

## La discriminazione algoritmica

### *Algorithmic discrimination*

**Marco Peruzzi**

*Associato di Diritto del Lavoro nell'Università di Verona*

**SOMMARIO:** 1. Premessa. – 2. La discriminazione algoritmica e la risposta regolativa Ue alle sfide dell'IA. – 2.1. Ruolo del GDPR e rilevanza del raccordo con l'*AI Act*. – 2.2. Discriminazioni e processi automatizzati: le specifiche garanzie previste nella proposta di direttiva in materia di lavoro su piattaforma digitale. – 3. La tutela contro la discriminazione algoritmica: riflessioni a partire dalla dir. Ue 2023/970... – 4. ... e guardando oltreoceano.

*Sinossi:* Il saggio cerca di individuare, *de iure condito* e *condendo*, il quadro normativo a tutela del lavoratore rispetto al rischio di discriminazione derivante dall'uso datoriale di sistemi di IA. Particolare attenzione è dedicata ad alcune fonti in via di adozione a livello Ue, segnatamente l'*AI Act* e la direttiva in materia di lavoro su piattaforma digitale, e alla necessità di considerare, in termini di reciproco raccordo, una pluralità di fonti, anche di matrice non lavoristica, quale anzitutto il GDPR. Alcune prospettive di possibile evoluzione del sistema di garanzie sono formulate analizzando il modello regolativo previsto dalla direttiva in materia di trasparenza retributiva. Spunti di riflessione sono tratti altresì osservando il modello di contrasto alla discriminazione algoritmica utilizzato dalla *New York City Local Law 144/2021*.

*Abstract:* *The essay aims at identifying, de iure condito and condendo, the legal framework protecting the employee against the risk of discrimination arising from the employer's use of AI systems. Particular attention is given to some sources being adopted at the European level, namely the AI Act and the Platform Work Directive, and to the need to consider, also in terms of mutual connection, a plurality of sources, such as the GDPR. Some perspectives on the possible evolution of the system of guarantees are drawn by analysing the regulatory instruments laid down in Directive 2023/970 on pay transparency. Food for thought is also drawn by looking at the algorithmic anti-discrimination model used by New York City Local Law 144/2021.*

**Parole-chiave:** Discriminazione – Algoritmo – Intelligenza artificiale – Approccio basato sul rischio – Trasparenza – Sorveglianza umana – Trattamento dei dati personali – Lavoro su piattaforma digitale

**Key-words:** Discrimination – Algorithm – Artificial intelligence – Risk-based approach- transparency – Human oversight – Processing of personal data – Platform work

## 1. Premessa

Per articolare una riflessione sul tema della discriminazione algoritmica, è necessario partire da una considerazione: l'elemento dell'intenzionalità è irrilevante ai fini dell'integrazione della fattispecie illecita, la nozione di discriminazione prevista dalle fonti di matrice Ue è oggettiva. Per nulla incide, quindi, sulla sua configurazione il fatto che la regola sia consegnata da un sistema algoritmico, «cieco» e «incosciente». A rilevare è l'effetto che l'applicazione e l'operatività di tale regola determina.

Ancora, nessuna garanzia contro il rischio di discriminazione è fornita dal «crisma» della «neutralità» attribuito all'azione automatizzata, governata da algoritmi e/o sistemi di IA<sup>1</sup>: come ci insegna il diritto antidiscriminatorio indicando gli elementi costitutivi della discriminazione indiretta, anche una sequenza di istruzioni ad applicazione generalizzata e «apparentemente neutra», ossia priva di riferimenti a un fattore di rischio, può integrare una fattispecie illecita a fronte dell'impatto differenziato che determina o può determinare a detrimento dei portatori di un fattore di rischio. È questa l'ipotesi che si è verificata nei più noti casi di discriminazione algoritmica conosciuti a livello di ordinamento italiano, ossia quelli degli algoritmi utilizzati da Deliveroo e da Foodinho per la profilazione e il *ranking* dei propri *rider*, nonché la conseguente organizzazione del sistema di prenotazione degli *slot* di lavoro. Come affermato rispettivamente dai Tribunali di Bologna e di Palermo<sup>2</sup>, l'incoscienza della macchina e l'applicazione indifferenziata dei parametri di valutazione non escludono l'accertamento della discriminazione, proprio per il particolare svantaggio che tali parametri implicano nei confronti dei portatori di determinati fattori di rischio, quali anzitutto – in questi casi – l'affiliazione sindacale, la religione e la disabilità.

Una discriminazione algoritmica può, d'altra parte, verificarsi a valle di un processo automatizzato molto differente da quello poc'anzi descritto. Si pensi, per richiamare il più celebre esempio, al caso Amazon: il sistema di IA utilizzato per il *ranking* dei candidati a posizioni di sviluppatori di *software* era stato addestrato sulla base dei curricula ricevuti dall'azienda negli ultimi dieci anni in modo che “imparasse” i parametri per individuare i soggetti “statisticamente migliori”. Di conseguenza aveva posizionato ai primi posti della classifica soltanto uomini.

Il caso Amazon, messo a confronto con quelli menzionati di Deliveroo e Foodinho, evidenzia come le forme di discriminazione che si possono configurare nei processi algoritmici sono molteplici e questo dipende anzitutto dalla diversa tipologia di algoritmo utilizzato. In altre parole, quando si parla di discriminazione algoritmica, è necessario tenere in considerazione che non tutti gli algoritmi sono uguali.

<sup>1</sup> ABRIANI, SCHNEIDER, *Diritto delle imprese e intelligenza artificiale. Dalla Fintech alla Corptech*, Bologna, 2021, 45.

<sup>2</sup> Trib. Bologna, 31 dicembre 2020, in *RIDL* 2021, II, 188 con nota di GAUDIO; Trib. Palermo, 17 novembre 2023.

Nei casi Deliveroo e Foodinho, l'algoritmo è *logic-based*, quindi è costituito o comunque muove da una sequenza di regole di inferenza logica scritte in sede di programmazione: per quanto complesso e articolato sia il processo computazionale nel calcolo e nella ponderazione dei diversi e numerosi parametri implicati<sup>3</sup>, il sistema rimane di carattere deterministico (ossia, a parità di dati d'ingresso, può condurre a un solo possibile percorso e a un solo possibile risultato), opera sulla base di istruzioni e rapporti di condizionalità (*if/then*) precodificati, nonché entro i confini della rappresentazione formale della conoscenza fornita. La discriminazione qui si può annidare nella regola tradotta in linguaggio di programmazione: fondamentalmente, lo scenario non è così distante da quello che si potrebbe configurare alla previsione e applicazione di un regolamento aziendale che riduce l'ammontare del premio di produttività in proporzione alle assenze, senza distinguerle in base alle ragioni che le giustificano; ovvero che fissa il medesimo periodo di comportamento per lavoratori disabili e non. La particolarità risiede nella digitalizzazione della regola all'interno di un algoritmo e nella costruzione di un processo decisionale automatizzato a partire dalla sua esecuzione.

Diverso è il caso Amazon, dove a essere utilizzato è un algoritmo di *machine learning*: i rapporti di condizionalità lasciano il posto a rapporti di correlazione, la consequenzialità logica lascia il posto alla statistica e si passa dal ragionamento automatico degli algoritmi deterministici a un apprendimento automatico, più o meno profondo a seconda dell'utilizzo di reti neurali e dei livelli di stratificazione delle stesse. Qui la macchina non dispone di un modello di relazioni logiche precodificato per estrarre e ottimizzare l'*output*: è l'obiettivo del sistema elaborarne uno, individuandone e ottimizzandone i parametri a partire da un ampio serbatoio di dati. È quindi la qualità dei dati di *training* utilizzati in sede di addestramento, nonché dei dati di *input* in caso di adattabilità e apprendimento continuo della macchina, a incidere non solo sul livello di accuratezza dell'*output* (quindi di funzionalità ed efficienza del processo rispetto all'obiettivo) ma anche sul rischio di *bias*, di distorsioni, di discriminazione. Si può configurare tanto la possibilità che i dati che addestrano l'algoritmo siano intrisi di pregiudizi e pertanto "insegnino" alla macchina a riprodurre una regola discriminatoria (come nel caso Amazon, il *dataset* registra per una determinata posizione una percentuale minima di donne e il sistema "impara" che per tale posizione è meglio scegliere un uomo)<sup>4</sup>, quanto la possibilità che le correlazioni (anche imprevedibili) che l'algoritmo di *machine learning* individua all'interno dei *dataset* consegnati nascondano degli impatti discriminatori. A tal ultimo riguardo, la letteratura riporta gli esempi di un algoritmo che associa una maggior capacità di programmazione informatica alla frequentazione di un determinato *website* di manga giapponesi, con conseguente

<sup>3</sup> Lo dimostra Trib. Torino, 5 agosto 2023, in *RGL*, 2023, II, 416, con nota di DONINI laddove affronta il problema della trasparenza dell'algoritmo utilizzato da Foodinho.

<sup>4</sup> Parla al riguardo di «algoritmo "specchio"» LO FARO, *Algorithmic Decision Making e gestione dei rapporti di lavoro: cosa abbiamo imparato dalle piattaforme*, in *federalismi.it*, 2022, 25, 189 ss.

selezione di candidati prevalentemente maschi; o di un algoritmo che correli l'acquisto di *e-book* con una migliore performance lavorativa andando a escludere o penalizzare candidati o lavoratori ipovedenti<sup>5</sup>.

Ferma l'applicabilità della strumentazione giuridica fornita dal diritto antidiscriminatorio – dalle definizioni al sistema di eccezioni e cause di giustificazione, fino al regime probatorio speciale – la determinazione algoritmica della fattispecie richiede di verificare quanto il modello regolativo che si sta configurando in risposta alle sfide dell'intelligenza artificiale (IA) possa essere funzionale all'applicazione e all'effettività della tutela contro il rischio di discriminazione. Proprio tale profilo costituirà l'oggetto dell'analisi che si intende sviluppare.

## 2. La discriminazione algoritmica e la risposta regolativa Ue alle sfide dell'IA

Ragionare di (ri)costruzione delle tutele del lavoratore di fronte all'utilizzo datoriale di sistemi algoritmici significa, da un lato, comprendere le scelte di metodo e di contenuto che connotano la risposta regolativa Ue alle sfide dell'IA, dall'altro, verificare come si integrino e si saldino lungo i meridiani di tali scelte i diversi plessi regolativi coinvolti, siano essi di matrice lavoristica o meno. Questo secondo passaggio è tanto più necessario se si considera il carattere orizzontale del Regolamento sull'IA (c.d. *AI Act*, di prossima adozione)<sup>6</sup>, dichiaratamente volto a stabilire un livello minimo e complementare di garanzie, ossia a integrare le altre fonti Ue e/o interne, senza pregiudicarne l'applicazione (v. *infra*)<sup>7</sup>.

Rinviando ad altra sede un'analisi più esaustiva di tali profili<sup>8</sup>, basti qui evidenziare come in tema di IA i diversi livelli di azione, euro-unitari e internazionali, condividano alcune scelte di fondo, metodologiche e di contenuto, in particolare un approccio basato sul ri-

<sup>5</sup> PECK, *They're watching you at work*, in *theatlantic.com*, dicembre 2013; KING, MRKONICH, "Big Data" and the Risk of Employment Discrimination, in *Oklahoma Law Review*, 2016, 3, 557 ss.; DAGNINO, *People Analytics: lavoro e tutele al tempo del management tramite big data*, in *Labour & Law Issues*, 2017, 1, I.1 ss.

<sup>6</sup> Cfr. ADINOLFI, *L'intelligenza artificiale tra rischi di violazione dei diritti fondamentali e sostegno della loro promozione: considerazioni sulla (difficile) costruzione di un quadro normativo dell'Unione*, in PAJNO, DONATI, PERRUCCI (a cura di), *Intelligenza artificiale e diritto: una rivoluzione? Diritti fondamentali, dati personali e regolazione*, vol. I, Bologna, 2022, 127 ss.

<sup>7</sup> Nella prospettiva lavoristica, può richiamarsi, a tal ultimo riguardo, il Cons. 9, ove si afferma che non deve essere intaccata la protezione dei lavoratori a livello Ue e nazionale, né la possibilità di esercitare il diritto di concludere e applicare contratti collettivi o di intraprendere azioni collettive. Soprattutto è importante segnalare l'art. 2, par. 11, secondo cui quanto disposto dal Regolamento funge da livello minimo di tutela e non impedisce agli Stati membri o all'Unione di mantenere o introdurre leggi, regolamenti o disposizioni amministrative più favorevoli ai lavoratori in termini di protezione dei loro diritti in relazione all'uso dei sistemi di IA da parte dei datori di lavoro, o di incoraggiare o consentire l'applicazione di contratti collettivi più favorevoli.

<sup>8</sup> Si consenta il rinvio a PERUZZI, *Intelligenza artificiale e lavoro: l'impatto dell'AI Act nella ricostruzione del sistema regolativo Ue di tutela*, in BIASI (a cura di), *Diritto del lavoro e intelligenza artificiale*, Milano, 2024 (in corso di pubblicazione).

schio e l'individuazione, quali garanzie imprescindibili di una prospettiva antropocentrica, della trasparenza e supervisione umana.

Sono scelte che trovano specifica declinazione nell'*AI Act*<sup>9</sup>.

In tale quadro di regole, il fornitore di un sistema di IA è gravato, prima, dall'obbligo di verificare se il sistema debba essere classificato come "ad alto rischio"<sup>10</sup>, quindi, in tal caso<sup>11</sup>, dall'obbligo di mappare e valutare i rischi e di predisporre un sistema di misure per la loro gestione (art. 9). Trattasi, in particolare, dei rischi noti e ragionevolmente prevedibili in condizioni di corretto utilizzo del sistema ovvero di quelli che potrebbero emergere in caso di un suo uso improprio ragionevolmente prevedibile. Rispetto alla qualità dei dati di addestramento, convalida e prova, al fornitore si richiede di adottare adeguate pratiche di *governance* e gestione, che comprendano, tra gli altri, un esame atto a valutare i possibili *bias* (molto rilevanti, come visto, nell'ottica antidiscriminatoria, soprattutto in caso di *machine learning*), nonché misure volte a localizzarli, prevenirli e mitigarli (art. 10). Si prevede segnatamente che tali pratiche riguardino «un esame atto a valutare le possibili distorsioni suscettibili di incidere sulla salute e sulla sicurezza delle persone, di avere un impatto negativo sui diritti fondamentali o di comportare discriminazioni vietate dal diritto dell'Unione, specie laddove gli *output* di dati influenzano gli *input* per operazioni future» (art. 10, par. 2, lett. f) e «le misure adeguate per individuare, prevenire e attenuare le possibili distorsioni individuate» (art. 10, par. 2, lett. g). Al riguardo, è interessante segnalare che l'art. 10, par. 5, dell'*AI Act*, autorizza «eccezionalmente» (nonché a determinate condizioni) i fornitori a trattare categorie particolari di dati di cui all'art. 9 GDPR nella misura in cui sia strettamente necessario per garantire il rilevamento e la correzione delle distorsioni. Il fornitore ha, quindi, l'obbligo di istituire un sistema di gestione della qualità a garanzia della *compliance* con il Regolamento (art. 17), comprensivo dei sopraccitati sistemi di gestione dei rischi e dei dati.

<sup>9</sup> Cfr. LOI, *Il rischio proporzionato nella proposta di regolamento sull'IA e i suoi effetti nel rapporto di lavoro*, in *federalismi.it* 2023, 29, 171 ss.; BARBERA, "La nave deve navigare". *Rischio e responsabilità al tempo dell'impresa digitale*, in *Labour & Law Issues*, 2023, 2, 1 ss.; ALAIMO, *Il Regolamento sull'Intelligenza Artificiale: dalla proposta della Commissione al testo approvato dal Parlamento. Ha ancora senso il pensiero pessimistico?*, in *federalismi.it*, 2023, 25.

<sup>10</sup> La classificazione del sistema come "ad alto rischio" non deriva soltanto dalla sua inclusione all'interno di un determinato elenco di aree critiche e casi d'uso, ma anche da una valutazione effettuata dal fornitore circa la sussistenza di un effettivo rischio significativo di danno alla salute, alla sicurezza o ai diritti fondamentali (art. 6, par. 3). Ai fini dell'analisi che si intende qui sviluppare, è importante evidenziare che i sistemi utilizzati nel campo dell'occupazione, della gestione dei lavoratori e dell'accesso al lavoro autonomo (punto 4, Allegato III) sono inclusi nell'elenco delle aree critiche. Come detto, detti sistemi devono comunque essere sottoposti alla valutazione di rischio effettivo da parte del fornitore. Ai sensi del Regolamento, tuttavia, un sistema di IA deve essere sempre considerato ad alto rischio laddove effettui una profilazione di persone fisiche (art. 6, par. 3) e questo è molto probabile che si verifichi in caso di utilizzo di processi automatizzati di monitoraggio o decisionali sul lavoro. Può ritenersi, quindi, tendenzialmente da escludere una classificazione come "a basso rischio" dei sistemi destinati a essere utilizzati nell'esercizio di poteri datoriali e nei confronti dei lavoratori.

<sup>11</sup> Specifici obblighi di trasparenza, a carico del fornitore o del *deployer*, sono stabiliti - a prescindere dalla qualificazione ad alto rischio e, in tal caso, senza pregiudizio per i relativi requisiti - per i sistemi destinati a interagire direttamente con persone fisiche e per quelli a carattere generativo, compresi i sistemi di IA per finalità generali (*general purpose AI* - GPAD), a tutela di chi sia esposto ai c.d. *deep fakes* (Capo IV, Art. 50).

In tale prospettiva, può essere utile richiamare la linea di lettura che qualifica l'assenza di blocchi contro deviazioni discriminatorie come difetto del prodotto ai fini della *product liability*<sup>12</sup>, ma anche l'avvertimento di chi ritiene che interventi di progettazione finalizzati a escludere formalmente la rilevanza dei fattori di rischio potrebbero poi paradossalmente andare a depotenziare la prova del nesso di causalità tra il trattamento subito e il fattore di rischio stesso<sup>13</sup>.

A interessare la prospettiva lavoristica nel quadro dell'*AI Act* è, tuttavia, soprattutto la figura del *deployer*, posto che tale sarà, in tale contesto normativo, la qualificazione assunta da un datore di lavoro che faccia uso di sistemi di IA. Proprio grazie alle informazioni ricevute dal fornitore, egli deve essere messo nelle condizioni di comprendere la logica di funzionamento del sistema nonché di interpretare e spiegare gli *outcome* che consegna<sup>14</sup>. Vari sono gli elementi di cui deve essere messo a conoscenza: il livello di accuratezza, comprese le metriche, robustezza e cybersicurezza del sistema; qualsiasi circostanza conosciuta o prevedibile che possa comportare rischi per la salute e la sicurezza o per i diritti fondamentali; ove opportuno le prestazioni del sistema per quanto riguarda le persone o i gruppi di persone specifici sui quali il sistema è destinato a essere utilizzato; le specifiche dei dati di ingresso o qualsiasi altra informazione pertinente in termini di serie di dati di formazione, convalida e prova utilizzati (art. 13).

In linea generale, il *deployer* è obbligato ad attuare le misure di sorveglianza umana indicate dal fornitore nelle istruzioni per l'uso<sup>15</sup> (art. 26) ed è soggetto, a sua volta, a un obbligo di valutazione dell'impatto del sistema sui diritti fondamentali (art. 27). Quest'ultimo

<sup>12</sup> RUFFOLO OSSERVA, in particolare, che tra i difetti del prodotto può essere ricompresa «la mancata introduzione *by design*, nella componente algoritmo di autoapprendimento, di un codice macchina con “blocchi” idonei a inibire le future deviazioni della IA (ivi compresa la “discriminazione algoritmica”)» e a fissare delle «soglie di comportamento invalicabili» (*Artificial intelligence e responsabilità. «Persona elettronica» e teoria dell'illecito*, in PAJNO, DONATI, PERRUCCI (a cura di), *Intelligenza artificiale e diritto: una rivoluzione? Amministrazione, responsabilità, giurisdizione*, vol. II, Bologna, 2022, 247). Come si legge nella relazione di accompagnamento alla proposta di direttiva volta ad aggiornare la normativa Ue sulla responsabilità per danno da prodotti difettosi, «la proposta di direttiva sulla responsabilità per danno da prodotti difettosi garantirà che, in caso di sistemi di IA difettosi che provocano danni fisici o patrimoniali o la perdita di dati, sia possibile chiedere il risarcimento al fornitore del sistema di IA o a un fabbricante che integri un sistema di IA in un altro prodotto» (Bruxelles, 28.9.2022 COM(2022) 495 final, par. 1.1).

<sup>13</sup> GIACOMELLI, *Big brother is «gendering» you. Il diritto antidiscriminatorio alla prova dell'intelligenza artificiale: quale tutela per il corpo digitale?*, in *BioLaw Journal – Rivista di Bio Diritto*, 2019, 269 ss.; AGRAWAL GANS, GOLDFARB, *Macchine predittive*, Milano, 2018.

<sup>14</sup> Cfr. CIUCCIOVINO, *La disciplina nazionale sulla utilizzazione della intelligenza artificiale nel rapporto di lavoro*, in *Lavoro Diritti Europa*, 2024, n. 1, 1 ss. L'Autrice evidenzia, in particolare, come in tal senso l'*AI Act* rafforzi una prospettiva di controllo *ex post*, basata sulla «spiegabilità degli effetti e dell'impatto (cioè sul “cosa fa” la IA, anziché sul “come fa”» (15).

<sup>15</sup> Le misure di supervisione umana sono finalizzate a prevenire o ridurre al minimo i rischi che possono emergere quando il sistema viene utilizzato in conformità allo scopo previsto o in condizioni di uso improprio ragionevolmente prevedibili. Ai sensi dell'art. 26, il *deployer* ha discrezionalità nell'organizzazione delle proprie risorse ai fini dell'implementazione di tali misure, ma deve assegnare i relativi compiti a persone sufficientemente competenti. Come evidenzia CIUCCIOVINO, il concetto «di sottoposizione alla sorveglianza umana e quello di integrale automatizzazione [...] non sono sovrapponibili. Il primo è riferibile al controllo sulla sicurezza del funzionamento del sistema, dal punto di vista della sua coerenza e compatibilità con i principi di tutela e con i limiti posti all'impiego dell'IA», mentre il secondo «al modo con cui avviene l'incorporazione dell'*output* algoritmico nell'atto decisionale» di chi utilizza il sistema (op. cit., 16-17).

obbligo, proposto dal Parlamento, è stato invero circoscritto, nell'accordo interistituzionale raggiunto, agli organismi di diritto pubblico o agli operatori privati che forniscono servizi pubblici, come ospedali, scuole, banche e compagnie di assicurazione, ciò suggerendo un'attenzione specifica ai diritti fondamentali dell'utenza. Pare significativo tuttavia segnalare come, al di là di questa delimitazione dell'obbligo di valutazione di impatto, tutti i *deployer* di sistemi ad alto rischio siano chiamati a garantire che i dati di *input* siano pertinenti e sufficientemente rappresentativi, nella misura in cui esercitano il controllo sugli stessi (art. 26, par. 4), ma soprattutto a monitorare il funzionamento del sistema e a sospenderne l'utilizzo, con informazione al fornitore o distributore e all'organismo di vigilanza nazionale, laddove abbiano ragione di ritenere che l'uso, pur conforme alle istruzioni ricevute, possa determinare un rischio per la salute e sicurezza e i diritti fondamentali (art. 26, par. 5).

Le disposizioni che regolano la sorveglianza umana e l'obbligo, pur circoscritto, di valutazione d'impatto confermano la natura orizzontale del Regolamento.

Le prime chiariscono che l'obbligo di utilizzare il sistema in conformità alle istruzioni d'uso e alle misure di sorveglianza umana indicate dal fornitore non pregiudica gli obblighi che gravano sul *deployer* ai sensi di altre fonti Ue o interne (art. 26, par. 3); anzi, l'art. 26, par. 9, specifica che il *deployer* utilizza le informazioni ricevute dal fornitore per adempiere all'obbligo di valutazione d'impatto sulla protezione dei dati personali (DPIA) previsto dal GDPR. Sulla medesima linea di raccordo, l'art. 27, par. 4 stabilisce che se uno qualsiasi degli obblighi ivi previsti è già rispettato a seguito della DPIA, la valutazione d'impatto sui diritti fondamentali integra la valutazione condotta a norma del GDPR. Del resto, come spiega il cons. n. 9, il Regolamento mira a rafforzare l'efficacia dei diritti e dei rimedi esistenti, introducendo requisiti, vincoli e garanzie che possono svolgere un ruolo funzionale, facilitante e abilitante. Per quanto riguarda la protezione dei lavoratori, il riferimento è anzitutto alle disposizioni del GDPR, alla legislazione in materia di salute e sicurezza e al diritto antidiscriminatorio.

### **2.1. Ruolo del GDPR e rilevanza del raccordo con l'AI Act**

La sinergia e l'importanza del raccordo tra l'*AI Act* e le altre fonti Ue o interne, evidenziata nel paragrafo precedente, si coglie, anzitutto, con riguardo alla disciplina in materia di protezione dei dati personali, soprattutto con riguardo agli obblighi di proceduralizzazione *ex ante* previsti in tale quadro di regole. Centrale risulta, in particolare, nella prospettiva d'analisi che ci occupa, la funzione della DPIA in termini di verifica e prevenzione della lesione dei diritti fondamentali, tra cui quello di non discriminazione, derivanti dal trattamento (anche automatizzato) dei dati personali.

Si può segnalare, al riguardo, come il rischio di discriminazione sia individuato, nelle linee guida adottate a livello europeo, tra gli elementi che consentono di qualificare in termini di significatività l'impatto di un trattamento automatizzato e come le stesse linee guida, soprattutto nel contesto di applicazione dell'art. 22 del GDPR in tema di processi decisionali integralmente automatizzati, collochino, tra le garanzie da adottare, anche procedure e misure adeguate per prevenire discriminazioni sulla base di categorie particolari di dati

(ferma l'inutilizzabilità di tali dati nell'ambito dei processi di cui all'art. 22, se non laddove siano integrate le stringenti condizioni indicate nel paragrafo 4)<sup>16</sup>.

Si rinvia ad altra sede una più compiuta analisi del ruolo del GDPR nella ricostruzione delle garanzie del lavoratore di fronte all'uso datoriale di sistemi algoritmici.

Si può qui brevemente sottolineare come l'obbligo di effettuare una DPIA ai sensi dell'art. 35 GDPR sia inevitabile se il datore di lavoro utilizza sistemi decisionali o di monitoraggio automatizzati<sup>17</sup>. La DPIA richiede al titolare del trattamento non solo di adottare tutte le misure necessarie per assicurare e dimostrare la conformità ai principi del GDPR, compreso il principio di trasparenza: richiede anche di valutare, mitigare e gestire i rischi per *tutti* i diritti e le libertà degli interessati. Il doppio passaggio è di cruciale rilevanza: se è vero, infatti, che la *compliance* al principio di trasparenza impone già di modulare il contenuto delle protezioni in rapporto alle caratteristiche specifiche del trattamento, le misure da adottare devono rispondere altresì a una più ampia e ancor più pregnante esigenza di tutela, tra cui rientra quella del diritto alla dignità delle persone coinvolte<sup>18</sup>. All'esito di una valutazione d'impatto, al datore potrebbe essere richiesto, quindi, di adottare misure di trasparenza ovvero di sorveglianza e riesame umani simili o addirittura equivalenti a quelle espressamente prescritte dal GDPR per i processi di cui all'art. 22 (ossia quelli decisionali *integralmente* automatizzati) o a quelle previste dalla proposta di direttiva sul lavoro su piattaforma digitale (v. *infra*) o, ancora, a quelle stabilite a livello interno dall'art. 1-bis d.lgs. 26 maggio 1997, n. 152 (per i processi decisionali e di monitoraggio integralmente automatizzati), che da tali fonti e dal Regolamento sull'IA trae ispirazione<sup>19</sup>.

La DPIA include, inoltre, anche «una valutazione della necessità e della proporzionalità del trattamento in relazione alle finalità» (art. 35, par. 7, lett. b): si può ritenere che tale verifica imponga al datore di giustificare la scelta di un certo tipo di algoritmo rispetto a un determinato obiettivo organizzativo, a fronte dell'assenza di alternative meno invasive per

<sup>16</sup> Gruppo di lavoro Articolo 29, *Linee guida in materia di processi decisionali automatizzati e profilazione*, WP 251 rev.01, adottate il 3 ottobre 2017, versione emendata e adottata il 6 febbraio 2018.

<sup>17</sup> L'elenco allegato al provvedimento del garante italiano n. 467 dell'11.10.2018 prevede direttamente l'obbligo di DPIA in caso di «trattamenti automatizzati finalizzati ad assumere decisioni che producono "effetti giuridici" oppure che incidono "in modo analogo significativamente" sull'interessato» nonché «trattamenti effettuati nell'ambito del rapporto di lavoro mediante sistemi tecnologici» da cui derivi la possibilità di controllo a distanza dei dipendenti.

<sup>18</sup> Risulta, al riguardo, particolarmente significativa la segnalazione di PIZZETTI, *La protezione dei dati personali e la sfida dell'Intelligenza Artificiale*, in PIZZETTI (a cura di), *Intelligenza artificiale, protezione dei dati personali e regolazione*, Torino, 2018, 5 ss., secondo cui al cuore del GDPR «sta la persona fisica con i suoi diritti e le sue libertà. La tutela dei dati personali è, si potrebbe dire, il contenuto "specialistico" di questa normativa che però, nelle sue caratteristiche fondamentali, è, per così dire, una "legge di rango costituzionale" di diritto europeo. [...] In sostanza, in questa prospettiva, la tutela dei dati personali costituisce, a partire dal livello europeo, l'attuazione di un diritto di spessore costituzionale, funzionale alla tutela "integrale" della persona e dei suoi diritti di libertà» (p. 165).

<sup>19</sup> Per un'analisi della disposizione interna cfr., *ex multis*, ZILLI, *La via italiana per condizioni di lavoro trasparenti e prevedibili*, in *DRI*, 2023, 1, 25 ss.; DONINI, *Informazione sui sistemi decisionali e di monitoraggio automatizzati: poteri datoriali e assetti organizzativi*, in *DLM*, 2023, 85; DAGNINO, *Il diritto interno: i sistemi decisionali e di monitoraggio (integralmente) automatizzati tra trasparenza e coinvolgimento*, in BIASI (a cura di), *Diritto del lavoro e intelligenza artificiale*, Milano, 2024 (in corso di pubblicazione).

i diritti e le libertà coinvolti. In mancanza di alternative, la questione si sposta sull'adeguatezza delle misure da adottare e adattare per garantire la trasparenza e mitigare i rischi per i diritti dei lavoratori. Questo significa che, laddove l'algoritmo lasci spazio all'incertezza e all'opacità (ipotesi che il Regolamento sull'IA invero dovrebbe scongiurare, posto che al fornitore è richiesto di garantire la comprensibilità del sistema e dell'*outcome* da parte del *deployer*, v. *supra*), o la questione viene risolta, con i relativi costi, anche utilizzando le c.d. tecniche di *XAI* (*eXplainable AI*) per "aprire la scatola nera" (ad es. modelli algoritmici che generano risultati controfattuali utili non solo per spiegare la decisione, ma anche per esaminare la *fairness* del sistema, su cui v. *infra*), oppure il datore di lavoro deve abbandonare l'approccio algoritmico in questione per orientarsi verso un obiettivo organizzativo e un algoritmo diversi e meno complessi.

Particolarmente rilevante risulta, sul punto, l'innesto *de iure condendo* dei contenuti dell'*AI Act*: come già evidenziato, ai sensi dell'art. 13, la documentazione pertinente e le istruzioni per l'uso consegnate al datore, in qualità di *deployer*, devono infatti includere «qualsiasi circostanza nota o prevedibile [...] che possa comportare rischi [...] per i diritti fondamentali» (così come mappati dal fornitore all'interno del sistema di gestione dei rischi di cui all'art. 9), quindi anche gli impatti potenzialmente discriminatori delle metriche utilizzate, nonché le specifiche relative ai dati di input e ai set di dati di addestramento, convalida e prova, i quali, come detto, devono essere obbligatoriamente sottoposti, a opera del fornitore, a pratiche di *governance* e gestione, tra cui un esame atto a valutare le possibili distorsioni suscettibili di comportare discriminazioni vietate dal diritto dell'Unione (v. *supra*).

## **2.2. Discriminazioni e processi automatizzati: le specifiche garanzie previste nella proposta di direttiva in materia di lavoro su piattaforma digitale**

Pur attraverso un percorso impervio, segnato da faticosi e complicati tavoli di negoziazione, l'11 marzo 2024 Consiglio e Parlamento europeo hanno trovato un accordo anche sulla proposta di direttiva in materia di lavoro su piattaforme digitali<sup>20</sup>.

Ai fini dell'analisi che ci occupa, pare opportuno segnalare come all'interno del quadro di regole così definito si prevedano «norme più specifiche in relazione al regolamento (UE) 2016/679 per quanto riguarda l'uso e la trasparenza del processo decisionale automatizzato», nonché «misure aggiuntive [...] per salvaguardare la protezione dei [...] dati personali,

<sup>20</sup> Commissione europea, *Proposta di direttiva del Parlamento europeo e del Consiglio relativa al miglioramento delle condizioni di lavoro nel lavoro mediante piattaforme digitali*, Bruxelles, 9 dicembre 2021 COM(2021) 762 final (v. *infra*). Si segnala, nel dibattito a livello Ue, l'attenzione posta all'opportunità di introdurre una regolazione ad hoc in materia di gestione algoritmica del lavoro (oltre l'ambito delle piattaforme), prospettiva promossa dall'Etuc in una risoluzione del 6 dicembre 2022 e condivisa sia da Gualmini e Benifei all'interno del Parlamento europeo sia da Schmit, commissario europeo per il lavoro e i diritti sociali nella commissione von der Leyen.

in particolare quando le decisioni sono adottate o sostenute dal trattamento automatizzato di dati personali», ai sensi dell'88 GDPR (così il cons. n. 39)<sup>21</sup>.

In questa prospettiva si pongono, anzitutto, le limitazioni sul trattamento di dati personali stabilite dall'art. 7, in alcuni casi speculari alle proibizioni previste dall'*AI Act* nell'ambito della categoria dei sistemi a rischio inaccettabile: si pensi, in particolare, al divieto di trattare dati relativi allo stato emotivo o psicologico ovvero per desumere alcune caratteristiche "sensibili", tra cui l'origine razziale o etnica, la *status* di migrante, le opinioni politiche, le convinzioni religiose, la disabilità, l'adesione a un sindacato, l'orientamento sessuale<sup>22</sup>. Si chiarisce che tali limitazioni si applichino a tutte le persone che svolgono un lavoro mediante piattaforme digitali (subordinati e autonomi) dall'inizio della procedura di assunzione o di selezione, nonché a prescindere dal fatto che il trattamento sia effettuato attraverso un sistema integralmente o parzialmente automatizzato.

Rispetto all'uso di sistemi di monitoraggio e decisionali automatizzati (qualificante, peraltro, la stessa definizione di piattaforma digitale di cui all'art. 2, par. 1, n. 1), l'art. 9 prevede specifici obblighi di trasparenza, anche a tutela dei candidati in fase di *recruitment*: l'informativa include, tra gli altri, le categorie di dati e i parametri utilizzati dall'algoritmo, nonché la spiegazione dei motivi alla base di alcune decisioni.

Si stabilisce, quale garanzia di sorveglianza umana aggiuntiva rispetto alla DPIA del GDPR (a sua volta opportunamente oggetto di una normativa di dettaglio a rafforzamento della tutela dei lavoratori, ai sensi dell'art. 8), una periodica valutazione dell'impatto delle decisioni individuali prese o (solo) sostenute dai sistemi automatizzati, in particolare sulla parità di trattamento sul lavoro (art. 10, par. 1). A tal ultimo riguardo, si chiarisce che «qualora dalla sorveglianza o dalla valutazione [...] emerga un elevato rischio di discriminazione sul lavoro nell'uso di sistemi decisionali e di monitoraggio automatizzati [...], la piattaforma di lavoro digitale adotta le misure necessarie, compresa, se del caso, una modifica del sistema decisionale e di monitoraggio automatizzato o la cessazione del suo utilizzo, al fine di evitare tali decisioni in futuro» (art. 10, par. 3).

Il quadro di garanzie è completato dal diritto dei lavoratori (autonomi e subordinati) alla spiegazione e al riesame umano di qualsiasi decisione presa o sostenuta da un sistema decisionale automatizzato (art. 11).

Rinviando ad altra sede l'analisi della declinazione collettiva delle tutele, è importante quantomeno segnalare come il sistema di garanzie così predisposto, oltre a promuovere il ruolo delle parti sociali e della contrattazione collettiva (artt. 25 e 28), preveda specifici

<sup>21</sup> La base giuridica della proposta di direttiva è costituita dall'art. 153, par. 2, lett. b, in combinato disposto con l'art. 153, par. 1, lett. b), e dall'art. 16 par. 2 TFUE.

<sup>22</sup> L'art. 5 dell'*AI Act* inserisce tra le pratiche di IA vietate i sistemi di riconoscimento delle emozioni sul posto di lavoro, ad eccezione di quelli che riconoscono il dolore o la fatica per motivi medici o di sicurezza, ovvero i sistemi di categorizzazione biometrica delle persone volti a dedurre caratteristiche sensibili come la razza, le opinioni politiche, l'affiliazione sindacale, l'orientamento sessuale e il credo religioso.

diritti per le rappresentanze sindacali: il coinvolgimento espresso e incondizionato nella DPIA (art. 8); il diritto alla trasparenza sui processi automatizzati (art. 9, par. 1 e 4); la partecipazione alla valutazione di impatto delle decisioni automatizzate nel contesto della sorveglianza umana (art. 10, par. 1, con diritto di informazione ai sensi del par. 4); il diritto di richiedere il riesame (art. 11, par. 2); la previsione, per l'esercizio dei diritti di cui alla dir. 2002/14/CE (espressamente richiamati affinché riguardino anche la decisione di introdurre o modificare sistemi automatizzati) dell'assistenza di un esperto, pagato dalla piattaforma se si supera la soglia dei 250 lavoratori (art. 13, par. 3). I rappresentanti degli autonomi sono inclusi nel raggio di operatività della direttiva nella misura in cui siano previsti dal diritto e dalle prassi nazionali (così la definizione di cui all'art. 2, par. 1, n. 7)<sup>23</sup> e possono esercitare i diritti stabiliti in punto di trasparenza e sorveglianza umana (tranne, e questo appare molto significativo, il diritto di partecipazione alla valutazione di impatto delle decisioni automatizzate) solo nella misura in cui agiscano per conto di autonomi per quanto riguarda la protezione dei loro dati personali (art. 15).

Si può evidenziare, infine, come in seno all'art. 12, in tema di tutela della salute e sicurezza dei lavoratori subordinati (con apertura alla possibilità di introdurre disposizioni diverse, ai sensi dell'art. 28), si ponga l'accento sulla necessità di garantire la protezione contro violenza e molestie (da considerare come discriminazioni, ai sensi delle fonti del diritto antidiscriminatorio), richiedendosi che, a livello interno, si disponga che le piattaforme adottino misure preventive, compresi canali di segnalazione efficaci.

### **3. Discriminazione algoritmica e trasparenza: riflessioni a partire dalla dir. Ue 2023/970...**

Al di là delle problematiche che possono derivare dall'opacità (ossia carenza di conoscibilità e/o spiegabilità) dei processi automatizzati e/o dalla minore visibilità, all'interno di un'architettura organizzativa algoritmica, delle prerogative datoriali ivi esercitate<sup>24</sup>, la tutela antidiscriminatoria presenta già di per sé una stretta e inevitabile interconnessione con la questione della trasparenza.

<sup>23</sup> Gli stessi Orientamenti della Commissione europea a sostegno degli accordi collettivi dei lavoratori autonomi individuali con le piattaforme di lavoro digitali (esclusi dall'ambito di applicazione dell'articolo 101 TFUE) fanno salva la discrezionalità degli Stati membri nel determinare i canali di rappresentanza collettiva dei lavoratori autonomi, applicandosi a tutte le forme di trattative collettive condotte conformemente alla pratica e al diritto nazionali. Cfr. Commissione europea, *Orientamenti sull'applicazione del diritto della concorrenza dell'Unione agli accordi collettivi concernenti le condizioni di lavoro dei lavoratori autonomi individuali*, Bruxelles, 29 settembre 2022 C(2022) 6846 final.

<sup>24</sup> Cfr. BELLAVISTA, *La questione del potere 'trasparente' nei rapporti di lavoro*, in *DML*, 2023, 577 ss.; NOVELLA, *Impresa*, in NOVELLA, TULLINI (a cura di), *Lavoro digitale*, Torino, 2022, 5 ss.; TEBANO, *Controlli digitali*, in BORELLI, BRINO, FALERI, LAZZERONI, TEBANO, ZAPPALÀ, *Lavoro e tecnologie*, Torino, 2022, 49; TULLINI, *La questione del potere nell'impresa. Una retrospettiva lunga mezzo secolo*, in *LD*, 2021, 429 ss.

Basti solo ricordare come sia proprio l'assenza di trasparenza, segnatamente dei sistemi classificatori e retributivi, e quindi la conseguente inaccessibilità delle cause delle differenze di trattamento da parte delle lavoratrici, a portare la Corte di giustizia, in *Danfoss*, a individuare a favore della ricorrente un onere di prova alleggerito ai fini dell'accertamento della fattispecie discriminatoria (in tal caso, retributiva di genere)<sup>25</sup>.

Da *Danfoss* in poi, le argomentazioni che la Corte sviluppa in punto di onere probatorio della discriminazione e, in particolare, di utilizzo del dato statistico evidenziano in più occasioni l'incidenza del principio di vicinanza della fonte probatoria e/o dell'oggettiva difficoltà di accesso a determinate informazioni da parte del ricorrente, ora per opacità dei criteri sottesi al trattamento ricevuto, ora per indisponibilità dei dati. Sono elementi che, in vario grado, investono inevitabilmente ogni vicenda discriminatoria e che giustificano, appunto, il *favor* accordato al ricorrente sotto il profilo probatorio, necessario per «non privare i lavoratori di qualsiasi strumento utile ad imporre il rispetto del principio della parità»<sup>26</sup>. L'incremento di opacità o l'inaccessibilità delle informazioni può intensificare l'interazione del principio di vicinanza delle fonti di prova e/o del canone interpretativo dell'effetto utile con il regime di *favor*, con conseguente ammissione di dati più sintetici o generali, ma parimenti significativi, nella costruzione della prova semi-piena. Pare comunque difficile, in generale, scindere l'accertamento della discriminazione da questioni attinenti alla trasparenza e/o alla accessibilità e prossimità delle fonti di prova. A essere necessariamente coinvolti nella comparazione, tra singoli soggetti o gruppi, sono, infatti, dati non riferibili soltanto alla sfera del ricorrente<sup>27</sup>, potenzialmente anche difficili da reperire, comunque più facilmente rientranti nella disponibilità del datore o afferenti alla sua sfera d'azione<sup>28</sup>; ciò è tanto più vero in un sistema di organizzazione digitale del lavoro caratterizzato dall'utilizzo di algoritmi e approcci di IA.

<sup>25</sup> C. giust., 17 ottobre 1989, causa C-109/88, *Danfoss*, ECLI:EU:C:1989:383.

<sup>26</sup> Cfr. C. giust., 31 maggio 1995, causa C-400/93, *Royal Copenhagen*, in *Racc.*, 1995, I-1275, punto 26; C. giust., 27 ottobre 1993, causa C-127/92, *Enderby*, in *Racc.*, 1993, I-5535. Il cons. n. 17 della dir. 97/80 spiega il regime probatorio speciale osservando che «le parti attrici in giudizio potrebbero essere private [altrimenti] di mezzi efficaci per far rispettare il principio della parità di trattamento».

<sup>27</sup> Al riguardo, si può segnalare che la dir. Ue 2023/970 sulla trasparenza retributiva prevede, all'art. 7, il diritto dei lavoratori di chiedere e ricevere per iscritto informazioni sul proprio livello retributivo individuale e sui livelli retributivi medi, suddivisi per sesso, relativi alle categorie di lavoratori che svolgono il loro stesso lavoro o un lavoro di pari valore rispetto al loro (par. 1). Si specifica, altresì, che ai lavoratori non deve essere impedito di rivelare la propria retribuzione ai fini dell'applicazione del principio di parità di retribuzione (par. 5). All'art. 20, si stabilisce che le autorità competenti o i tribunali nazionali debbano essere in grado di ordinare alla parte convenuta di rivelare qualsiasi prova pertinente che si trovi sotto il suo controllo, disponendo misure efficaci per la protezione di tali informazioni (v. *infra*).

<sup>28</sup> Per una valorizzazione del collegamento tra l'agevolazione e il principio di vicinanza della fonte probatoria, cfr. Cass., 26 febbraio 2021, n. 5476 in *LG*, 2021, 5, 551, secondo cui «nei giudizi antidiscriminatori i criteri di riparto dell'onere probatorio non seguono i canoni dell'art. 2729 c.c., essendo prevista un'agevolazione in favore del ricorrente, anche in ragione del principio di vicinanza della prova». In tal senso anche Trib. Roma, 12 marzo 2022; Trib. Roma, 23 marzo 2022. Rispetto all'operatività del principio di vicinanza della prova nell'accertamento di una discriminazione algoritmica, cfr. l'ordinanza Trib. Bologna, 31 dicembre 2020, cit., dove tuttavia, a parere di chi scrive, risulta criticabile l'affermazione secondo cui, per operatività del generale principio di vicinanza della fonte, sarebbe addossata al datore la prova del «concreto meccanismo di funzionamento dell'algoritmo». Si potrebbe, infatti, ritenere che ciò equivalga a spostare sul convenuto la prova del collega-

Se le considerazioni svolte sottolineano il legame inscindibile tra il problema della trasparenza e la questione della tutela antidiscriminatoria, il rapporto tra rischio di discriminazione e utilizzo di sistemi algoritmici è altrettanto stretto, come dimostrano i plurimi riferimenti a detto rischio tracciabili all'interno delle varie fonti regolative analizzate nel corso della trattazione o comunque il rilievo che tale rischio può assumere all'interno di esse<sup>29</sup>. In tale prospettiva, la repressione della discriminazione derivante dall'uso di un trattamento automatizzato di dati può seguire percorsi e basi normative diverse e parallele. È quanto si è potuto, ad esempio, osservare nella vicenda dell'algoritmo Frank di Deliveroo: i profili di illiceità sono stati affrontati sia dal Tribunale di Bologna utilizzando la strumentazione giuridica del diritto antidiscriminatorio, grazie all'esercizio dell'azione collettiva da parte del sindacato (ordinanza del 31 dicembre 2020), sia dal Garante della privacy, a partire da un'istruttoria avviata d'ufficio, al termine della quale è stata accertata la violazione dell'art. 22, par. 3, GDPR e dell'art. 47-*quinquies*, d.lgs. n. 81/2015 (ordinanza-ingiunzione del 22 luglio 2021).

Gli obblighi di trasparenza sui processi automatizzati costituiscono una garanzia imprescindibile per l'operatività degli strumenti del diritto antidiscriminatorio.

È sì vero che nel regime probatorio speciale ivi previsto può essere individuato un meccanismo di parziale inversione dell'onere a favore del ricorrente, chiamato a fornire soltanto una prova semi-piena del fatto<sup>30</sup>. Anche tale offerta di prova può, tuttavia, risultare ostacolata dalla difficoltà di risalire la trama del processo di calcolo e le ragioni del risultato prodotto dalla macchina: l'opacità algoritmica complica non solo il tracciamento del dato storico della disparità, ma anche l'individuazione del collegamento, diretto o indiretto, del trattamento subito con il fattore di rischio, per quanto, come detto, al ricorrente sia richiesta soltanto una prova semi-piena e, quindi, il grado di conferma probatoria richiesto, per questo secondo fatto, sia limitato, ossia tale da non escludere necessariamente la plausibilità di ipotesi opposte.

Vi è da dire che laddove si attivi il meccanismo di parziale inversione, si sposta sul convenuto il rischio di soccombenza e, quindi, anche il rischio derivante dalle correlazioni che un sistema di *machine learning* abbia posto alla base della regola di governo del processo "imparata". La modalità con cui tali correlazioni sono costruite può portare, infatti, il sistema ad associare all'obiettivo un fattore che, seppur "statisticamente efficace", non si

---

mento tra il criterio "neutro" e il trattamento subito dai lavoratori. Non si discute della possibilità per il convenuto di produrre una prova contraria che neghi la sussistenza di detto collegamento o dello svantaggio particolare dedotti dal ricorrente. Si ritiene problematico far gravare sul convenuto la prova del nesso in ragione del principio di vicinanza, senza contestualizzare questa distribuzione degli oneri nell'ambito di un regime probatorio che, pur alleggerito, deve comunque muovere da un'offerta di prova del ricorrente su tale fatto.

<sup>29</sup> Lo evidenzia anche Cons. Stato, sez. VI, 4 febbraio 2020, n. 881.

<sup>30</sup> Cfr. Cass., 28 marzo 2022, n. 9870, in *Leggi d'Italia*; Cass., 3 febbraio 2023, n. 3361, in *LG*, 2023, 5, 522. Si consenta sul punto il rinvio a PERUZZI, *Le discriminazioni*, in ANELLI, BRIGUGLIO, CHIZZINI, DE POLI, GRAGNOLI, ORLANDI, TOSI (a cura di), *L'onere della prova*, Padova, 2024, 951 ss.

pone di per sé in un rapporto di causa-effetto con l'esigenza lavorativa. Questo esito rende problematica la prova datoriale sia laddove volta a dimostrare la necessità e appropriatezza della misura in rapporto all'obiettivo e, quindi, la sussistenza di una giustificazione per la discriminazione indiretta; sia laddove funzionale a invocare le clausole di eccezione al divieto, con dimostrazione della natura essenziale e determinante, per lo svolgimento dell'attività lavorativa, della caratteristica correlata dalla macchina al risultato di interesse. La medesima problematica incide peraltro sulla possibilità del datore di dimostrare la conformità del processo ai divieti di indagini su «fatti non rilevanti ai fini della valutazione dell'attitudine professionale del lavoratore» ovvero su «dati personali dei lavoratori che non siano strettamente attinenti alle loro attitudini professionali e al loro inserimento lavorativo», stabiliti rispettivamente dall'art. 8 St. Lav. e, per le agenzie per il lavoro, dall'art. 10 d.lgs. 10 settembre 2003, n. 276<sup>31</sup>.

Tornando alle difficoltà che possono incidere sull'offerta di una prova, pur semi-piena, da parte del lavoratore e ribadita l'importanza, a superamento dell'opacità algoritmica, dei diritti di informazione sui processi automatizzati, si deve d'altra parte rilevare come gli stessi non bastino per assicurare l'effettività della tutela. Essi dovrebbero essere affiancati quantomeno dagli ulteriori strumenti disposti dalla direttiva 2023/970 in materia di trasparenza retributiva di genere, ossia la previsione di un'inversione completa dell'onere in caso di violazione degli obblighi di trasparenza datoriali, nonché la possibilità di un ordine di *disclosure* in giudizio (quest'ultimo previsto espressamente anche nella proposta di direttiva sul lavoro su piattaforma digitale, all'art. 21)<sup>32</sup>.

La stessa direttiva 2023/970 invita, inoltre, a non concentrarsi soltanto sulla soluzione delle difficoltà che si possono raccogliere, a valle, sul piano della prova della (in)sussistenza di una discriminazione, ma a spostare il ragionamento e la ricerca di una garanzia di tutela a monte, sul piano della prevenzione e dell'osservazione degli effetti derivanti dall'utilizzo di determinati criteri da parte del datore di lavoro. L'art. 10 prevede, nello specifico, una procedura di esame congiunto (*joint assessment*) con le rappresentanze dei lavoratori, laddove dal *reporting* sulle retribuzioni emerga un differenziale sentinella che il datore di lavoro non riesca a giustificare sulla base di elementi non collegabili (direttamente o indirettamente) al genere ovvero non abbia tempestivamente corretto. La direttiva specifica che la valutazione congiunta deve essere effettuata al fine di individuare, correggere e prevenire le differenze retributive tra uomini e donne, non giustificate sulla base di criteri oggettivi *gender-neutral*, deve essere messa a disposizione dei lavoratori e dei loro rappresentanti,

<sup>31</sup> Cfr. TULLINI, *Divieto di indagini sulle opinioni e trattamenti discriminatori*, in PEDRAZZOLI (coordinato da), *Il nuovo mercato del lavoro*, Zanichelli, 2004, 145 ss. Come già più volte segnalato, dette disposizioni sono espressamente richiamate dall'art. 113, d.lgs. 30 giugno 2003, n. 1963, rubricato «Raccolta di dati e pertinenza», e rientrano tra le «norme più specifiche» per l'ambito dei rapporti di lavoro, rilevanti ai fini dell'art. 88 GDPR.

<sup>32</sup> Per una più compiuta analisi del sistema di regole della direttiva, si rinvia a ZILLI, *La trasparenza nel lavoro subordinato. Principi e tecniche di tutela*, Pisa, 2022.

nonché dell'ispettorato del lavoro e dell'organismo di parità qualora lo richiedano. Inoltre, deve essere comunicata all'organismo di controllo designato dallo Stato membro per il monitoraggio sull'applicazione della normativa in questione. Si stabilisce, infine, che l'implementazione delle misure derivanti dalla valutazione congiunta debba avvenire in stretta collaborazione con i rappresentanti dei lavoratori e con possibilità che anche l'ispettorato del lavoro e/o l'organismo di parità siano chiamati a partecipare al processo.

Il modello proposto dalla direttiva per affrontare il rischio di discriminazione è tanto più rilevante nel contesto dei processi automatizzati, soprattutto laddove si tratti di sistemi di apprendimento automatico, associati all'uso di reti neurali profonde. In tal caso, infatti, è proprio l'osservazione dall'esterno del sistema e dei suoi *outcome* che permette di comprendere il suo funzionamento, nonché di individuare e mitigare i rischi che ne derivano. È peraltro un modello in grado di saldarsi con forza all'approccio *risk-based* che caratterizza la risposta regolativa all'IA e che connota l'architettura normativa implicata nella ricostruzione delle garanzie per il lavoratore di fronte all'uso di sistemi algoritmici. Basti ricordare la riflessione sviluppata nel corso della trattazione in merito agli obblighi di proceduralizzazione *ex ante* previsti nel contesto del GDPR, in particolare con riferimento alla funzione della DPIA in termini di verifica e prevenzione della lesione dei diritti fondamentali, tra cui quello di non discriminazione, derivanti dal trattamento dei dati personali. Ovvero anche alle garanzie di sorveglianza umana e di valutazione dell'impatto delle decisioni individuali, prese o sostenute dai sistemi automatizzati, previste dalla proposta di direttiva in materia di lavoro su piattaforma, in cui si evidenzia peraltro, come visto, una specifica attenzione per l'impatto sulla parità di trattamento e per il rischio di discriminazione.

La soluzione adottata dalla direttiva 2023/970 prevede, d'altra parte, una declinazione congiunta e partecipata del processo che i quadri normativi ora richiamati non raggiungono, per quanto in entrambi i casi, pur in diverso grado e soprattutto nel secondo, sia individuabile un ruolo per i rappresentanti dei lavoratori.

#### 4. ... e guardando oltreoceano.

Il tema del rischio di discriminazione collegato all'uso dei processi automatizzati sta assumendo sempre più rilievo nell'attuale sviluppo normativo, non solo a livello di Unione europea ma anche interno. A tal ultimo riguardo, basti evidenziare come la questione sia affrontata dal legislatore non soltanto nel citato e noto art. 1-*bis* del d.lgs. 152/97, con specifico riguardo alla tutela dei lavoratori (subordinati e autonomi<sup>33</sup>), ma anche nel con-

<sup>33</sup> Il comma 7 dell'art. 1-*bis* prevede che gli obblighi informativi gravano anche sul committente nell'ambito dei rapporti di lavoro di cui all'art. 409, n. 3, c.p.c. e di cui all'art. 2, comma 1, del d.lgs. 15 giugno 2015, n. 81. Ai sensi dell'art. 47-*ter* d.lgs. 81/2015, i lavoratori autonomi di cui all'art. 47-*bis* devono ricevere le informazioni di cui all'art. 1-*bis* in quanto compatibili.

testo del codice dei contratti pubblici, a favore degli operatori economici: all'art. 30, posta la possibilità per le stazioni appaltanti e gli enti concedenti di automatizzare le proprie attività ricorrendo a soluzioni tecnologiche, ivi incluse l'intelligenza artificiale, il d.lgs. 31 marzo 2023, n. 36 prevede che «le decisioni assunte mediante automazione rispettano i principi di [...] non discriminazione algoritmica, per cui il titolare mette in atto misure tecniche e organizzative adeguate al fine di impedire effetti discriminatori nei confronti degli operatori economici»; ancora, al comma quarto si stabilisce che «le stazioni appaltanti e gli enti concedenti adottano ogni misura tecnica e organizzativa atta a garantire che siano rettificati i fattori che comportano inesattezze dei dati e sia minimizzato il rischio di errori, nonché a impedire effetti discriminatori nei confronti di persone fisiche sulla base della nazionalità, dell'origine etnica, delle opinioni politiche, della religione, delle convinzioni personali, dell'appartenenza sindacale, dei caratteri somatici, dello status genetico, dello stato di salute, del genere o dell'orientamento sessuale».

Il tema sta entrando anche nella grammatica della contrattazione collettiva, come dimostra l'accordo stipulato il 19 febbraio 2024 da Assogrocery con Cgil, Cisl e Uil per la disciplina del lavoro su piattaforma digitale degli shopper. All'art. 5 si dispone che «al fine di permettere una verifica, da parte degli Shopper e del sindacato, del rispetto della normativa privacy e di evitare l'utilizzo di sistemi di profilazione e gestione automatizzata degli incarichi che discriminino i collaboratori, il committente è tenuto ad informare lo Shopper e le organizzazioni sindacali e le rappresentanze sindacali, dell'utilizzo di sistemi decisionali o di monitoraggio automatizzati deputati a fornire indicazioni rilevanti» per la gestione del rapporto.

Se la questione della discriminazione algoritmica, specie a fronte di un'evoluzione tecnologica sempre più accelerata, non potrà che richiedere nuove riflessioni e azioni normative, nonché la ricerca di adeguate misure di prevenzione, a livello di processi partecipati ma anche di tecniche algoritmiche di c.d. *Fair-XAI*, a conclusione dell'indagine che si è inteso qui sviluppare si consenta una breve proiezione oltreoceano del raggio d'analisi. Pare, infatti, utile e significativo raccogliere alcuni spunti di riflessione da un intervento normativo effettuato dalla Città di New York: la legge n. 144/2011, entrata in vigore il 6 luglio 2023. Tale legge permette ai datori di lavoro o all'agenzie di collocamento di utilizzare strumenti automatizzati per assumere decisioni sui candidati all'assunzione o a una promozione, basando la scelta (quantomeno in via preponderante) sui punteggi, classificazioni e raccomandazioni consegnate dalla macchina, a condizione che sia stata effettuata (e da non più di un anno) una verifica dei *bias* da parte di un *auditor* indipendente.

Se il processo automatizzato è usato per la selezione dei candidati, la verifica deve includere, almeno, il tasso di selezione e l'indice di impatto (ossia il rapporto con il tasso di selezione della categoria più selezionata) per le categorie di sesso (uomini e donne), per una serie di categorie di razza/etnia (ispanici o latini, bianchi, afroamericani, nativi hawaiani o delle isole del Pacifico, asiatici, nativi americani o dell'Alaska, due o più razze), nonché per categorie intersezionali di sesso, etnia e razza. La medesima operazione deve essere effettuata se il processo automatizzato è usato per assegnare un punteggio ai candidati,

con calcolo in questo caso del tasso di punteggio e dell'indice di impatto in rapporto al tasso di punteggio della categoria che ha ottenuto il punteggio più alto.

La normativa non prevede indici sentinella né specifiche conseguenze a fronte di determinati differenziali. Richiede “soltanto” ai datori di lavoro e alle agenzie di pubblicare una sintesi dei risultati sul proprio sito *web*, in modo chiaro e visibile, indicando il numero di individui, valutati dal processo, che non sono stati inclusi nei calcoli richiesti, perché rientranti in una categoria sconosciuta.

Si prevede, altresì, che i datori di lavoro o agenzie di collocamento (che utilizzino detti strumenti automatizzati) comunichino ai candidati residenti nella città di New York sia che verrà utilizzato lo strumento in questione, consentendo agli stessi di richiedere un processo di selezione alternativo o un accomodamento; sia le qualifiche e le caratteristiche del lavoro che tale strumento utilizzerà nella valutazione.

La disciplina descritta non può che far riflettere, necessitando peraltro della doverosa premessa che le categorie indicate sono quelle previste nel c.d. «*EEO-1 Component 1*» report, una relazione annuale sui dati demografici della forza lavoro, richiesta dalla US Equal Employment Opportunity Commission (EEOC) ai datori di lavoro con almeno 100 dipendenti o agli appaltatori federali con almeno 50 dipendenti, al ricorrere di determinati criteri<sup>34</sup>.

Ci si può d'altra parte domandare, riportando lo sguardo nel contesto europeo, quanto possano incidere sull'utilizzo di sistemi automatizzati in fase di *recruitment* o di promozione alcuni obblighi stabiliti dalla più volte citata direttiva 2023/970. Si pensi, in particolare, all'obbligo dei datori di lavoro di provvedere affinché le procedure di assunzione siano condotte in modo non discriminatorio, così da non compromettere il diritto alla parità di retribuzione per uno stesso lavoro o per un lavoro di pari valore (art. 5, par. 3). Ovvero all'obbligo dei datori di lavoro di adottare, per la progressione economica, criteri oggettivi e neutri sotto il profilo del genere, rendendoli facilmente accessibili ai propri lavoratori (art. 6, par. 1).

E se la soluzione fosse quella prevista dalla *New York City Local Law 144*?

<sup>34</sup> Si rinvia, sul punto, al *2023 EEO-1 Component 1 Data Collection Instruction Booklet*.

