di un sistema di IA sui diritti fondamentali protetti dalla Carta classificandolo "ad alto rischio". Tali diritti comprendono il diritto alla dignità umana, il rispetto della vita privata e della vita familiare, la protezione dei dati personali, la libertà di espressione e di informazione, la libertà di riunione e di associazione e il diritto alla non discriminazione, il diritto all'istruzione, la protezione dei consumatori, i diritti dei lavoratori, i diritti delle persone con disabilità, l'uguaglianza di genere, i diritti di proprietà intellettuale, il diritto a un ricorso effettivo e a un giudice imparziale, i diritti della difesa e la presunzione di innocenza e il diritto a una buona amministrazione (…). Nel considerando (57) del Reg. si mette in guardia sul pericolo che «Se progettati e utilizzati in modo inadeguato, tali sistemi possono essere particolarmente intrusivi e violare il diritto all'istruzione e alla formazione, nonché il diritto alla non discriminazione, e perpetuare modelli storici di discriminazione, ad esempio nei confronti delle donne, di talune fasce di età, delle persone con disabilità o delle persone aventi determinate origini razziali o etniche o un determinato orientamento sessuale».

[24] Nel considerando (27) del Reg. si richiama uno dei sette principi etici (elaborati del 2019 dall'AI High level Expert Group HLEG, indipendente, nominato dalla Commissione) posto a presidio della «diversità, non discriminazione ed equità» per un'IA affidabile «si intende che i sistemi di IA sono 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 o nazionale».

[25] E si legge nel preambolo, par. 26, del Reg. "Al fine di introdurre un insieme proporzionato ed efficace di regole vincolanti per i sistemi di IA è opportuno avvalersi di un approccio basato sul rischio definito in modo chiaro. Tale approccio dovrebbe adattare la tipologia e il contenuto di dette regole all'intensità e alla portata dei rischi che possono essere generati dai sistemi di IA. È pertanto necessario vietare determinate pratiche di IA inaccettabili, stabilire requisiti per i sistemi di IA ad alto rischio e obblighi per gli operatori pertinenti, nonché obblighi di trasparenza per determinati sistemi di IA".

La dottrina ha posto in luce la *contradictio in terminis* che la relazione tra rischio e violazione di un diritto fondamentale può innescare; come può essere coordinato il concetto di "alto rischio" quale soglia di verifica dei sistemi di IA in ambito europeo con i principi di eguaglianza e di solidarietà posti dalla Carta costituzionale in un ambito come quello dei sistemi di intelligenza artificiale ove non è prevedibile se il risultato dello sviluppo del meccanismo tecnologico sia compiuto nel pieno rispetto della *dignità* e della *personalità* di ogni individuo - inteso sia come persona fisica sia come appartenente ad un gruppo.

Una scelta condivisibile è quella secondo cui alla costruzione funzionamento e vigilanza sui sistemi di IA debbano partecipare studiosi delle più diverse estrazioni scientifiche.

Con la legge 23 settembre 2025 n. 135 (entrata in vigore a ottobre 2025) l'Italia si è munita di una disciplina generale sull'IA che, in conformità con il Reg europeo 1689/2024, stabilisce *disposizioni e deleghe* in materia di IA in vari settori (informazione e riservatezza dei dati personali, sviluppo economico, sicurezza e difesa nazionale, ambito sanitario e disabilità, lavoro, professioni intellettuali, pubblica amministrazione, attività giudiziaria, inclusi quello lavorativo e di protezione dei minori, vietando l'uso di sistemi IA che violino la dignità umana e il principio di non discriminazione, etc.).

In ambito convenzionale, va ricordato il Trattato quadro del Consiglio d'Europa sull'intelligenza artificiale adottato in data 17 maggio 2024 che costituisce il primo trattato internazionale giuridicamente vincolante volto a garantire che l'IA rispetti i diritti umani, la democrazia e lo Stato di diritto. Basata su un approccio antropocentrico, impone obblighi per l'intero ciclo di vita dell'IA, concentrandosi su trasparenza, non discriminazione e protezione dei dati.

Si segnalano l'art. 7 "*Dignità umana* e *Autonomia individuale"* (nozione aperta) tanto più importante l'autonomia individuale al cospetto di sistemi di IA dotati di abilità di imitazione e manipolazione, l'art. 8 *Controllo umano sul funzionamento delle IA e* l'*a*rt. 10 *Eguaglianza e non discriminazione.*

3. L'accesso alla tutela e l'accertamento del fenomeno discriminatorio derivante da sistemi di IA

I casi di violazione dei diritti fondamentali in questo ambito portati all'attenzione dei giudici e quelli in cui è stata accertata una violazione sono pochi sia a livello internazionale che nazionale.

Una giurisprudenza così scarna dipende, per un verso, dalla difficoltà di raccordo tra le norme di diritto positivo in materia di diritto discriminatorio e i principi costituzionali e sovranazionali che

riconoscono e garantiscono il diritto dell'individuo all'accesso ad un giudice e, per l'altro, dal vuoto normativo sul carattere discriminatorio legato all'azione della macchina ai danni del singolo o dei gruppi sociali.

Non si discute più del fatto che il sistema di IA discrimina, o meglio, se può discriminare, bensì il *come* lo fa, ai danni di *chi,* in base a *che cosa*.

Da una parte, le difficoltà insorgono anche per le nuove modalità attraverso cui la discriminazione da IA si manifesta e che vanno dalla «molestia, al trattamento discriminatorio con un *comparator* soltanto ipotetico[26], tipologie di discriminazione che sfociano in violenza, l'ordine di discriminazione»[27].

D'altra parte, ulteriore fattore di crisi delle categorie tradizionali di discriminazione si annida nelle difficoltà di accertamento effettivo della discriminazione subita e della sanzionabilità, poi, della condotta in quanto discriminatoria, difficoltà riscontrabile tradizionalmente anche nel diritto discriminatorio classico e che si acuisce al cospetto di quelle derivanti dall'impiego delle tecnologie di intelligenza artificiale.

Sempre in tema di accesso alla tutela giudiziaria, la discriminazione derivante dal IA acuisce e accentua discriminazioni già esistenti e contribuisce a formare nuovi gruppi che possono anche non essere consapevoli di essere stati destinatari di una discriminazione di gruppo.

Tre criteri si mostrano utili per *accertare il funzionamento discriminatorio* della tecnologia della intelligenza artificiale: diritto alla conoscenza, identificazione del soggetto attivo del trattamento se il sistema sia utilizzato dallo Stato e l'eventuale responsabilità dello Stato nell'impiego della IA.

Il primo criterio *c.d. "Right to Know"* può aiutare ad accertare un eventuale *agere* discriminatorio dell'algoritmo e fa da corollario al *principio di trasparenza.* Consente non soltanto di conoscere il funzionamento della macchina, ma anche se si stia utilizzando una intelligenza artificiale. Questo ultimo elemento di conoscenza è particolarmente importante allorquando sia le autorità pubbliche ad utilizzare le tecnologie di IA: si segnala, ad esempio, l'esperienza delle città di Amsterdam e di Helsinki che hanno reso pubblico il registro degli strumenti di intelligenza artificiale impiegati ove sono elencate le

[26] Se non esiste una persona reale che ha ricevuto un trattamento migliore, è possibile utilizzare un "comparator ipotetico" per mostrare come una persona senza quella caratteristica sarebbe stata trattata in situazioni analoghe. Il *comparator* è la persona o il gruppo che non condivide la caratteristica protetta (es. razza, genere, età, disabilità, orientamento sessuale) e che viene trattato più favorevolmente in una situazione comparabile. Ad esempio, una donna paga un abbonamento in palestra più di un uomo (comparator) per gli stessi servizi (discriminazione diretta per genere); un lavoratore con disabilità riceve meno formazione rispetto a un collega senza disabilità in una posizione simile (comparator).

[27] C. Nardocci, *Algoritmi*, *op. cit*., 368.

applicazioni, i dati utilizzati, la loro logica operativa e il livello di supervisione umana[28] .

Il secondo ricerca *il nesso tra ricerca dei dati e funzionamento del sistema di IA.*

Il terzo e ultimo attiene al suo *impiego da parte del potere pubblico o privato*.

Necessità e obbligatorietà delle valutazioni di impatto (già previste dall'AI ACT artt. 9 e 27 e 77) che possono assurgere a strumenti non vincolante per l'interprete (Avvocato, Giudice) al fine di isolare il trattamento discriminatorio e accertarlo.

L'utilizzo del sistema di IA diviene presupposto dell'accertamento di una responsabilità da parte dell'utilizzatore (ma anche del programmatore che ha costruito il sistema; o di chi ne ha consentito la diffusione nel mercato, oltre a chi lo ha effettivamente impiegato nell'esercizio delle proprie funzioni).

La dottrina si è chiesta se si possono utilizzare per la IA, le tradizionali categorie della discriminazione diretta e indiretta.

Nel diritto discriminatorio tradizionale[29], si intende per discriminazione diretta, la situazione nella quale una persona è trattata, in base ad un fattore di discriminazione classico ai sensi dell'art. 3 della Cost. (genere, razza, opinione politica etc.) o per caratteristiche individuali individuate dall' art. 21 della Carta dei diritti fondamentali dell'Unione europea (orientamento sessuale, identità sessuale, disabilità), meno favorevolmente di quanto sia, sia stata o sarebbe trattata un'altra in una situazione analoga (CGUE C.-177/1988, 8 novembre 1990, rifiuto di assunzione per motivo di gravidanza può opporsi solo alle donne).

Si intende per discriminazione indiretta, la situazione nella quale una disposizione, un criterio o una prassi apparentemente neutri possono mettere in una situazione di particolare svantaggio le persone di un determinato sesso rispetto a persone dell'altro, a meno che tale disposizione, criterio o prassi siano oggettivamente giustificati da una finalità legittima e i mezzi impiegati per il conseguimento della stessa finalità siano appropriati e necessari (CGUE C-38/2024 11 settembre

[28] C. Nardocci, *op. cit.* nota 38 pag. 389.

[29] In via generale, la Corte EDU ha da tempo affermato che «una differenza di trattamento può consistere nell'effetto sproporzionatamente pregiudizievole di una politica o di una misura generale che, *se pur formulata in termini neutri*, produce una discriminazione nei confronti di un determinato gruppo» (Cedu, 9 giugno 2009, Opuz c. Turchia (n. 33401/02), punto 183. Cedu, 20 giugno 2006, Zarb Adami c. Malta (n. 17209/02), punto 80). Nella stessa prospettiva, la Corte di cassazione ha chiarito che la discriminazione diretta è la condotta, il comportamento tenuto, che determina la disparità di trattamento, mentre in quella indiretta, la disparità vietata è l'effetto di un atto, di un patto di una disposizione di una prassi in sé legittima; di un comportamento che è corretto in astratto e che, in quanto destinato a produrre i suoi effetti nei confronti di un soggetto con particolari caratteristiche, che costituiscono il fattore di rischio della discriminazione, determina invece una situazione di disparità che l'ordinamento sanziona (tra tante, Cass. Sez. L, 25/07/2019, n. 20204).

2025 ha esteso "per associazione" ai lavoratori *caregiver* le previsioni della Direttiva 2000/78/CE relative alle discriminazioni indirette sul luogo di lavoro e agli accomodamenti ragionevoli[30]).

Nel diritto antidiscriminatorio, il divieto di fare espresso riferimento ai fattori di discriminazione tradizionali ha prodotto così un mutamento nella selezione delle qualità individuali e nelle modalità entro cui si definisce l'appartenenza identitaria del singolo al gruppo. Essa non si fonda più, almeno esplicitamente, su elementi come la razza, l'etnia, il sesso oppure il genere, bensì su criteri che vanno da fattori quantificabili e misurabili, "il quartiere di residenza", "il codice postale", "le preferenze di siti internet, film, serie televisive", ad altri dai contorni discrezionali, come "l'affidabilità del debitore", il "buon lavoratore", "l'automobilista a basso rischio di sinistri stradali".

La discriminazione algoritmica non si sottrae a tale meccanismo di torsione della fenomenologia discriminatoria, ed è in grado di creare minoranze "inedite" che poggiano su criteri di appartenenza individuale e collettiva "diversi" finendo per costruire nuove forme identitarie. Allo stesso tempo, se è vero che la discriminazione algoritmica ha portato ad una ridefinizione del concetto di affiliazione individuale, di gruppo, di identità, non è altrettanto vero che le sue vittime siano nuove; o meglio, lo siano solo in parte.

Il criterio della discriminazione diretta risulta non bene attagliarsi ai sistemi di IA. Difatti, a meno che non ricorra una ipotesi di c.d. *masking*, ovvero di una discriminazione intenzionale, cioè consapevolmente posta in essere dal programmatore e che discenda da una volontaria e parziale selezione di dati idonea a riverberarsi negativamente ai danni di una o più classi protette. Come esempio, può farsi quello del fenomeno del *redlining*, forma specifica di *proxy discrimination,* che veniva impiegata dagli istituti finanziari statunitensi per evitare di dover servire zone ad alta densità abitativa di afro-americani; non potendo assumere la razza ad elemento in base al quale discriminare, gli istituti finanziari utilizzavano come *proxy* le aree geografiche, discriminando la minoranza afro-americana, pur non facendo esplicitamente riferimento al fattore di discriminazione vietato.

Di regola, a meno che non si tratti di discriminazione intenzionale, il criterio di discriminazione diretta (treatment) non rileva per il sistema di IA, atteso che nel combinarsi della *caratteristica* che il sistema ricerca e la *categoria* che le viene associata, non viene mai inserito un fattore discriminatorio (razza, sesso etc.) e la scelta del *"proxy"* cade su un elemento sul quale si incardina il processo di decisione della macchina.

[30] C. Berti I, *La Corte UE sui «caregiver»: tutela antidiscriminatoria indiretta e diritto agli accomodamenti ragionevoli*, in Rivista Labor, 2025.

Ciò rende difficoltoso, per la parte che invoca tutela, dimostrare il trattamento irragionevolmente differenziato subito, ma anche per lo stesso giudice accertare il carattere discriminatorio.

Pure il criterio della discriminazione indiretta derivante dalla IA (impact) stride con i criteri classici; la peculiarità del criterio classico risiede, difatti, in un criterio *neutro* o apparentemente tale, dal quale poi si genera il trattamento discriminatorio; nel sistema di IA il criterio non è affatto neutro, tenuto conto che è congegnato per fare perno su qualità individuali altamente predittive dell'appartenenza a categorie protette. Ad esempio, può portarsi l'esempio di un algoritmo impiegato per selezionare possibili candidati a ricoprire una posizione lavorativa per la quale l'altezza è requisito necessario e, tuttavia, la macchina non dispone di dati sull'altezza dei candidati e delle candidate, in ragione della appurata correlazione tra altezza e sesso (altrimenti ci troveremmo davanti ad un caso di discriminazione diretta). Neppure il sesso potrà essere impiegato quale *proxy* e non verrà inserito nel *data-set* trattandosi di fattore di discriminazione vietato oltre che dato sensibile. Potrebbe allora accadere che la macchina, al fine di raggiungere il proprio obiettivo (la selezione del/la candidato/a migliore), utilizzi un proxy diverso, come le preferenze in tema di serie tv. In questo caso, il ricorso ad un elemento apparentemente neutrale e non direttamente predittivo del fattore di discriminazione vietato (il sesso) può tuttavia produrre un effetto proporzionalmente deteriore ai danni delle donne, che saranno escluse dalla selezione[31].

Nei sistemi di intelligenza artificiale progressivamente più autonomi, la *proxy discrimination,* tipologia particolarmente insidiosa di discriminazione, perché ricorre in tutte le ipotesi in cui la categoria che si vuole colpire non è immediatamente individuabile perché "coperta" da un elemento di affiliazione individuale diverso e fuorviante e rappresenta la principale sfida al diritto anti discriminatorio tradizionale che, viceversa, mira a prevenire e contrastare disparità di trattamento fondate su qualità individuali immediatamente predittive dell'appartenenza a gruppi protetti.

È stato posto in luce sia dalla dottrina che dalla giurisprudenza che il criterio di discriminazione indiretta, depurato dall'elemento neutro (che non c'è nei sistemi di IA) potrebbe essere recuperato e valorizzato in funzione sull'effetto che la macchina *learning* determina, ciò potrebbe rendere meno difficoltoso l'accertamento del fenomeno discriminatorio, prescindendo dalla condotta discriminatoria.

---

[31] L'esempio è citato in C. Nardocci, *Intelligenza artificiale e discriminazioni*, op cit. pag. 22, nota 81, ed è tratto da A.E.R. Prince, D. Schwarcz, *Proxy discrimination in the age of artificial intelligence and big data*, ivi citato, 1280.

La discriminazione "artificiale" può, ancora, atteggiarsi quale prodotto di condotte plurali che si sommano e si intrecciano a formare una spirale in continuo movimento, con una discriminazione risultante all'esito dal funzionamento della macchina che costituisce il prodotto dell'azione correlata di più *proxy discriminations* che agiscono simultaneamente. In tale caso viene detta anche *intersezionale* ove l'algoritmo combina fattori come genere, cittadinanza e vulnerabilità [32].

Criticità si ravvisano anche in ordine alle regole processuali classiche in tema di discriminazione (inversione dell'onere della prova e utilizzazione a fini probatori di evidenze statistiche) che potrebbero rivelarsi di poco ausilio per i sistemi di IA (per l'impossibilità e difficoltà di leggere il meccanismo discriminatorio operato dal sistema sia per l'agente che per la vittima).

In dottrina si propone di tipizzare la discriminazione che deriva dall'intelligenza artificiale, così da isolarne con precisione i caratteri, concentrando l'attenzione sull'effetto discriminatorio (come accade per la discriminazione indiretta) e sulla capacità predittiva del proxy istituendo correlazioni con caratteristiche protette e a tal fine soccorre l'art. 77 del Reg. con il richiamo alle Autorità nazionali chiamate ad assolvere funzione di supervisione e cooperazione nell'attuazione delle regole unionali, invitando gli Stati membri a predisporre una lista di autorità al fine della tutela dei diritti fondamentali "compreso il diritto alla non discriminazione" [33]. Le Autorità competenti italiane (Art. 77 IA Act) designate sono: l'Agenzia per la Cybersicurezza Nazionale (ACN) e l'Agenzia per l'Italia Digitale (AgID); la prima vigila sulla sicurezza dei sistemi di IA, mentre l'AgID gestisce le notifiche, operando in coordinamento con altre autorità settoriali.

4. Casistica

Tra la esigua casistica interna, molto rilevante è il caso "Deliveroo" deciso dal Tribunale di Bologna che riguardava il meccanismo automatico di accesso, prenotazione e cancellazione delle sessioni di lavoro da parte dei *riders* impiegati dalla società.

---

[32] C. Nardocci, *Algoritmi, op.* cit, *123 e ss.*

[33] Art. 77 (Poteri delle autorità che tutelano i diritti fondamentali) 1. Le autorità o gli organismi pubblici nazionali che controllano o fanno rispettare gli obblighi previsti dal diritto dell'Unione a tutela dei diritti fondamentali, compreso il diritto alla non discriminazione, in relazione all'uso dei sistemi di IA ad alto rischio di cui all'allegato III hanno il potere di richiedere qualsiasi documentazione creata o mantenuta a norma del presente regolamento o di accedervi, in una lingua e un formato accessibili, quando l'accesso a tale documentazione è necessario per l'efficace adempimento dei loro mandati entro i limiti della loro giurisdizione. L'autorità pubblica o l'organismo pubblico pertinente informa l'autorità di vigilanza del mercato dello Stato membro interessato di qualsiasi richiesta in tal senso.

Il *software* (*machine non learning*) non consentiva, tra l'altro, la cancellazione della prenotazione indipendentemente dalla ragione addotta dal lavoratore, che si vedeva pregiudicata la scelta del turno successivo anche qualora quest'ultima fosse stata motivata da ragioni connesse all'esercizio di diritti di rilievo costituzionale, come, in quel caso, il diritto di sciopero[34].

Il primo problema posto dalle associazioni sindacali ricorrenti aveva ad oggetto il sistema di IA e il funzionamento discriminatorio dell'accesso tramite la piattaforma digitale[35] utilizzata dalla società datrice di lavoro.

Il secondo era quello della responsabilità della società convenuta per la discriminazione derivante dall'impiego di tale sistema di IA.

Il Tribunale di Bologna ha accolto la domanda ravvisando una discriminazione indiretta ai sensi dell'art. 2 d. lgs 216 del 2003 (in attuazione direttiva 2000/78/CEE per la parità di trattamento in materia di occupazione e di condizioni di lavoro) con condanna al risarcimento dei danni.

Due importanti precisazioni fissate nella citata ordinanza: 1) la qualificazione della condotta come discriminazione indiretta; 2) l'identificazione del soggetto responsabile con la società Deliveroo cioè

---

[34] Tribunale di Bologna, ordinanza del 31 dicembre 2020; il meccanismo prenotativo, in caso di cancellazione della prenotazione, si era dimostrato ad alto impatto discriminatorio per i riflessi sul "ranking" reputazionale del singolo rider.
Un altro caso da segnalare, sebbene non direttamente incentrato sulla dimensione discriminatoria, è offerto dalla sentenza del Tribunale Amministrativo del Lazio, Sez. III bis n. 3769 del 2017 in relazione all'impiego di un software, incaricato di stabilire la sede di assegnazione di ciascun/a docente. A commento v., I. FORGIONE, *Il caso dell'accesso al software MIUR per l'assegnazione dei docenti* - T.A.R. Lazio Sez. III bis, 14 febbraio 2017, n. 3769, in *Giornale di diritto amministrativo*, 2018, 647 ss.; Consiglio di Stato, Sez. VI, n. 2270 del 2019. Secondo il Giudice amministrativo «è con il software che si concretizza la volontà finale dell'amministrazione procedente che costituisce, modifica o estingue le situazioni giuridiche individuali anche se lo stesso non produce effetti in via diretta all'esterno. Il software finisce per identificarsi e concretizzare lo stesso procedimento». In dottrina si è osservato che tale lettura non tiene, forse, in adeguato conto che l'azione umana si inserisce anche a monte, cioè nella fase di programmazione del programma, cosa che sconsiglierebbe una piana identificabilità tra macchina e utilizzatore finale.

[35] In tema di "sharing economy o economia collaborativa", nel tentativo di risolvere, almeno apparentemente, le incongruenze presenti nelle rigide strutture dell'architettura tradizionale del mercato capitalistico *offline*, proiettandosi verso prospettive maggiormente dinamiche e innovative che sfruttano, da una parte, l'assenza di barriere fisiche e la possibilità di interazione costante tra gli utenti, e, dall'altra, l'interferenza strutturale tra circolazione dei dati e dei beni e, infine, la capacità di intermediazione delle piattaforme stesse. Al riguardo, possono essere formulate due considerazioni: la prima è che il fenomeno ha ottenuto una prima embrionale definizione normativa attraverso il citato d. L n. 101 del 2019 (conv. in l. 128 del 2019), che ha introdotto nel d. lgs. n. 81 del 2015 un nuovo art. 47 bis, il cui comma 2 prevede: "*2. Ai fini del presente decreto si considerano piattaforme digitali i programmi e le procedure informatiche delle imprese che, indipendentemente dal luogo di stabilimento, organizzano le attività di consegna di beni, fissandone il prezzo e determinando le modalità di esecuzione della prestazione*"; la seconda, è che il lavoro mediante piattaforme digitali entra nella sfera d'influenza del diritto del lavoro quando "*la piattaforma digitale, lungi dall'essere mero luogo di incontri fra fornitori e fruitori di servizi, funga da vero datore di lavoro, esercitando i relativi poteri*".

con l'utilizzatore finale. Sotto il primo profilo, il tribunale fonda la natura indiretta della discriminazione sul carattere apparentemente *neutro* dei criteri utilizzati dal sistema (cecità) di prenotazione delle sessioni di lavoro senza distinguere le ragioni sottostanti alla richiesta di cancellazione della sessione prenotata (partecipazione a manifestazioni sindacali, esigenze familiari o di salute) si risolve in un trattamento discriminatorio irragionevole e in una disparità di trattamento perché il sistema di IA tratterebbe in modo identico situazioni tra loro diverse (equiparando chi non partecipa alla sessione lavorativa per futili motivi e chi per giustificate esigenze personali) violando il principio di eguaglianza; il fattore discriminatorio nel concreto è quello della disparità di trattamento per motivi sindacali (ascrivibile nella categoria delle convinzioni personali come si legge nell'ordinanza). Sotto il secondo profilo, interessante l'operazione logico-giuridica compiuta, che addossa la responsabilità sull'utilizzatore finale senza coinvolgimento del programmatore della piattaforma accertata come discriminatoria.

In tal modo, si risolve il problema del nesso causale ritenendo, da un lato, la consapevolezza della società datrice di lavoro sul fatto che il sistema fosse discriminatorio e che essa stessa avrebbe potuto intervenire sul funzionamento della piattaforma per correggerne le rigidità; difatti, in due ipotesi, la società datrice di lavoro aveva ammesso di "togliersi la benda che la rendeva cieca" (ovvero nei casi di infortunio del rider su turni consecutivi e malfunzionamento del sistema che impedisce al lavoratore il *log-in*); in sostanza, è l'uomo che conosce la rigidità del funzionamento del sistema della macchina, ma la utilizza e se ne avvantaggia (non volendo discriminare, ma assumendo il rischio e disinteressandosene). Questa decisione è stata seguita da successivi provvedimenti del Garante (ordinanze-ingiunzioni del giugno e del luglio 2021 su altre società di distribuzione di generi alimentari mediante *riders* come FOODINHO e GLOVO), Il Garante per la protezione dei dati personali parla di *profilazione* rifacendosi al contesto regolatorio *ratione temporis* vigente [36].

Dal raffronto tra le decisioni, risulta una diversità di approccio tra giurisdizione e Autorità garante per la protezione dei dati personali, la prima legata alle tradizionali categorie classiche del diritto discriminatorio e la seconda rivolta quasi totalmente ai profili discriminatori della *profilazione* nel trattamento dei dati personali (che

[36] *Regolamento europeo generale sulla protezione dei dati,* Reg. UE 679 del 2016 (vedi punto 71 del considerando ove la profilazione è definita una «forma di trattamento automatizzato dei dati personali che valuta aspetti personali concernenti una persona fisica, in particolare al fine di analizzare o prevedere aspetti riguardanti il rendimento professionale, la situazione economica, la salute, le preferenze o gli interessi personali, l'affidabilità o il comportamento, l'ubicazione o gli spostamenti dell'interessato

costituisce una delle forme della discriminazione algoritmica). Ciò in larga misura può essere ricondotto all'evoluzione della regolamentazione normativa che ha trovato in tema di protezione dei dati personali una più rapida e organica definizione in ambito europeo rispetto a quella concernente il fenomeno discriminatorio al cospetto degli strumenti di IA, messa a punto soltanto nel 2025[37].

Anche a seguito dei recenti interventi regolatori, si avverte la necessità di una nozione giuridica di discriminazione artificiale che comprenda in sé sia le classiche tradizionali categorie di discriminazione indiretta sia quelle derivanti da categorie di discriminazione inedita derivanti dai sistemi di IA ove le possibilità discriminatorie si amplificano sia "per associazione" sia perché vengono a manifestarsi in maniera multipla e intersezionale.

La Corte di Giustizia ha affermato il diritto ad ottenere una spiegazione con la decisione C 203-22, 27 febbraio 2025 CK v. Magistrat der Stadt Wien.

Un giudice austriaco ha disposto il rinvio pregiudiziale alla Corte di Giustizia in un caso in cui, un operatore di telefonia mobile aveva

[37] Si vedano, in proposito tre recenti arresti della Corte di cassazione in merito al problema della trasparenza dell'algoritmo e trattamento dei dati personali: 1) il primo sulla questione del Rating reputazionale e il caso "Mevaluate", Cass. Sez. 1 25/05/2021 n. 14381 secondo cui in tema di trattamento di dati personali, il consenso è validamente prestato solo se espresso liberamente e specificamente in riferimento ad un trattamento chiaramente individuato; ne consegue che nel caso di una piattaforma *web* (con annesso archivio informatico) preordinata all'elaborazione di profili reputazionali di singole persone fisiche o giuridiche, incentrata su un sistema di calcolo con alla base un algoritmo finalizzato a stabilire punteggi di affidabilità, il requisito della consapevolezza non può considerarsi soddisfatto ove lo schema esecutivo dell'algoritmo e gli elementi di cui si compone restino ignoti o non conoscibili da parte degli interessati; 2) il secondo, Cass Sez. 1 10/10/2023 n. 28358, secondo cui il consenso dell'interessato può ritenersi validamente prestato ove sia espresso in riferimento ad informazioni circa le finalità e le modalità del trattamento; ne consegue che, ove i dati personali siano inseriti in un algoritmo, il consenso dovrà avere ad oggetto anche le modalità di tale procedimento, non occorrendo, invece, che il linguaggio matematico ed informatico sia esteso agli utenti, né che dagli stessi sia effettivamente compreso. (Nella specie, la S.C. ha affermato che, in relazione ad un trattamento di dati da parte di un'associazione, effettuato al fine di elaborare profili reputazionali concernenti persone fisiche e giuridiche attraverso un algoritmo di calcolo, doveva ritenersi validamente prestato un consenso che aveva ad oggetto un determinato sistema di parametri, non ritenendo invece necessario che fosse indicato il peso specifico delle singole componenti considerate o i meccanismi matematici di interazione tra i vari fattori); 3) Cass. Sez. 1, 13/05/2024, n. 12967 ove è stato affermato che ai sensi dell'art. 9 del Regolamento (UE) n. 679 del 2016, ricorre un trattamento di dati biometrici, come definiti dall'art. 4, n. 14 del medesimo regolamento, quando i dati personali sono ottenuti mediante un trattamento tecnico automatizzato specifico, realizzato con un s*oftware* che, sulla base di riprese e analisi delle caratteristiche fisiche, fisiologiche o comportamentali di una persona fisica, le elabora, evidenziando comportamenti o elementi anomali, e che perviene a un esito conclusivo, costituito da una elaborato video o foto che consente (o che conferma) l'identificazione univoca della persona fisica, restando irrilevante la circostanza che l'esito finale del trattamento sia successivamente sottoposto alla verifica finale di una persona fisica. (Fattispecie relativa all'impiego di un sistema di supervisione proctoring, nell'ambito dello svolgimento delle prove scritte d'esame di studenti universitari, al fine di identificare questi ultimi o di verificarne il corretto comportamento durante lo svolgimento della prova d'esame in video conferenza).

negato ad un cliente la stipulazione di un contratto, di appena 10 euro mensili, adducendo che non era sufficientemente solvibile; l'operatore si basava su una valutazione di presunta affidabilità creditizia della cliente, effettuata per via automatizzata da una terza società (Dun & Bradstreet Austria) specializzata nella fornitura di valutazioni di tale tipo.

Nella controversia nazionale che ne è scaturita, un giudice austriaco ha dichiarato, con decisione definitiva, che la società terza, la Dun & Bradstreet Austria, aveva violato il Regolamento generale sulla protezione dei dati (GDPR), in particolare perché non avrebbe fornito alla cliente «informazioni significative sulla logica utilizzata» nel processo decisionale automatizzato in questione.

Poiché la Dun & Bradstreet Austria era rimasto inattiva nonostante la sentenza definitiva, CK ha chiesto al Comune di Vienna di eseguire la sentenza. L'autorità esecutiva ha respinto la richiesta in quanto D&B ritenendo che avesse già adempiuto all'obbligo di fornire informazioni. CK si è quindi rivolta nuovamente al Tribunale amministrativo di Vienna, che ha rinviato il caso alla CGUE per una pronuncia pregiudiziale.

La Corte di Giustizia ha affermato che "l'interessato può pretendere dal titolare del trattamento, a titolo di «informazioni significative sulla logica utilizzata», che quest'ultimo gli spieghi, mediante informazioni pertinenti e in forma concisa, trasparente, comprensibile e facilmente accessibile, la procedura e i principi concretamente applicati per utilizzare, con mezzi automatizzati, i dati personali relativi a tale interessato al fine di ottenerne un risultato determinato, come un profilo di solvibilità".

Altra precisazione degna di nota, è quella secondo cui il diritto di ottenere «informazioni significative sulla logica utilizzata» non si può soddisfare "né la semplice comunicazione di una formula matematica complessa, come un algoritmo, né la descrizione dettagliata di tutte le fasi di un processo decisionale automatizzato, in quanto nessuno di questi metodi costituirebbe una spiegazione sufficientemente concisa e comprensibile" (par. 59 della decisione) dovendo invece l'avente diritto essere posto nelle condizioni di conoscere quali dei sui dati personali siano stati impiegati dal sistema automatizzato.

5. La necessità della supervisione umana dei sistemi di IA

La necessità della *human oversight* è stata affermata da una sentenza della Corte d'appello distretto della Columbia Ross v. USA, 25 febbraio 2025, sull'impiego di Chat GPT; in un caso giudiziario relativo alla condanna di una donna che aveva lasciato il proprio cane in automobile a forti temperature esterne, il giudice, per risolvere un caso,

aveva interpellato Chat GPT al fine di stabilire che cosa significasse *commom knowledge* (fatto notorio) e l'opinione di maggioranza ha espresso perplessità in merito all'impiego del sistema di IA per accertare la nozione di fatto notorio :"nutriamo dubbi sul fatto che ChatGPT sia "un buon indicatore di ciò che è, e ciò che non è, conoscenza comune".

Ma la stessa *human oversight* può non rivelarsi un antidoto sufficiente; ne è un esempio il caso deciso dalla Corte federale australiana Tickle v. Giggle for Girls Pty Ltd. 23 agosto 2024.

Roxanne Tickle, una donna *transgender*, ha fatto causa all'App per sole donne "Giggle for Girls" per incontri tra donne, dopo che il suo account era stato sospeso, posto che dal sistema di identificazione biometrica impiegato la sua identificazione era stata definita come "uomo" e l'addetto umano alla sorveglianza del sistema per verificare la correttezza aveva confermato l'errore della macchina nel non riconoscere la persona come donna transgender. La Corte federale ha censurato come discriminazione indiretta la fattispecie esaminata, ma non imputandola a quella compiuta dal sistema di IA, piuttosto al controllore umano responsabile di non aver assolto al suo compito di supervisione.

Una prima risposta al problema del rischio connesso ai sistemi di riconoscimento facciale in ambito europeo (importante precedente nella futura giurisprudenza europea sul tema della sorveglianza di massa) è stata data dalla Corte europea dei diritti dell'uomo, 4 luglio 2023, ricorso n. 11519/20, Glukhin c. Russia.

Un cittadino russo, autore di una manifestazione politica solitaria e pacifica tenuta all'interno della metropolitana di Mosca (recando con sè un cartello ritraente un dissidente, Konstanin Kotov e uno striscione che ne lamentava l'incarcerazione), veniva condannato per un illecito amministrativo non avendo provveduto a comunicare in via preventiva la sua protesta. Ai fini dell'accertamento di responsabilità erano utilizzate le riprese delle telecamere di sorveglianza a circuito chiuso, presenti nella stazione della metropolitana, che avevano registrato il suo passaggio; in particolare, esaminando i relativi fotogrammi registrati tramite tecnologia di riconoscimento facciale, la polizia era stata in grado di identificare, localizzare e arrestare il manifestante.

Secondo la Corte di Strasburgo, il trattamento dei dati personali biometrici del soggetto si è tradotto in una interferenza ingiustificata nel rispetto del suo diritto alla vita privata (art. 8 Convenzione EDU) e una illegittima compressione della libera manifestazione del pensiero (art. 10 Convenzione EDU). La tecnologia del riconoscimento facciale dal vivo, infatti, in quanto strumento particolarmente intrusivo, esige un alto livello di giustificazione per risultare necessario in una società democratica. Tale requisito - affermano i Giudici - non è tuttavia ravvisabile nel caso di

specie, in quanto il manifestante veniva multato per un illecito minore, che non aveva determinato alcuna situazione di pericolo, risultando la misura del tutto sproporzionata e non giustificata da alcuna esigenza di protezione sociale.

Sotto il profilo della violazione dell'art. 8 CEDU la Corte afferma che ogni raccolta e conservazione di dati personali costituisce sempre e di per sé stessa una ingerenza nel diritto ella vita privata. Il contenuto della vita privata si estende anche quando associato all'impiego di tecnologie di AI identificazione biometrica, la Corte afferma il diritto dell'individuo non soltanto di rifiutare "la pubblicazione della propria immagine, ma comprende anche il diritto dell'individuo di opporsi alla registrazione, alla conservazione e alla riproduzione dell'immagine da parte di un'altra persona" (par. 66 della decisione).

Ancora la Corte ammonisce sull'estrema necessità che ciascun ordinamento nazionale si munisca di sistemi normativi adeguati nel disciplinare l'impiego di tali tecnologie tenuto conto che "i dati personali che rivelano opinioni politiche, come le informazioni sulla partecipazione a proteste pacifiche, rientrano nelle categorie speciali di dati sensibili che richiedono un livello di protezione più elevato" e che "nel contesto della raccolta e del trattamento dei dati personali, è pertanto essenziale disporre di norme chiare e dettagliate che disciplinino la portata e l'applicazione delle misure, nonché di garanzie minime riguardanti, tra l'altro, la durata, la conservazione, l'utilizzo, l'accesso di terzi, le procedure per preservare l'integrità e la riservatezza dei dati e le procedure per la loro distruzione, fornendo così garanzie sufficienti contro il rischio di abusi e arbitrarietà" (par.76 e 77 della decisione. Evidenzia che nel caso in esame "Il diritto interno non contiene alcuna limitazione alla natura delle situazioni che possono dare luogo all'uso della tecnologia di riconoscimento facciale, alle finalità previste, alle categorie di persone che possono essere prese di mira o al trattamento di dati personali sensibili" (par. 83 della decisione).

La Corte di Strasburgo, pertanto, non ha ritenuto che l'impiego di sistemi di identificazione biometrica sia vietato in assoluto in linea con quanto previsto dal Reg Act IA e dalla Convenzione Quadro del Consiglio d'Europa, ha considerato, però, necessaria una regolamentazione normativa che autorizzi l'utilizzo dei sistemi di riconoscimento facciale in tipologie definite di casi (di pubblica sorveglianza e sicurezza) e che assicuri la supervisione da parte dell'ago. Stigmatizza l'uso sproporzionato dello strumento di controllo e la sua non conformità rispetto alle caratteristiche di una Stato democratico (il rapido sviluppo delle tecnologie di IA come la nuova frontiera dei diritti civili).