



**LaBoUR & Law Issues**  
Rights | Identity | Rules | Equality

## **Il diritto antidiscriminatorio al test di intelligenza artificiale**

**MARCO PERUZZI**  
Università di Verona

**vol. 7, no. 1, 2021**

ISSN: 2421-2695



# Il diritto antidiscriminatorio al test di intelligenza artificiale

**MARCO PERUZZI**

Università di Verona

Associato di Diritto del Lavoro

marco.peruzzi@univr.it

---

## ABSTRACT

---

Moving from the analysis of the case concerning the algorithm Frank, as adjudicated by the Tribunal of Bologna, the essay points out the complex challenges that the application of antidiscrimination law faces when the systems of Artificial Intelligence are based on machine-learning approaches. In particular, the problems arising as a consequence of algorithmic opacity, namely the so called «black box» effect, are highlighted. First, the topic of transparency is examined as it is addressed by multiple regulatory fields, with regard to Data, Computer Programs and Trade Secrets Protection. Then, the investigation focuses on the reconstruction of possible solutions aimed at straightening and ensuring the antidiscrimination protection. The reflection builds on the assessment of the measures contained in the proposal for a Regulation on Artificial Intelligence and of the techniques than could be borrowed from the proposal of Directive on pay transparency in the context of gender equality, both presented by the European Commission in 2021.

**Keywords:** artificial intelligence; machine learning; algorithms; black box; antidiscrimination law; transparency; burden of proof.

<https://doi.org/10.6092/issn.2421-2695/13117>

---

## Il diritto antidiscriminatorio al test di intelligenza artificiale

SOMMARIO: 1. 1. Gli anticorpi del diritto antidiscriminatorio rispondono alla sfida dell’algoritmo Frank. – 2. Il diritto antidiscriminatorio di fronte alle sfide del *machine learning*. – 3. Trasparenza, decisioni automatizzate e discriminazioni nella rete dei livelli normativi. – 4. Ragionando di possibili prospettive di tutela, a partire dalla proposta di Regolamento sull’IA e dalla proposta di direttiva su trasparenza e parità retributiva di genere.

### 1. Gli anticorpi del diritto antidiscriminatorio rispondono alla sfida dell’algoritmo Frank

Il caso deciso dal Tribunale di Bologna con l’ordinanza del 31 dicembre 2020, relativo all’utilizzo dell’algoritmo Frank da parte di Deliveroo, si inserisce nel dibattito sull’applicabilità, adeguatezza ed efficacia delle categorie e degli strumenti del diritto antidiscriminatorio di matrice europea di fronte alle sfide poste dai processi decisionali digitalizzati e automatizzati<sup>(1)</sup>.

La vicenda, in sintesi, verteva sull’uso di un «algoritmo di assegnazione», che partendo dallo storico delle sessioni di lavoro prenotate e non cancellate dal rider nel termine previsto dal contratto, attribuiva al lavoratore un punteggio, che condizionava poi le sue chance di inserire nel sistema ulteriori prenotazioni,

---

<sup>(1)</sup> Cfr. B. Caruso - L. Zappalà, *Un diritto del lavoro “tridimensionale”: valori e tecniche di fronte ai mutamenti dei luoghi di lavoro*, WP CSDLE, It. n. 439/2021; P. Tullini, *La salvaguardia dei diritti fondamentali della persona che lavora nella gig-economy*, *costituzionalismo.it*, 2020, 1, 39 ss.; J. Adams-Prassl, *What If Your Boss Was an Algorithm? Economic Incentives, Legal Challenges, and the Rise of Artificial Intelligence at Work*, *CLL&PJ*, 2019, 41, 1, 123 ss.; L. Giacomelli, *Big brother is «gendering» you. Il diritto antidiscriminatorio alla prova dell’intelligenza artificiale: quale tutela per il corpo digitale?*, *BioLaw Journal*, 2019, 269 ss.; G. Centamore - L. Ratti, *Oltre il dilemma qualificatorio: potenzialità e limiti del diritto antidiscriminatorio nella protezione del lavoratore on-demand*, in C. Alessi, M. Barbera, L. Guaglianone (a cura di), *Impresa, lavoro e non lavoro nell’economia digitale*, Cacucci, 2019, 663 ss.; M. Kullman, *Platform Work, Algorithm Decision-Making, and EU Gender Equality Law*, *IJLLIR*, 2018, 34, 11 ss.; P. Hacker, *Teaching fairness to artificial intelligence: existing and novel strategies against discrimination under EU law*, *CMLR*, 2018, 1143 ss.; V. Maio, *Il diritto del lavoro e le nuove sfide della rivoluzione robotica*, *ADL*, 2018, 6, 1414 ss. Per un’analisi dell’ordinanza si rinvia ai contributi in questo numero della rivista nonché, *ex multis*, a R. Santagata De Castro, *Anti-discrimination Law in the Italian Courts: the new frontiers of the topic in the age of algorithms*, WP CSDLE, It. n. 440/2021; M. V. Ballestrero, *Ancora sui rider. La cecità discriminatoria della piattaforma*, *Labor*, 2021, 1, 104 ss.; A. Perulli, *La discriminazione algoritmica: brevi note introduttive a margine dell’Ordinanza del Tribunale di Bologna*, *LDE*, 2021, n. 1; M. Faioli, *Discriminazioni digitali e tutela giudiziaria su iniziativa delle organizzazioni sindacali*, *DRI*, 2021, 204 ss.

determinando così una possibile riduzione delle sue occasioni di accesso al lavoro.

L'ordinanza rileva il carattere discriminatorio della condotta datoriale: da un lato, il sistema adottato dava «applicazione ad una disposizione apparentemente neutra (la normativa contrattuale sulla cancellazione anticipata delle sessioni prenotate)» idonea a mettere «una determinata categoria di lavoratori (quelli partecipanti ad iniziative sindacali di astensione del lavoro) in una posizione di potenziale particolare svantaggio»; dall'altro, la società convenuta non ha adeguatamente assolto l'onere probatorio relativo ai fatti impeditivi della fattispecie, ossia alla giustificazione del trattamento, non dimostrando la sussistenza di una finalità legittima e il carattere di appropriatezza e necessità dei mezzi impiegati per conseguirla, ai sensi dell'art. 3, co. 6, d.lgs. n. 216/03.

Nel caso *de quo*, le fonti del diritto antidiscriminatorio si dimostrano idonee a consentire l'accertamento della fattispecie illecita in un contesto organizzativo caratterizzato dall'uso di algoritmi. Le ragioni sono plurime.

Anzitutto, sono fonti applicabili a prescindere dalla soluzione data a una tipica questione connessa alla digitalizzazione, soprattutto nel contesto delle piattaforme di *delivery*: quella della qualificazione giuridica del rapporto. Non soltanto, infatti, l'ambito di tutte le direttive europee in materia include l'accesso al lavoro autonomo<sup>(2)</sup> e, pur solo con riguardo al genere, la dir. 2010/41/UE estende il principio di parità a soggetti che esercitano un'attività autonoma<sup>(3)</sup>. L'art. 47-quinquies, d.lgs. n. 81/2015 dispone, inoltre, che – anche laddove non operi l'estensione dello statuto protettivo del lavoro subordinato *ex art. 2, co. 1* – ai lavoratori autonomi che svolgono attività di consegna di beni per conto altrui, in ambito urbano e con l'ausilio di velocipedi o veicoli a motore, attraverso piattaforme anche digitali, «si applicano la disciplina antidiscriminatoria e quella

---

(2) Cfr. art. 14, par. 1, lett. a), dir. 2006/54/CE; art. 3, par. 1, lett. a), dir. 2000/43/CE; art. 3, par. 1, lett. a), dir. 2000/78/CE. Cfr. B. Caruso - L. Zappalà, *op. cit.* Sull'estensione della nozione di lavoratore dipendente nel diritto UE cfr. *ex multis* G. Pacella, *La nozione euro-unitaria di lavoratore dipendente alla prova della gig-economy: si pronuncia la Corte di giustizia europea*, LLI, 2020, 1, R.17.

(3) Come evidenzia C. Alessi (*Lavoro tramite piattaforma e divieti di discriminazione nell'UE*, in C. Alessi - M. Barbera - L. Guaglianone (a cura di), *Impresa, lavoro e non lavoro nell'economia digitale*, *op. cit.*, 695), gli artt. 3 e 4 della dir. 2010/41/UE «contengono, rispettivamente, le definizioni di discriminazione diretta e indiretta e di molestie, nonché l'affermazione del principio di parità di trattamento “ad esempio per quanto riguarda la creazione, la fornitura di attrezzature o l'ampliamento di un'impresa o l'avvio o l'ampliamento di ogni forma di attività autonoma”»; «si tratta di una previsione molto ampia, che senza dubbio si può considerare applicabile anche alle prestazioni di lavoro (genuinamente) autonome svolte dai lavoratori tramite piattaforma».

a tutela della libertà e dignità del lavoratore previste per i lavoratori subordinati, ivi compreso l'accesso alla piattaforma»<sup>(4)</sup>.

In secondo luogo, la traduzione della regola in una ricetta digitalizzata e della decisione in un processo automatizzato così come l'«incoscienza» e «cecità» del programma informatico sono dati che non incidono sulla costruzione della fattispecie normativa della discriminazione.

Non rileva il tipo di condotta, ma la sua riconducibilità al datore di lavoro<sup>(5)</sup>, oltre che la sua attinenza, in termini non meramente ipotetici<sup>(6)</sup>, all'ambito di applicazione della normativa, nel caso *de quo* le «condizioni di accesso all'occupazione e al lavoro, sia dipendente che autonomo» (art. 3, par. 1, dir. 2000/78/CE).

Non rileva parimenti l'apparente neutralità delle istruzioni eseguite dal sistema digitale, ma l'impatto differenziato che provocano. Soprattutto, è l'effetto prodotto, nella sua oggettività, a consentire il perfezionamento della fattispecie e non l'intenzionalità del soggetto agente<sup>(7)</sup>.

Si segnala, infine, come la vicenda consenta di valorizzare due ulteriori elementi del quadro normativo antidiscriminatorio che è utile ricordare in vista del ragionamento che sarà sviluppato nel prosieguo della trattazione.

---

(4) La medesima disposizione, al secondo comma, prevede inoltre che «l'esclusione dalla piattaforma e le riduzioni delle occasioni di lavoro ascrivibili alla mancata accettazione della prestazione sono vietate».

(5) Come osserva A. Rota (*Rapporto di lavoro e big data analytics: profili critici e risposte possibili*, LLI, 2017, I.42), «nella disciplina anti-discriminatoria [...] non si configurano aree d'impunità per [il datore di lavoro] al verificarsi di comportamenti ad esempio posti in essere dai suoi sottoposti o [...] da macchine intelligenti. Tale logica pare non troppo dissimile da quella già accolta da alcune norme del codice civile», quali l'art. 2049 c.c. o l'art. 2050 c.c.

(6) Sul punto, cfr. Corte giust. 23 aprile 2020, C-507/18, *NH c. Associazione Avvocatura per i diritti LGBTI – Rete Lenford*.

(7) La faticosa acquisizione della nozione oggettiva di discriminazione da parte della giurisprudenza italiana, nonostante la sua affermazione a livello europeo già nella sentenza *Dekker* della Corte di giustizia (8 novembre 1990, C-177/88) e a livello normativo interno già nell'art. 4 della l. 125/91, si spiega a fronte del persistente utilizzo del modello civilistico della nullità per motivo illecito *ex* art. 1345 c.c. nell'interpretazione della fattispecie del licenziamento discriminatorio, al fine di estendere le tutele normativamente previste per quest'ultimo anche al licenziamento ritorsivo. Come spiega M. V. Ballestrero (*Sul licenziamento discriminatorio*, GDLRI, 2016, 248), «animata dalla buona intenzione di allargare la protezione antidiscriminatoria, questa giurisprudenza finisce per restringerla, facendo gravare per intero sul lavoratore l'onere di provare l'intento discriminatorio». Ha contribuito al superamento di questa lettura la formulazione dell'art. 18 St. Lav. derivante dagli interventi di modifica della cd. riforma Fornero. Non è un caso che sia successiva, del 2016, la pronuncia con cui la Cassazione ha definitivamente chiarito la natura oggettiva della discriminazione, prendendo le distanze dall'inquadramento di matrice civilistica (Cass. 5 aprile 2016, n. 6575).

Per quanto, nel caso in esame, l'identificazione di singoli lavoratori svantaggiati (ossia che erano stati o potevano essere penalizzati dall'esercizio dello sciopero) fosse possibile<sup>(8)</sup>, l'azione di contrasto – proposta da tre organizzazioni sindacali ai sensi dell'art. 5, co. 2, d.lgs. n. 216/03 – è agevolata dal fatto che l'accertamento della discriminazione e del danno non richiede necessariamente che le persone lese siano individuabili in modo diretto e immediato, riconoscendosi comunque in tale ipotesi agli enti esponenziali la titolarità di un'azione collettiva.

Ancora, l'accertamento non richiede la presenza di un termine di confronto reale, bastando – come nel caso *de quo*, relativo a una discriminazione indiretta – che la regola *possa* implicare un particolare svantaggio per i portatori di un fattore di rischio, ovvero potendosi effettuare la comparazione con un soggetto ipotetico, per quanto attiene alla discriminazione diretta.

Se nel perimetro circoscritto della decisione in esame il diritto antidiscriminatorio dimostra di possedere tutti gli anticorpi necessari per l'accertamento di una fattispecie mediata e/o veicolata dallo strumento tecnologico, la vicenda non appare, tuttavia, davvero indicativa delle più complesse problematiche implicate dall'utilizzo di sistemi di intelligenza artificiale (d'ora in avanti IA) nel lavoro.

Si premette che, ai fini dell'analisi che ci occupa, l'insieme di tali sistemi sarà individuato facendo riferimento all'ampia nozione adottata dalla recente proposta di Regolamento sull'IA, presentata dalla Commissione europea il 21 aprile 2021<sup>(9)</sup>, comprensiva non soltanto degli approcci e delle tecniche di

---

<sup>(8)</sup> Come riporta un dirigente sindacale in sede di deposizione, nelle occasioni in cui le organizzazioni sindacali avevano proposto, come modalità di protesta, quella dello sciopero, i riders avevano «rappresentato che scioperare avrebbe comportato non solo la perdita della retribuzione per quella giornata, ma anche una forte penalizzazione nella attività lavorativa futura», spiegando che «la problematica nasceva da quello che loro chiamano ranking reputazionale e dal sistema che c'è dietro».

<sup>(9)</sup> Commissione europea, *Proposta di Regolamento che stabilisce regole armonizzate sull'Intelligenza Artificiale (Legge sull'Intelligenza Artificiale)*, Brussels, 21.4.2021 COM(2021) 206 final. La nozione è prevista dall'annex 1 e si dispone all'art. 4 Reg. che la Commissione possa adottare atti delegati per modificarne il contenuto a fronte dell'evoluzione tecnologica e del mercato. Segnatamente, sono inclusi all'interno della nozione di AI: «a) approcci di apprendimento automatico, compresi l'apprendimento supervisionato, l'apprendimento non supervisionato e l'apprendimento per rinforzo, con utilizzo di un'ampia gamma di metodi, tra cui l'apprendimento profondo (*deep learning*); b) approcci basati sulla logica e approcci basati sulla conoscenza, compresi la rappresentazione della conoscenza, la programmazione induttiva (logica), le basi di conoscenze, i motori inferenziali e deduttivi, il ragionamento (simbolico) e i sistemi esperti; c) approcci statistici, stima bayesiana, metodi di ricerca e ottimizzazione».

apprendimento automatico, ma anche, ad esempio, dei modelli basati sul ragionamento meccanico<sup>(10)</sup>.

Per quanto riguarda il caso Deliveroo, è sì vero che la società presenta, sul proprio sito, l'algoritmo Frank come «una potente tecnologia predittiva che gestisce nel modo più efficiente l'assegnazione degli ordini in base alla posizione dei ristoranti, dei rider e dei clienti», aggiungendo che «l'apprendimento automatico aiuta a stimare il tempo necessario per la preparazione di un ordine, aiutando a ottimizzare l'esperienza complessiva». Ed è altrettanto vero che il giudice di Bologna osserva che «la mancata allegazione e prova, da parte della società resistente, del concreto meccanismo di funzionamento dell'algoritmo che elabora le statistiche dei rider preclude in radice una più approfondita disamina della questione» e «la società [...] non ha mai chiarito quali specifici criteri di calcolo vengano adottati per determinare le statistiche di ciascun rider, né tali specifici criteri vengono pubblicizzati sulla piattaforma».

Nell'ordinanza, tuttavia, il giudice reputa comunque il quadro istruttorio idoneo a fondare il suo convincimento e, in particolare, a permettere l'individuazione della logica seguita dall'algoritmo nella selezione delle priorità di accesso al sistema di prenotazione: da un lato, il contratto-tipo depositato da Deliveroo prevedeva espressamente che la cancellazione tardiva potesse costituire elemento di preferenza per la prenotazione di sessioni successive; dall'altro, la prova testimoniale aveva confermato l'incidenza di tale dato sul *ranking* del rider.

Al di là della riflessione in punto di onere della prova che questo passaggio impone e che sarà approfondita oltre nella trattazione, è qui interesse evidenziare come nel giudizio, ai fini dell'accertamento della discriminazione, sia isolato e analizzato, all'interno del più ampio processo di funzionamento del sistema digitale utilizzato da Deliveroo, uno specifico segmento, quello che

---

<sup>(10)</sup> Cfr. Gruppo indipendente di esperti ad alto livello sull'intelligenza artificiale, *Una definizione di IA: principali capacità e discipline*, Bruxelles, 2019. Anche nelle Proposte per una Strategia italiana per l'intelligenza artificiale, elaborate dal Gruppo di Esperti MISE e pubblicate dal Ministero dello sviluppo economico il 2 luglio 2020 ([www.mise.gov.it](http://www.mise.gov.it)), si individua con il termine IA «una famiglia di tecnologie che spaziano dalla rappresentazione della conoscenza al ragionamento automatico che sottende aree quali la pianificazione e il supporto decisionale, fino alla percezione e all'apprendimento automatico. [...] È importante, dunque, soffermarsi sul fatto che le tecniche di AI vanno ben oltre quelle più tipicamente evocate (come il *machine learning*) e includono tecniche [...] che permettono di sviluppare sistemi pienamente trasparenti» (11). La nozione adottata in questi documenti ha un perimetro più ampio di quello tracciato da parte della letteratura in materia. H. Schildt, ad esempio, riporta che «*as a technology category, AI is used to capture machine learning algorithms or some subset of them*» (*The Data Imperative*, OUP, 2020, 134).

gestisce la sequenza decisionale che parte dal computo delle passate cancellazioni tardive e arriva a stabilire lo spettro delle occasioni di lavoro future. Quantomeno rispetto a questo segmento, l'unico dedotto e rilevante per il riconoscimento della fattispecie illecita, si potrebbe affermare che il sistema operi sulla base di un algoritmo deterministico, ossia di una successione informatizzata di istruzioni che, a parità di dati d'ingresso, produce un solo e unico output. Il processo decisionale segue un ragionamento meccanico e un approccio basato sulla logica.

In tale contesto, il quadro normativo antidiscriminatorio trova agile applicazione dal momento che la regola codificata dall'algoritmo ed eseguita dal sistema digitale è ricostruibile, così come sono individuabili il nesso di causa-effetto che collega tale regola alla disparità di trattamento e il suo impatto differenziato, valutabile anche a prescindere dalla diretta identificazione delle vittime. Fondamentalmente, lo scenario non è così distante da quello che si potrebbe configurare, ad esempio, 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. La particolarità qui risiede nella digitalizzazione della regola all'interno di un algoritmo e nella costruzione di un processo decisionale automatizzato a partire dalla sua esecuzione.

La valutazione di adeguatezza degli strumenti offerti dal quadro normativo antidiscriminatorio diventa ben più problematica laddove il modello di IA adotti le tecniche e gli approcci di apprendimento automatico, soprattutto in caso di *deep learning* e utilizzo di reti neurali profonde, stratificate.

## **2. Il diritto antidiscriminatorio di fronte alle sfide del *machine learning***

Al di là del rischio, ampiamente evidenziato nella riflessione scientifica, relativo alla riproducibilità rigida ed esponenziale dei pregiudizi e stereotipi insiti nei big data usati per addestrare e alimentare il sistema<sup>(11)</sup>, il tratto distintivo del

---

<sup>(11)</sup> Si rinvia sul punto a E. Dagnino, *People Analytics: lavoro e tutele al tempo del management tramite big data*, *LLI*, 2017, 3, I.2. ss.; L. Giacomelli, *op. cit.*; A. Rota, *op. cit.* Nelle Proposte per una Strategia italiana per l'intelligenza artificiale si evidenzia che «l'intelligenza artificiale non è intelligente *stricto sensu*. Se utilizzata in modo stupido, l'AI riproduce e spesso amplifica la stupidità. [...] un utilizzo maldestro dell'AI può amplificare forme di *bias* e discriminazione a livello sociale, specialmente se utilizzato in modo autonomo o se basato su *dataset* non rappresentativi; peraltro, dato che la nostra società è già caratterizzata da *bias* e



*machine learning* rispetto alla situazione configurata nel caso Deliveroo si registra soprattutto nella possibile opacità del criterio discretivo utilizzato dalla macchina per raggiungere il risultato.

Tale problema, che trova sintesi nell'espressione “*black box*” e a cui cerca di rispondere il ramo fiorente del cd. “*explainable AI*”, è dovuto, come noto, al meccanismo che alcuni sistemi basati sul metodo dell'apprendimento automatico (soprattutto il *deep learning*) seguono per consegnare l'output più efficace per gli obiettivi impartiti. Il processo muove dalla ricerca, all'interno dei dati forniti, di una regolarità statistica, una correlazione, dei pattern, dei modelli di ricorrenza, che consentano di estrarre informazioni predittive funzionali all'effetto richiesto. È «un processo analitico iterativo»<sup>(12)</sup>, che permette al sistema di progredire, generare in un ciclo continuo nuovi modelli sulla base dei dati di feedback, così da adeguare, modificare, perfezionare in modo autonomo le proprie azioni, in termini non sempre prevedibili, ma sempre funzionali. L'opacità che si riscontra in questi sistemi si riconnette alla difficoltà di ricostruire, conoscere, interpretare le associazioni, le correlazioni e i collegamenti probabilistici che, in quanto «statisticamente efficaci», si sono creati nelle reti neurali<sup>(13)</sup>. Ne consegue la difficoltà di comprendere il processo generativo della decisione.

In un iter d'indagine volto a verificare l'adeguatezza del diritto antidiscriminatorio rispetto a tali sistemi di IA, è interessante osservare come le categorie dell'associazione e della correlazione siano ben conosciute e consolidate al suo interno. Si pensi alla fattispecie della discriminazione associata, in cui a rilevare è la connessione oggettiva con il fattore, a prescindere dal fatto che la vittima ne sia portatrice<sup>(14)</sup>; alla discriminazione indiretta, dove la correlazione con il fattore si costruisce sulla base dell'impatto differenziato di una regola o criterio di applicazione generale; alla discriminazione intersezionale, dove la disparità di trattamento consegue alla combinazione di più fattori<sup>(15)</sup>;

---

diseguaglianza, utilizzare anche in modo corretto dati passati può portare a forme di discriminazione sempre più evidenti nel corso del tempo» (16-17).

<sup>(12)</sup> G. Comandè, *Intelligenza artificiale e responsabilità tra liability e accountability. Il carattere trasformativo dell'IA e il problema della responsabilità*, *Analisi giuridica dell'economia*, 2019, 172 ss.

<sup>(13)</sup> G. Sanguinetti, *Machine Learning: accuratezza, interpretabilità e incertezza*, *Ithaca: Viaggio nella Scienza*, 2020, 16, 74; U. Ruffolo, *Intelligenza artificiale, machine learning e responsabilità da algoritmo*, *GI*, 2019, 7, 1657 ss.

<sup>(14)</sup> Cfr. Corte giust. 17 luglio 2008, C-303/06, *Coleman*; Corte giust. 16 luglio 2015, C-83/14, *Chez*.

<sup>(15)</sup> Cfr. M. Militello - D. Strazzari, *I fattori di discriminazione*, in M. Barbera - A. Guariso (a cura di), *La tutela antidiscriminatoria*, Giappichelli, 2019, 85 ss.

all'eccezione del requisito essenziale, dove a rilevare è una disparità basata su caratteristiche collegate a un fattore di rischio (ad es. l'utilizzo di un determinato indumento rispetto alla religione).

La presenza di queste categorie non pare, tuttavia, fornire al quadro normativo anticorpi efficaci per rispondere ai rischi della *black box*.

Anzitutto, l'opacità del processo generativo della decisione rende più difficile l'individuazione della disparità di trattamento in assenza di vittime direttamente identificate o identificabili e in assenza di un termine di raffronto reale che palesi il fatto storico della disparità. Le stesse vittime saranno probabilmente inconsapevoli del trattamento subito, di esser parte di un gruppo penalizzato, quantomeno nella prima fase di maturazione della fattispecie discriminatoria. Il tracciamento dell'esclusione o dello svantaggio, necessitando di un confronto tra gruppi o soggetti in posizione analoga e/o di una misurabilità dei differenziali, potrebbe essere tanto più difficile in contesti lavorativi in cui proprio l'utilizzo di sistemi di IA, soprattutto se incorporati in entità robotiche, ha comportato una flessione del coinvolgimento umano e una sostituzione del lavoro, con impatto sul numero di dipendenti, ovvero riduzione e polarizzazione delle professionalità impiegate. O ancora una frammentazione dei compiti, una soggettivizzazione dei rapporti e delle regole, una morfologia reticolare e de-gerarchizzata dell'impresa<sup>(16)</sup>.

In secondo luogo, le combinazioni costruite nel contesto di funzionamento degli algoritmi di apprendimento automatico possono essere imprevedibili. La letteratura riporta gli esempi di un algoritmo che associ una maggior capacità di programmazione informatica alla frequentazione di un determinato website di manga giapponesi, con conseguente 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>(17)</sup>.

La presenza di abbinamenti inattesi, se non inconoscibili, rende più complesso l'accertamento della fattispecie discriminatoria in giudizio, incidendo negativamente sull'offerta di prova sia del lavoratore, sia del datore.

---

<sup>(16)</sup> Cfr. B. Caruso - L. Zappalà, *op. cit.*; A. Perulli, *La "soggettivazione regolativa" nel diritto del lavoro*, *DRI*, 2019, 1, 111 ss.; R. Del Punta, *Innovazioni tecnologiche e diritto del lavoro*, *Rivista degli infortuni e delle malattie professionali*, 2019, 106, 2/3, 261 ss.; A. Salento, *Digitalizzazione delle imprese e trasformazione delle competenze. Quadro analitico e riscontri empirici*, *Labor*, 2019, 2, 131 ss.; G. Bolego, *Intelligenza artificiale e regolazione delle relazioni di lavoro: prime riflessioni*, *Labor*, 2019, 1, 51 ss.

<sup>(17)</sup> D. Peck, *They're watching you at work*, *theatlantic.com*, dec. 2013; A. G. King - M. Mrkonich, *"Big Data" and the Risk of Employment Discrimination*, *Oklahoma Law Review*, 2016, 68, 3, 557 ss.; E. Dagnino, *People Analytics*, *cit.*

Va premesso che nel regime probatorio speciale previsto dal diritto antidiscriminatorio può essere individuato un meccanismo di parziale inversione dell'onere, che si attiva al raggiungimento da parte del ricorrente di una prova semi-piena: prova di una mera porzione della fattispecie, costituita dal dato storico della disparità rispetto a un soggetto comparabile reale, laddove presente (v. *infra* il caso *Danfoss*)<sup>(18)</sup>; prova di probabilità attenuata del nesso intercorrente tra il trattamento subito e il fattore di rischio, in caso di comparazione ipotetica o di discriminazione indiretta<sup>(19)</sup>.

E' sì vero che tale alleggerimento dell'onere consente al ricorrente di non confrontarsi con «l'impossibilità di controbattere a una decisione – quella del software algoritmico – che si fonda su dati statistici con altrettanti dati statistici che comprovino l'ingiustizia dei suoi risultati»<sup>(20)</sup>.

La difficoltà di risalire la trama del processo di calcolo e le ragioni del risultato prodotto riesce comunque a ostacolarne la prova, rendendo complicato 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 il grado di conferma probatoria richiesto, per questo secondo fatto, sia limitato, ossia tale da non escludere necessariamente la plausibilità di ipotesi opposte.

L'utilizzo delle correlazioni può operare, d'altra parte, a svantaggio anche dell'eventuale prova datoriale relativa ai fatti impeditivi della discriminazione. 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 pone in 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<sup>(21)</sup>.

---

<sup>(18)</sup> Si potrebbe, d'altra parte, riflettere sul fatto che l'eventuale utilizzo di una prova statistica che proietti la disparità su una dimensione collettiva e di sistematicità possa proiettare altresì l'oggetto della prova verso il nesso di causalità, oltre il perimetro del mero dato storico della differenza di trattamento.

<sup>(19)</sup> Si consenta il rinvio a M. Peruzzi, *La prova del licenziamento ingiustificato e discriminatorio*, Giappichelli, 2017.

<sup>(20)</sup> L. Giacomelli, *op. cit.*, 283.

<sup>(21)</sup> Ad es., art. 3, co. 3, d. lgs. n. 216/03. Sul passaggio dall'approccio della causalità a quello delle correlazioni, cfr. A.G. King - M. Mrkonich, *op. cit.*; A. Donini, *Tecniche avanzate di analisi dei dati e protezione dei lavoratori*, DRI, 2018, 222 ss.; E. Dagnino, *Dalla fisica all'algoritmo*:

Lo svantaggio sull'offerta di prova non incide in egual misura per le due parti sull'esito del giudizio, considerato che, pur in presenza di un regime alleggerito per il ricorrente, è comunque necessario che lo stesso fornisca una prova semi-piena della fattispecie affinché si attivi il meccanismo di parziale inversione dell'onere e la regola di giudizio versi a sfavore del convenuto<sup>(22)</sup>.

Si osservi come, per diretta conseguenza delle ragioni indicate in premessa, nel caso dell'ordinanza del Tribunale di Bologna l'applicazione del regime speciale probatorio non incontri i problemi ora citati e riscontrabili, invece, nel contesto di sistemi di IA ad apprendimento automatico caratterizzati dal fenomeno della *black box*.

Il giudice ivi ritiene «che siano emersi elementi sufficienti a far presumere, sia pure nei limiti della *sommatoria cognitio* che caratterizza il [...] procedimento, la discriminatorietà del sistema di accesso alle fasce di prenotazione delle sessioni di lavoro adottato da Deliveroo». Il sistema risulta, infatti, ricostruibile «sulla scorta delle allegazioni delle parti e dalle testimonianze assunte, che appaiono su molti aspetti sostanzialmente concordi»: la lettura del contratto-tipo sottoscritto dal rider evidenzia la rilevanza delle cancellazioni tardive e l'incidenza degli indici di affidabilità e partecipazione sull'accesso al sistema di prenotazione; dalle allegazioni anche di parte resistente, si può evincere come il valore dell'indice di affidabilità fosse determinato «dal numero delle occasioni in cui il rider, pur avendo prenotato una sessione, non ha partecipato, dove “partecipare” significa loggarsi entro i primi 15 minuti dall'inizio della sessione». L'ordinanza continua affermando come sia «pacifico, infatti, ed è stato confermato dai testi che la mancata partecipazione del rider alla sessione prenotata – e non previamente cancellata – incide negativamente sulle statistiche del rider e in particolare sul parametro della affidabilità». «Ed invero, è la stessa società convenuta a riferire che il rider può cancellare liberamente la sessione prenotata fino al momento del suo inizio e può “non

---

*una prospettiva di analisi giuslavoristica*, AUP, 2019. È importante segnalare come la medesima problematica incida 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. n. 276/2003. Cfr. P. Tullini, *Divieto di indagini sulle opinioni e trattamenti discriminatori*, in M. Pedrazzoli (coordinato da), *Il nuovo mercato del lavoro*, Zanichelli, 2004, 145 ss. Si segnala che dette disposizioni sono espressamente richiamate dalla normativa in tema di privacy all'art. 113, d.lgs. n. 196/03, rubricato «Raccolta di dati e pertinenza».

<sup>(22)</sup> Cfr. G. Gaudio, *Algorithmic management, poteri datoriali e oneri della prova: alla ricerca della verità materiale che si cela dietro l'algoritmo*, LLI, 2020, 2, 21 ss.

loggarsi fino a 14 minuti e 59 secondi dopo l'inizio della sessione” senza che ciò impatti negativamente sulle sue statistiche, con ciò implicitamente ma chiaramente ammettendo che, al contrario, il mancato log-in oltre i 15 minuti dall'inizio della sessione prenotata incide negativamente sul parametro della affidabilità. In questo senso anche le concordi dichiarazioni dei testi». In conclusione, per il Tribunale di Bologna «appare provato che l'adesione ad una iniziativa di astensione collettiva dal lavoro è idonea a pregiudicare le statistiche del rider, giacché il rider che aderisce ad uno sciopero, astenendosi dall'attività lavorativa – ciò in cui consiste il comportamento materiale definibile come sciopero – e dunque non partecipando ad una sessione prenotata, verrà inevitabilmente a subire una diminuzione del suo punteggio sotto il profilo della affidabilità».

Se la possibilità di ricostruire la regola eseguita dal sistema consente di raggiungere la prova del particolare svantaggio da essa derivante per i lavoratori in sciopero e quindi un livello di conferma probatoria della fattispecie discriminatoria indiretta sufficiente a spostare sul convenuto il rischio di soccombenza, è interessante osservare come per il giudice la prova «del concreto meccanismo di funzionamento dell'algoritmo che elabora le statistiche dei rider» gravasse sul datore «se non altro sulla base del generale principio di vicinanza».

Con questa affermazione, l'ordinanza pare cadere in una sorta di cortocircuito.

L'onere di prova individuato in capo ai ricorrenti è dal giudice espressamente collegato alla necessità che essi diano indicazione di «una condizione di svantaggio per un gruppo di lavoratori». Nell'ambito della fattispecie della discriminazione indiretta, la posizione di particolare svantaggio in cui sono posti i portatori di un determinato fattore di rischio a causa dell'impatto di un criterio ad applicazione generalizzata e indifferenziata costituisce il fatto che determina il collegamento tra il criterio apparentemente neutro e il fattore di rischio. Il collegamento c'è se la condotta datoriale può spiegare i propri effetti oltre il singolo lavoratore, nei confronti del gruppo di cui fa parte, ponendone i componenti in una posizione di particolare svantaggio rispetto al gruppo dei non portatori (senza che chi subisca lo svantaggio sia necessariamente portatore del fattore di rischio in questione, altrimenti si tratterebbe di discriminazione diretta)<sup>(23)</sup>. Per quanto l'alleggerimento dell'onere

---

<sup>(23)</sup> Cfr. M. Peruzzi, *op. cit.*, spec. 159-161 e 168-172. In caso di discriminazione indiretta, l'oggetto della prova di cui il ricorrente è onerato non può essere costituito da una porzione della fattispecie, ad esempio la mera disparità di trattamento, posto che la disparità ha in questo caso una precisa causa immediata che ne giustifica la portata e che in astratto non

consenta ai ricorrenti di offrire una prova di detto collegamento in termini di probabilità “attenuata”, tale prova presuppone comunque quella del collegamento tra il trattamento subito e l’applicazione del criterio apparentemente neutro. In altre parole, presuppone la prova del collegamento tra l’esclusione e il funzionamento dell’algoritmo (che potrebbe essere anche tratta dalla deduzione di specifici fatti storici che ne diano concreta rappresentazione, ad es. la sistematica esclusione dal gruppo prioritario di accesso al sistema di prenotazione dei lavoratori che abbiano effettuato cancellazioni tardive, tra cui i partecipanti a iniziative di astensione collettiva).

Se, tuttavia, per operatività del generale principio di vicinanza della fonte, si addossa al datore la prova del “concreto meccanismo di funzionamento dell’algoritmo”, si può ritenere che ciò equivalga a spostare sul convenuto la prova del collegamento tra il criterio apparentemente neutro e il trattamento subito dai lavoratori. In tal senso, l’assetto di distribuzione degli oneri così articolato va a incidere necessariamente sulla porzione a carico del ricorrente e sulla regola di giudizio che ne deriva.

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.

Il cortocircuito è forse la diretta conseguenza del confronto tra il quadro antidiscriminatorio e le possibili opacità dei sistemi algoritmici, una criticità che nel caso deciso dal Tribunale di Bologna è superata dalla ricostruibilità della regola a partire dal contenuto del contratto-tipo e dall’esperienza diretta dei lavoratori poi chiamati a testimoniare.

---

è connotata in termini discriminatori. Potrebbe essere solo la disparità di trattamento laddove il ricorrente affermasse la pretestuosità della causa immediata e la sussistenza di un nesso di causalità diretto con il fattore di rischio, quale reale causa immediata occultata. In tal senso, tuttavia, il ricorrente starebbe deducendo in giudizio una fattispecie di discriminazione diretta, e non indiretta. Si noti che nel caso deciso dall’ordinanza del Tribunale di Bologna, le ricorrenti avevano affermato che il sistema di prenotazione delle fasce di lavoro per priorità in ragione delle statistiche assegnate al rider era stato introdotto da Deliveroo come misura organizzativa in coincidenza con le prime iniziative di astensione dal lavoro attuate dai riders autorganizzati negli anni 2017/2018. Sulla configurabilità della discriminazione operata dall’algoritmo Frank come diretta cfr. il dibattito nell’ambito del webinar *Frank discrimina*, organizzato dalla *Labour Law Community* e tenutosi il 2 febbraio 2021, la cui registrazione è visualizzabile sul sito dell’associazione.

Più complesso, si ribadisce, è lo scenario laddove la regola sia difficile da decodificare.

Il rischio che l'opacità dei sistemi di IA neutralizzi, in questo caso, l'operatività degli strumenti disposti dal diritto antidiscriminatorio a protezione del lavoratore, anche in punto di prova, richiede una riflessione sulla necessità di ulteriori e specifiche misure di tutela<sup>(24)</sup>.

Proprio a tal riguardo, nel prosieguo della trattazione si cercheranno di individuare delle prospettive di soluzione. Partendo dal tracciamento dei livelli normativi interessati dalla questione della trasparenza nel contesto di sistemi di IA, si arriverà a verificare la funzionalità delle tecniche di tutela contenute in due testi normativi proposti dalla Commissione europea tra marzo e aprile 2021: da un lato, la proposta di direttiva volta a rafforzare l'applicazione del principio della parità retributiva di genere attraverso la trasparenza delle retribuzioni<sup>(25)</sup>; dall'altro, la proposta di Regolamento sull'Intelligenza Artificiale.

### **3. Trasparenza, decisioni automatizzate e discriminazioni nella rete dei livelli normativi**

La questione della trasparenza nel contesto di decisioni automatizzate con utilizzo di sistemi di IA implica una riflessione di carattere sistematico. Il tema intercetta, infatti, *de iure condito*, più nuclei normativi.

Innanzitutto, nella prospettiva della tutela della riservatezza nel trattamento dei dati personali, il Regolamento Ue 2016/679 prevede il diritto dell'interessato di informazione (art. 13) e di accesso (art. 15) nonché l'obbligo del titolare del trattamento di fornire informazioni specifiche e facilmente accessibili sul processo decisionale (artt. 14 e 15)<sup>(26)</sup>. L'applicazione di tali

---

<sup>(24)</sup> Come segnala G. Gaudio (*op. cit.*, 49), «il combinato disposto della parziale inversione degli oneri della prova e dei diritti di informazione e accesso del GDPR sembra, nella maggior parte dei casi, ridurre solo in astratto il rischio per un lavoratore di subire le conseguenze negative dell'opacità decisionale algoritmica. Nonostante sia stato effettivamente vittima di discriminazioni e sempre ammesso che se ne sia reso conto e abbia impugnato, il lavoratore, molto probabilmente, soccomberebbe quindi in giudizio, salvo che il giudice decida di esercitare i propri poteri istruttori per far emergere nel processo la verità materiale che si cela dietro l'algoritmo, seguendo eventuali “piste probatorie” aperte dai ricorrenti grazie all'esercizio, preventivo e stragiudiziale, dei diritti loro garantiti dal GDPR».

<sup>(25)</sup> Commissione europea, *Proposta di direttiva volta a rafforzare l'applicazione del principio della parità di retribuzione tra uomini e donne per uno stesso lavoro o per un lavoro di pari valore attraverso la trasparenza delle retribuzioni e meccanismi esecutivi*, Bruxelles, 4.3.2021 COM(2021) 93 final.

<sup>(26)</sup> Per il rapporto tra normativa sulla privacy e art. 4 St. Lav., basato sul reciproco richiamo (art. 114 d.lgs. n. 196/03 – art. 4, co. 3, St. Lav.) e sulla previsione al comma terzo della disposizione statutaria della necessità che sia data «adeguata informazione delle modalità

disposizioni è tanto più rilevante laddove siano integrate le eccezioni al divieto di decisioni basate unicamente sul trattamento automatizzato di cui all'art. 22<sup>(27)</sup>, tra cui rientra l'ipotesi in cui tale processo costituisca un mezzo appropriato e necessario<sup>(28)</sup> per la conclusione o l'esecuzione di un contratto tra l'interessato e il titolare del trattamento (ad es., nell'ambito del lavoro, profilazione ai fini dell'assunzione in caso di volume eccezionalmente elevato di candidature)<sup>(29)</sup>.

---

d'uso degli strumenti e di effettuazione dei controlli» per l'utilizzabilità dei dati raccolti a tutti i fini connessi al rapporto, cfr. *ex multis*: M. P. Monaco, *Controlli a distanza sui lavoratori: evoluzione, riforme e privacy*, *Labor*, 2021, 2, 155 ss.; A. Sartori, *Il controllo tecnologico sui lavoratori*, Giappichelli, 2020; P. Tullini, *La digitalizzazione del lavoro, la produzione intelligente e il controllo tecnologico nell'impresa*, in P. Tullini (a cura di), *Web e lavoro. Profili evolutivi e di tutela*, Giappichelli, 2017, 3 ss.; Ead. (a cura di), *Controlli a distanza e tutela dei dati personali del lavoratore*, Giappichelli, 2017; L. Tebano, *Tutela della privacy e potere di controllo del datore di lavoro tra l'ordinamento italiano e le fonti europee*, *LLI*, 2017, 2, C-1 ss. Sul tema generale del trattamento dei dati nell'ambito del lavoro, cfr. di recente l'analisi di L. D'Arcangelo, *L'obbligo di protezione dei dati del lavoratore: adempimento e sanzioni*, *DLM*, 2020, 1, 75 ss.; A. Sitzia - S. Crafa, *Impronte digitali, algoritmo e trattamento di dati personali: questioni di "law and technology"*, *LG*, 2019, 3, 245 ss.; A. Aloisi - E. Gramano, *Artificial Intelligence Is Watching You at Work: Digital Surveillance, Employee Monitoring, and Regulatory Issues in the EU Context*, *CLL&PJ*, 2019, 41, 1, 95 ss.; C. Ogriseg, *GDPR and Personal Data Protection in the Employment Context*, *LLI*, 2017, 2, R.1 ss.

<sup>(27)</sup> Cfr. il Protocollo di emendamento alla Convenzione sulla protezione delle persone rispetto al trattamento automatizzato di dati a carattere personale (Convenzione n. 108 del Consiglio d'Europa), fatto a Strasburgo il 10 ottobre 2018 e ratificato in Italia dalla l. 22 aprile 2021, n. 60.

<sup>(28)</sup> Per parte della dottrina, «sembra che il divieto generale di soggezione alla decisione unicamente automatizzata non sia facilmente superabile in ambito lavorativo. Pare difficile che essa sia “necessaria” – e quindi imprescindibile – nella gestione del rapporto [...] o che possa valere il consenso del candidato o del dipendente, visto lo squilibrio del rapporto sottostante», F. Costantini, *Profilazione e “automated decision making” in ambito lavorativo nella giurisprudenza italiana*, *LG*, 2019, 11, 992-993. Rispetto al consenso, sono le stesse Linee guida 5/2020 sul consenso, adottate dal Comitato europeo per la protezione dei dati, a ritenere «problematico per il datore di lavoro trattare i dati personali dei dipendenti attuali o futuri sulla base del consenso, in quanto è improbabile che questo venga prestato liberamente. Per la maggior parte delle attività di trattamento svolte sul posto di lavoro, la base legittima non può e non dovrebbe essere il consenso del dipendente [...] in considerazione della natura del rapporto tra datore di lavoro e dipendente» (10).

<sup>(29)</sup> Cfr. Linee guida WP 251, *Sul processo decisionale automatizzato relativo alle persone fisiche e sulla profilazione ai fini del regolamento 2016/679*, adottate il 3 ottobre 2017 dal “Gruppo di lavoro ex art. 29” (ora Comitato europeo per la protezione dei dati), p. 26. Mentre ai sensi dell'art. 22 «il titolare del trattamento attua misure appropriate per tutelare i diritti, libertà e i legittimi interessi dell'interessato, almeno il diritto di ottenere l'intervento umano da parte del titolare del trattamento, di esprimere la propria opinione e di contestare la decisione», il considerando n. 71, pur di natura non vincolante, indica la necessità di «garanzie adeguate, che dovrebbero comprendere», oltre quanto previsto dall'art. 22, «la specifica informazione all'interessato e il diritto di ottenere [...] una spiegazione della decisione». E' discussa in dottrina la possibilità di affiancare al diritto all'informazione il diritto alla spiegazione (nel senso di escluderlo, S. Wachter - B. Mittelstadt - L. Floridi, *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation, International Data Privacy Law*, 2017, 76 ss.;



Si noti, peraltro, che, ai sensi dell'art. 22, par. 4, il processo decisionale automatizzato può anche basarsi sulle categorie particolari di dati personali di cui all'art. 9, par. 1, del Regolamento (gli ex cd. «dati sensibili»), laddove sussistano «motivi di interesse pubblico rilevante» (art. 9, par. 2, lett. g). All'interno di tali motivi, l'art. 2 sexies, d.lgs. n. 196/03 (introdotto dal d.lgs. n. 101/2018) fa rientrare anche trattamenti effettuati da soggetti che svolgono compiti di interesse pubblico o connessi all'esercizio di pubblici poteri in materia di rapporti di lavoro<sup>(30)</sup>, ferma la necessità che specifiche norme di legge (o regolamentari, nei casi previsti dalla legge) dettagliino «i tipi di dati che possono essere trattati, le operazioni eseguibili e il motivo di interesse pubblico rilevante, nonché le misure appropriate e specifiche per tutelare i diritti fondamentali e gli interessi dell'interessato». Ai fini dell'analisi che ci occupa, è significativo rilevare come la lista delle categorie particolari di dati di cui all'art. 9<sup>(31)</sup> non coincida con l'elenco dei fattori di rischio previsto dal quadro normativo antidiscriminatorio, differendo sia in eccesso (caratteristiche genetiche, biometriche, salute, vita sessuale), sia in difetto (genere, età, disabilità), per quanto i fattori mancanti possano essere fatti rientrare in alcune delle categorie indicate. Il confronto rimanda al dibattito sulla tipicità dei fattori di rischio ai fini dell'applicazione del sistema normativo speciale antidiscriminatorio e sulla possibilità o meno di far derivare una sua *vis expansiva* dall'impatto dell'art. 21 della Carta di Nizza, che contempla ad esempio proprio il riferimento alle caratteristiche genetiche, o a livello interno dell'art. 3 Cost. Si consenta di

---

lo includono, invece, A. D. Selbst - J. Powles, *Meaningful information and the right to explanation*, *International Data Privacy Law*, 2017, 233 ss.). Sul punto si rinvia all'analisi di E. Falletti, *Decisioni automatizzate e diritto alla spiegazione: alcune riflessioni comparatistiche*, *Diritto dell'informazione e dell'informatica*, 2020, 2, 169 ss.; F. Hendrickx, *Privacy 4.0 at Work: Regulating Employment, Technology and Automation*, *CLL&PJ*, 2019, 41, 1, 147 ss.

<sup>(30)</sup> Segnatamente, «compiti di igiene e sicurezza sui luoghi di lavoro» (co. 2, lett. u), «instaurazione, gestione ed estinzione, di rapporti di lavoro di qualunque tipo, anche non retribuito o onorario, e di altre forme di impiego, materia sindacale, occupazione e collocamento obbligatorio, previdenza e assistenza, tutela delle minoranze e pari opportunità nell'ambito dei rapporti di lavoro, adempimento degli obblighi retributivi, fiscali e contabili, igiene e sicurezza del lavoro o di sicurezza o salute della popolazione, accertamento della responsabilità civile, disciplinare e contabile, attività ispettiva» (co. 2, lett. dd).

<sup>(31)</sup> Operando un confronto con quanto previsto dalla dir. 95/46/CE, M. Granieri (*Il trattamento di categorie particolari di dati personali nel regolamento UE 2016/679, NLCC*, 2017, 167) osserva che «il Regolamento è più specifico e aggiunge i dati genetici, i dati biometrici intesi a identificare in modo univoco una persona fisica, oltre a quelli relativi alla salute o alla vita sessuale o all'orientamento sessuale della persona. Il tempo trascorso tra il 1995 e il 2016 ha certamente consentito al legislatore di costruire sull'esperienza dei casi, delle discriminazioni e degli abusi che si sono verificati in vari ambiti. La tecnologia nel frattempo ha dimostrato che la sua evoluzione può effettivamente creare rischi crescenti per la tutela delle informazioni personali e, a maggior ragione, per talune categorie di dati».

aggiungere che allo stesso dibattito rinvia la riflessione sul rapporto tra la nozione di discriminazione (e di gruppo vulnerabile) utilizzata dalla proposta di Regolamento sull'IA e quella rilevante ai fini dell'applicazione del sistema di regole fornito dal diritto antidiscriminatorio (v. *infra*)<sup>(32)</sup>.

Proseguendo nel ragionamento sulla trasparenza nel contesto delle regole a tutela della privacy, ai sensi degli artt. 14 e 15 Reg. e come evidenziano le Linee guida WP 251, in caso di processi decisionali automatizzati l'obbligo di informazione e il diritto di accesso si traducono nell'obbligo del titolare di «fornire informazioni significative sulla logica utilizzata». In particolare, anche con riguardo all'uso di sistemi di apprendimento automatico, «il titolare del trattamento dovrebbe trovare modi semplici per comunicare all'interessato la logica o i criteri sui quali si basa l'adozione della decisione». Ciò non equivale necessariamente a «una spiegazione complessa degli algoritmi utilizzati o la divulgazione dell'algoritmo completo» (par. E, pt. 1).

Quest'ultima precisazione si ricollega a quanto affermato dal considerando n. 63 del Regolamento, ove si chiarisce che il diritto dell'interessato alle suddette informazioni «non dovrebbe ledere i diritti e le libertà altrui, compreso il segreto industriale e aziendale e la proprietà intellettuale, segnatamente i diritti d'autore che tutelano il software».

A essere intercettato, al riguardo, è un secondo nucleo normativo, ossia la disciplina europea in materia di diritti d'autore sul software (dir. 2009/24/CE, recepita nella legge sul diritto d'autore n. 633/41), la quale non impone mai un obbligo di ostensione del codice sorgente<sup>(33)</sup>, rendendo al più possibile l'attività di decompilazione necessaria per garantire l'interoperabilità del programma,

---

<sup>(32)</sup> Come spiega M. Barbera (*Il licenziamento alla luce del diritto antidiscriminatorio*, RGL, 2013, 161 e 152), la *ratio* del diritto antidiscriminatorio risiede «nel particolare disvalore attribuito a diseguaglianze che nascono da caratteristiche costituenti l'identità soggettiva della persona»; tali distinzioni sono «qualitativamente diverse da comportamenti scorretti e da distinzioni arbitrarie, ingiustificate o ingiuste che possono colpire occasionalmente e in modo indistinto tutti i lavoratori». Ancora, «è sempre il carattere strutturale, sistematico delle discriminazioni, la loro tipicità sociale, a legittimare un regime di prova alleggerito, a rendere “verosimile” la prova statistica, a richiedere forme specifiche di rimedi, di natura inibitoria e restitutoria (come i piani di azione positiva e di rimozione delle discriminazioni), a consentire il risarcimento del danno patrimoniale e non patrimoniale anche in funzione dissuasiva e in assenza di colpa, a istituire forme collettive di legittimazione ad agire» (*Il cavallo e l'asino. Ovvero dalla tecnica della norma inderogabile alla tecnica antidiscriminatoria*, in O. Bonardi (a cura di), *Eguaglianza e divieti di discriminazione nell'era del diritto del lavoro derogabile*, Ediesse, 2017, 24). Cfr. i contributi di M. Barbera, *Principio di eguaglianza e divieti di discriminazione*, e di M. Militello - D. Strazzari, *op. cit.*, in M. Barbera - A. Guariso (a cura di), *La tutela antidiscriminatoria*, Giappichelli, 2019, 5 ss. e 85 ss.

<sup>(33)</sup> Per una definizione, cfr. TAR Lazio, 21 marzo 2017 n. 3742.

senza compromissione della segretezza del codice e dei connessi diritti di privativa intellettuale<sup>(34)</sup>.

In tale prospettiva, è fortemente contestata la giurisprudenza amministrativa interna sviluppatasi sul tema con riguardo a procedimenti amministrativi a processo decisionale automatizzato e basato sull'utilizzo di strumenti informatici algoritmici (spesso procedure di assegnazione di sedi). Nello specifico, la maggior parte delle pronunce trae dal diritto di accesso agli atti amministrativi di cui all'art. 24, co. 7, l. n. 241/1990 ovvero dal più generale principio di trasparenza dell'agire amministrativo un obbligo di ostensione del "linguaggio sorgente" del software utilizzato nell'ambito del procedimento, anche quando la p.a. sia ricorsa alla fornitura di software di tipo proprietario acquisendo una mera facoltà d'uso, quindi con permanenza in capo al privato appaltatore dei diritti di proprietà intellettuale sul software. L'argomento adottato è che «non può assumere rilievo l'invocata riservatezza delle imprese produttrici dei meccanismi informatici utilizzati i quali, ponendo al servizio del potere autoritativo tali strumenti, all'evidenza ne accettano le relative conseguenze in termini di necessaria trasparenza» (Cons. St. 4 febbraio 2020, n. 881; Cons. St. 13 dicembre 2019, n. 8472). Oltre al contrasto con la regola giuridica in materia di tutela dei diritti d'autore, per parte della dottrina la posizione della giurisprudenza amministrativa tradirebbe anche uno scollamento rispetto alla regola tecnica, dal momento che l'ostensione del codice non consente di raggiungere l'obiettivo di tutela auspicato: «le istruzioni (in linguaggio di programmazione) con cui vengono scritti i programmi utilizzati per l'adozione di decisioni automatizzate non seguono una "logica lineare", ma modelli "euristici" basati su "sistemi esperti" o su reti "neurali": dovendo gestire quella che viene definita "un'esplosione combinatoria", forniscono "meta-regole" o altre strategie di funzionamento che consentono al software di *decision-making*, unitamente ad una base di conoscenze predeterminate ed incrementalmente (anche mediante sistemi di *machine learning* capaci di autoapprendimento), di raggiungere soluzioni non predeterminabili sulla base delle sole istruzioni contenute nel codice sorgente»<sup>(35)</sup>. Il nodo è complesso e le sue plurime articolazioni esulano dall'ambito del presente contributo. Basti solo evidenziare l'interconnessione anche con la rete normativa di protezione del segreto

---

<sup>(34)</sup> F. Bravo, *Trasparenza del codice sorgente e decisioni automatizzate*, *Diritto dell'informazione e dell'informatica*, 2020, 4, 693 ss. Cfr. in merito la stringente interpretazione adottata da Corte giust. 2 maggio 2012, causa C-406/10, *SAS Institute* nonché dall'AG M. Szpunar nelle Conclusioni, 10 marzo 2021, causa C-13/20, *Top System*.

<sup>(35)</sup> F. Bravo, *op. cit.*, 715.

commerciale, anch'essa di matrice europea (dir. 2016/943, recepita nel d.lgs. n. 30/2005, cd. codice della proprietà industriale), e di possibile rilevanza laddove prevede l'apprestamento di specifiche garanzie a tutela della riservatezza dei segreti commerciali nel corso di procedimenti giudiziari (art. 9 dir. 2016/943; art. 121 ter, d.lgs. n. 30/2005).

#### **4. Ragionando di possibili prospettive di tutela, a partire dalla proposta di Regolamento sull'IA e dalla proposta di direttiva su trasparenza e parità retributiva di genere**

Individuate le criticità che si riscontrano nell'applicazione del diritto antidiscriminatorio di fronte al problema della *black box* e verificato come il tema della trasparenza si irradi nella rete dei livelli normativi, l'iter d'indagine si svilupperà ora nella ricerca di possibili soluzioni *de iure condendo*.

Come anticipato, questo percorso muoverà dalla lettura di due testi normativi proposti dalla Commissione europea tra marzo e aprile 2021: non solo la proposta di Regolamento sull'IA, ma anche la proposta di direttiva volta a rafforzare l'applicazione del principio della parità retributiva di genere attraverso la trasparenza delle retribuzioni.

L'obiettivo è analizzare i modelli di tutela apprestati per poi valutare l'opportunità di mutuare e integrare alcune tecniche previste nella proposta di direttiva anche nell'ambito di studio che ci occupa.

Si osservi, anzitutto, come il tema della trasparenza e del contrasto all'opacità delle decisioni algoritmiche sia affrontato in modo specifico nella citata proposta di Regolamento sull'IA.

Segnatamente, si prevede, per i sistemi cd. "ad alto rischio", tra cui sono compresi quelli utilizzati per la gestione delle relazioni contrattuali di lavoro (anche autonomo)<sup>(36)</sup>, che la documentazione tecnica indichi, tra le varie informazioni, «la logica generale del sistema di IA e degli algoritmi», «le principali scelte di progettazione [...] anche per quanto riguarda le persone o i gruppi di persone sui quali il sistema è destinato a essere utilizzato», «le risorse computazionali usate per sviluppare, addestrare, sottoporre a prova e convalidare il sistema di IA», «le metriche utilizzate per misurare l'accuratezza, la robustezza, la cibersecurity e la conformità ad altri requisiti pertinenti»<sup>(37)</sup> così come «gli impatti potenzialmente discriminatori», «i prevedibili risultati

---

<sup>(36)</sup> V. art. 6 e annex III, punto 4.

<sup>(37)</sup> Cfr. R. Flor, *Cybersecurity ed il contrasto ai cyber-attacks a livello europeo: dalla CIA-Triad Protection ai più recenti sviluppi*, *Diritto di internet*, 2019, 3, 453 ss.

indesiderati e fonti di rischio per la salute, la sicurezza e i diritti fondamentali, nonché di rischio di discriminazione in considerazione della finalità prevista del sistema di IA; le misure di sorveglianza umana necessarie [...], comprese le misure tecniche poste in essere per facilitare l'interpretazione degli output dei sistemi di IA da parte degli utenti; le specifiche relative ai dati di input, se del caso»<sup>(38)</sup>. Come chiarito dal considerando n. 47, proprio la documentazione tecnica (insieme alle istruzioni d'uso) andrebbe a consegnare un certo grado di trasparenza ai sistemi ad alto rischio, consentendo di gestire il problema dell'opacità che potrebbe altrimenti renderli incomprensibili o troppo complessi per la lettura umana e permettendo agli utenti di «interpretare gli output del sistema e utilizzarlo in modo adeguato».

La *ratio* della proposta è quella di ridurre i rischi connessi all'uso dell'IA attraverso un sistema di garanzie procedurali che richiama, per l'approccio di *accountability* nonché i meccanismi di *compliance*, verifica di conformità<sup>(39)</sup> e valutazione/gestione del rischio previsti, il percorso tracciato dal Regolamento sulla privacy per il trattamento dei dati.

Particolare attenzione è rivolta al tema delle discriminazioni all'interno dei sistemi ad alto rischio. In linea generale, nella relazione accompagnatoria si precisa che la proposta integra «il diritto dell'Unione in vigore in materia di non discriminazione con requisiti specifici che mirano a ridurre al minimo il rischio di discriminazione algoritmica, in particolare in relazione alla progettazione e alla qualità dei set di dati utilizzati per lo sviluppo dei sistemi di IA, integrati con obblighi relativi alle prove, alla gestione dei rischi, alla documentazione e alla sorveglianza umana durante l'intero ciclo di vita dei sistemi di IA» (par. 1.2, p. 4).

Come già anticipato, è necessario evidenziare come il testo faccia riferimento, in alcuni casi, al concetto generale di “*bias*”, ossia di “distorsione”, “parzialità”, quale possibile causa di discriminazione (ad es. cons. 44), in altri direttamente a quello di discriminazione, senza offrirne tuttavia una definizione, identificando alcuni fattori di rischio in termini non sistematici e meramente esemplificativi (genere, età, disabilità, razza e origine etnica, orientamento sessuale, v. *infra*), menzionando la possibilità di nuove forme di discriminazione (cons. 37), individuando gruppi di soggetti vulnerabili, in rapporto ai quali

---

<sup>(38)</sup> V. art. 11 e annex IV, punto 2 (in particolare, lett. b, c, g) e punto 3.

<sup>(39)</sup> Nella posizione pubblicata il 28 maggio 2021, l'ETUC contesta la proposta sul punto, nella misura in cui prevede una mera autovalutazione da parte del fornitore, evidenziando la necessità che, quantomeno per i sistemi di IA concernenti il lavoro, la valutazione di conformità sia realizzata da una parte terza autorizzata. Raccomanda, inoltre, una combinazione tra una fase di *ex-ante compliance* ed una di *ex-post enforcement*.

vietare le pratiche di IA (art. 5)<sup>(40)</sup> o misurare il livello di rischio dei sistemi ai fini della loro classificazione (art. 7)<sup>(41)</sup>. Se la relazione accompagnatoria segnala la complementarità di questo quadro di regole rispetto alle fonti antidiscriminatorie UE vigenti, già si è potuto osservare come sull'interazione di questi due sistemi normativi si innesti il dibattito sulla tipicità o meno dei fattori di rischio ai fini dell'applicazione degli speciali strumenti di tutela apprestati dal diritto antidiscriminatorio (v. *supra*).

Con specifico riferimento all'ambito del lavoro, il considerando n. 36 evidenzia come attraverso il processo di assunzione e selezione ovvero nella valutazione, promozione o gestione delle persone in rapporti contrattuali di lavoro (compresi quelli che prevedono la fornitura di servizi su piattaforma digitale), i sistemi di IA «possono 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». Al considerando n. 44, si sottolinea la necessità, a tal fine, che i dati utilizzati siano di alta qualità. Segnatamente, «i set di dati di addestramento, convalida e prova dovrebbero tenere conto, nella misura necessaria alla luce della finalità prevista, delle caratteristiche o degli elementi peculiari dello specifico contesto o ambito geografico, comportamentale o funzionale all'interno del quale il sistema di IA ad alto rischio è destinato a essere usato. Al fine di proteggere i diritti di terzi dalla discriminazione che potrebbe derivare dalla distorsione nei sistemi di IA, è opportuno che i fornitori siano in grado di trattare anche categorie particolari di dati personali, come questione di rilevante interesse pubblico, al fine di garantire il monitoraggio, il rilevamento e la correzione delle distorsioni in relazione ai sistemi di IA ad alto rischio».

Prima di analizzare il modello di tutela apprestato nel secondo testo normativo indicato in premessa, ossia la proposta di direttiva sulla parità retributiva di genere<sup>(42)</sup>, è utile ricordare come il tema della trasparenza non sia

---

<sup>(40)</sup> È stabilito che siano vietati «l'immissione sul mercato, la messa in servizio o l'uso di un sistema di IA che sfrutta le vulnerabilità di uno specifico gruppo di persone, dovute all'età o alla disabilità fisica o mentale» (art. 5, par. 1, lett. b).

<sup>(41)</sup> Nell'aggiornamento della lista dei sistemi ad alto rischio di cui all'annex 3, la Commissione deve tenere in considerazione anche «la misura in cui le persone che potrebbero subire il danno o l'impatto negativo si trovano in una posizione vulnerabile rispetto all'utente di un sistema di IA, in particolare a causa di uno squilibrio di potere, conoscenza, situazione economica o sociale o età» (art. 7, par. 2, lett. f).

<sup>(42)</sup> La proposta di direttiva deve essere coordinata con quanto previsto, in termini generali, dalla dir. 2019/1152/UE relativa a condizioni di lavoro trasparenti e prevedibili nell'Unione europea.

affatto inedito per il diritto antidiscriminatorio e sia anzi risalente, trovando una specifica declinazione nell'ambito del contrasto al *gender pay gap*.

A evidenziarsi in tale contesto, oltre alle recenti prospettive presentate dalla Commissione, è il ragionamento che vi si ricollega in punto di onere della prova, elaborato dalla Corte di giustizia già alla fine degli anni Ottanta nel caso *Danfoss*<sup>(43)</sup>.

È proprio l'assenza di trasparenza, segnatamente dei sistemi classificatori e retributivi, e quindi la conseguente inconoscibilità delle cause delle differenze di trattamento da parte delle lavoratrici, a portare la Corte, in *Danfoss*, a individuare a favore della ricorrente (nel caso di specie, la Federazione degli impiegati) un onere di prova alleggerito, limitato cioè al dato storico della disparità. Rinviando ad altra sede l'analisi della questione statistica implicata nell'offerta di tale prova e non priva a sua volta di problematicità<sup>(44)</sup>, preme qui sottolineare come lo schema di distribuzione dell'onere fissato dalla Corte in questa sentenza, utile a interpretare il regime probatorio speciale del diritto antidiscriminatorio in termini di parziale inversione dell'onere, sia connesso a un problema di opacità dei criteri sottesi al trattamento<sup>(45)</sup>.

---

<sup>(43)</sup> CgCE 17 ottobre 1989, C-109/88, *Danfoss*.

<sup>(44)</sup> Come evidenzia la dottrina, la questione della quantificazione della disparità di trattamento evidenziata dai dati statistici è ampiamente discussa, «in quanto non è stabilita per legge una soglia oltre la quale la sproporzione dello svantaggio possa essere considerata discriminatoria». Il problema è tanto più complicato laddove «le differenze non sono così nette e l'analisi statistica si fonda su dati meramente aziendali» (nel caso *Danfoss*, la prova statistica fornita dall'attore rivela una differenza retributiva tra uomini e donne in azienda pari al 7%). Si osserva come, d'altra parte, «non occorre fare riferimento solo a una differenza quantitativamente significativa, ma all'effetto sostanziale e pratico prodotto», nella prospettiva indicata dalla riformulazione della nozione di discriminazione indiretta adottata dalle direttive del 2000 (O. Bonardi - C. Meraviglia, *Dati statistici e onere della prova nel diritto antidiscriminatorio*, in O. Bonardi (a cura di), *Eguaglianza e divieti di discriminazione nell'era del diritto del lavoro derogabile*, *op. cit.*, 375-376; C. Barnard - B. Hepple, *Indirect Discrimination: Interpreting Seymour-Smith*, *CLJ*, 1999, 399 ss.). Può essere rilevante tale questione per evidenziare la difficoltà di offrire una prova del dato storico della disparità – tra soggetti in posizione analoga in caso di discriminazione diretta (anche) collettiva ovvero tra portatori o meno di un fattore in caso di applicazione di un criterio “neutro” a impatto differenziato – in assenza di un numero adeguato di soggetti da includere nella comparazione o di un differenziale quantificabile in termini significativi.

<sup>(45)</sup> In merito alla funzione prodromica del *leading case Danfoss* rispetto agli interventi normativi in materia e alla sua importanza per l'interpretazione del regime probatorio speciale previsto dal diritto antidiscriminatorio, cfr. M. Peruzzi, *op. cit.*, spec. 164 ss.; come osservano anche O. Bonardi - C. Meraviglia (*op. cit.*, 359-360), richiamando l'attenzione sul caso *Danfoss*, «il principio dell'inversione dell'onere della prova si afferma in presenza di una situazione tipica delle fattispecie discriminatorie data dal fatto che la persona che lamenta la discriminazione dimostra l'effetto oggettivamente discriminatorio, ovvero la situazione in cui una persona è trattata meno favorevolmente rispetto a un'altra o ad altri, ma non è in grado di dimostrare la

Si noti, tuttavia, come la tecnica di tutela dell'alleggerimento dell'onere non consenta necessariamente di superare il problema dell'assenza di trasparenza, soprattutto nei contesti in cui non sia possibile per il ricorrente fornire nemmeno la prova di una mera porzione di fattispecie, ossia del fatto storico della disparità rispetto a un soggetto comparabile reale.

Deve esserne consapevole la Commissione europea nel momento in cui, all'interno della recente proposta di direttiva, prevede che, all'inosservanza da parte del datore di specifici obblighi di informativa, reporting e confronto, consegua l'applicazione di una tecnica di tutela ben più ampia, ossia un'inversione completa dell'onere di prova, con passaggio del rischio di soccombenza esclusivamente a carico del convenuto. Oltretutto si specifica, pur non risultando chiaro se solo rispetto a questo tipo di inversione, che "la parte ricorrente beneficia di qualsiasi dubbio residuale" (art. 16, par. 3).

Il modello proposto dalla Commissione è di particolare interesse per il tema che ci occupa.

È significativo, anzitutto, perché stabilisce, a monte e nella prospettiva della prevenzione, non solo diritti soggettivi individuali di informazione e un processo di rendicontazione di dati aggregati (comunque con più variabili rispetto a quello richiesto dal nostro art. 46 d.lgs. n. 198/06). A seguire, prevede inoltre la possibilità per i lavoratori e i loro rappresentanti di chiedere chiarimenti e dettagli ulteriori, comprese spiegazioni su eventuali disparità, nonché l'effettuazione di una valutazione congiunta in cooperazione con i rappresentanti dei lavoratori, al ricorrere di un divario medio (su dati aggregati) pari almeno al 5%, non giustificato dal datore con fattori oggettivi e "neutri" sotto il profilo del genere<sup>(46)</sup>.

Se mutuata a garanzia della trasparenza e del contrasto alle discriminazioni nei sistemi di IA, la tecnica procedurale descritta può orientare il sistema di regole verso almeno due importanti obiettivi.

In primo luogo, permette di integrare la dimensione collettiva nella rete dei diritti di informazione e dei processi di trasparenza contemplati dagli altri livelli normativi coinvolti, quello a protezione della privacy e quello delineato nella proposta di Regolamento sull'IA. In tali ambiti, infatti, gli strumenti

---

ragione. [...]. È dunque fondamentale la condizione di ignoranza rispetto ai criteri di classificazione utilizzati o rispetto ad altri elementi determinanti ai fini della valutazione della situazione concreta a giustificare l'inversione dell'onere della prova». Sul rapporto tra trasparenza e prova della discriminazione, cfr. anche Corte giust. 19 aprile 2012, C-415/10, *Meister*.

<sup>(46)</sup> Per alcuni rilievi critici, cfr. D. Gottardi, *Giustizia retributiva e trasparenza*, in *Studi in Memoria di Massimo Roccella*, ESI, 2021, in corso di pubblicazione.



apprestati rischiano di non rispondere all'esigenza di coinvolgere nella tutela detta dimensione, propria non solo in linea generale del diritto del lavoro ma nello specifico del diritto antidiscriminatorio, rimanendo confinati sul piano della gestione datoriale unilaterale e/o dell'esercizio individuale (fatta eccezione per il rinvio alla contrattazione collettiva di cui all'art. 88 Reg. privacy, che potrebbe fornire un punto di contatto con la prospettiva qui presentata)<sup>(47)</sup>. In merito alla proposta di Regolamento sull'IA, è la stessa ETUC nella posizione pubblicata il 28 maggio 2021 a contestare l'assenza di regole volte a garantire il coinvolgimento di sindacati e rappresentanze dei lavoratori nella costruzione e monitoraggio dei sistemi di IA nel lavoro (dalle strutture di dialogo sociale, alla contrattazione collettiva ai diritti di informazione, consultazione e partecipazione). Nella stessa ottica, si potrebbe evidenziare l'opportunità di estendere a tutti i sistemi di IA, anche a quelli ad alto rischio, quanto stabilito dalla proposta di Regolamento per quelli a basso-rischio, ossia la previsione all'interno di Codici di condotta di impegni relativi alla partecipazione degli stakeholder nella progettazione e sviluppo dei sistemi di IA e alla garanzia di diversità nella composizione dei team di sviluppo (Titolo IX, art. 69, par. 2).

In secondo luogo, nella costruzione di un processo partecipato di monitoraggio e analisi congiunta delle disparità, il modello individuato nella proposta di direttiva risponderebbe alle sollecitazioni di quella parte della dottrina che, a soluzione del problema dell'opacità nei sistemi di IA, invita a «concentrarsi sulla trasparenza degli input ed output dell'algoritmo, ovvero sulla qualità dei dati in ingresso e dei risultati in uscita», e non invece su interventi di (ri)programmazione dell'algoritmo mirati a escludere formalmente la rilevanza dei fattori di rischio. Si osserva, infatti, come non solo tali interventi non scongiurerebbero la possibilità che le discriminazioni comunque si insinuino

---

<sup>(47)</sup> Cfr. altresì le raccomandazioni in merito da parte del Gruppo di lavoro *ex art. 29*. All'interno dell'Opinion n. 2/2017, con riguardo ai trattamenti risultanti dal monitoraggio dell'uso delle tecnologie informatiche sul posto di lavoro, si «raccomanda di coinvolgere sempre un campione rappresentativo di dipendenti nella valutazione della necessità del monitoraggio, nonché della logica e dell'accessibilità» delle politiche aziendali in tema (par. 5.3); ancora «le politiche e le norme riguardanti il monitoraggio legittimo devono essere chiare e facilmente accessibili. Il Gruppo di lavoro raccomanda di coinvolgere un campione rappresentativo di dipendenti nel processo di creazione e valutazione di tali norme e politiche, in quanto la maggior parte dei monitoraggi può potenzialmente violare la vita privata dei dipendenti» (par. 6.3). Sulla prospettiva del coinvolgimento delle parti sociali nella gestione del processo di trasformazione digitale, cfr. A. Ingraio, *Data-Driven management e strategie collettive di coinvolgimento dei lavoratori per la tutela della privacy*, LLI, 2019, 2, 128 ss.; V. De Stefano, *"Negotiating the Algorithm": Automation, Artificial Intelligence, and Labor Protection*, CLL&PJ, 2019, 41, 1, 15 ss.; E. Dagnino - I. Armaroli, *A Seat at the Table: Negotiating Data Processing in the Workplace: A National Case Study and Comparative Insights*, CLL&PJ, 2019, 41, 1, 173 ss.

nelle correlazioni statistiche, ma paradossalmente andrebbero poi a depotenziare la prova del nesso di causalità tra il trattamento subito e il fattore di rischio<sup>(48)</sup>.

Proprio al riguardo, come già evidenziato, la previsione all'interno della proposta di direttiva delle garanzie procedurali è tanto più rilevante nel momento in cui vi è agganciata una specifica disciplina del regime probatorio, prevedendosi che «qualora un datore di lavoro non abbia rispettato uno qualsiasi dei diritti o degli obblighi in materia di trasparenza retributiva [...], gli Stati membri provvedono affinché spetti al datore di lavoro dimostrare, in qualsiasi procedimento giudiziario o amministrativo riguardante una discriminazione diretta o indiretta, che non vi è stata discriminazione».

La combinazione tra un sistema di garanzie procedurali che declina le tecniche di tutela della trasparenza su un piano collettivo e partecipato e un rafforzamento della protezione del ricorrente in punto di prova in caso di datore “*non-compliant*” potrebbe inoculare all'interno del diritto antidiscriminatorio anticorpi più efficaci contro le sfide del *machine learning*. Potrebbe consentire, cioè, di superare alcune delle criticità individuate nel corso della trattazione in merito all'operatività degli strumenti ivi disposti, tra cui, appunto, il meccanismo di parziale inversione dell'onere probatorio, che rischia di rimanere paralizzato laddove l'opacità algoritmica ostacoli l'offerta di prova del ricorrente necessaria per spostare sul convenuto il rischio di soccombenza.

Il confronto con le tecniche contenute nella proposta di direttiva sulla parità retributiva di genere interroga, d'altra parte, sulla validità della scelta di vincolare alcuni obblighi di trasparenza al raggiungimento di determinate soglie occupazionali, in contesti in cui proprio l'impatto della tecnologia invita a riflettere su altri criteri di misurazione della dimensione datoriale.

---

<sup>(48)</sup> Cfr. L. Giacomelli, *op. cit.*, 296; A. Agrawal - J. Gans - A. Goldfarb, *Macchine predittive*, Franco Angeli, 2018, in cui si evidenzia che «per sapere se l'IA discrimina si deve guardare a ciò che produce. Gli uomini ottengono risultati diversi da quelli delle donne? Gli ispanici ottengono risultati diversi dagli altri? E gli anziani e i disabili? Questi risultati diversi limitano le loro opportunità? [...] Per utilizzare l'IA servono investimenti aggiuntivi in consulenza sulla discriminazione e quindi lavorare per ridurre i risultati discriminatori» (202). Sul ragionamento aleggia il dubbio di chi rileva che «pensare di intervenire sui risultati dell'algoritmo ci fa mettere in discussione l'utilità di utilizzarlo» (T. Numerico, *Social network e algoritmi di machine learning: problemi cognitivi e propagazione dei pregiudizi*, Sistemi intelligenti, 2019, 476).

## Bibliografia

- Adams-Prassl J., *What if Your Boss was an Algorithm?*, in *CLL&PJ*, 2019, 41, 1, 123 ss.
- Agrawal A. - Gans J. - Goldfarb A., *Macchine predittive*, Franco Angeli, 2018.
- Alessi C., *Lavoro tramite piattaforma e divieti di discriminazione nell'UE*, in C. Alessi - M. Barbera - L. Guaglianone (a cura di), *Impresa, lavoro e non lavoro nell'economia digitale*, Cacucci, 2019, 683 ss.
- Aloisi A. - Gramano E., *Artificial Intelligence Is Watching You at Work: Digital Surveillance, Employee Monitoring, and Regulatory Issues in the EU Context*, in *CLL&PJ*, 2019, 41, 1, 95 ss.
- Ballestrero M. V., *Ancora sui rider. La cecità discriminatoria della piattaforma*, in *Labor*, 2021, 1, 104 ss.
- Barbera M., *Il cavallo e l'asino. Ovvero dalla tecnica della norma inderogabile alla tecnica antidiscriminatoria*, in O. Bonardi (a cura di), *Eguaglianza e divieti di discriminazione nell'era del diritto del lavoro derogabile*, Ediesse, 2017, 17 ss.
- Barbera M., *Il licenziamento alla luce del diritto antidiscriminatorio*, in *RGL*, 2013, 139 ss.
- Barbera M., *Principio di eguaglianza e divieti di discriminazione*, in M. Barbera – A. Guariso (a cura di), *La tutela antidiscriminatoria*, Giappichelli, 2019, 5 ss.
- Barnard C. - Hepple B., *Indirect Discrimination: Interpreting Seymour-Smith*, in *CLJ*, 1999, 399 ss.
- Bolego G., *Intelligenza artificiale e regolazione delle relazioni di lavoro: prime riflessioni*, in *Labor*, 2019, 1, 51 ss.
- Bonardi O. - Meraviglia C., *Dati statistici e onere della prova nel diritto antidiscriminatorio*, in O. Bonardi (a cura di), *Eguaglianza e divieti di discriminazione nell'era del diritto del lavoro derogabile*, Ediesse, 2017, 351 ss.
- Bravo F., *Trasparenza del codice sorgente e decisioni automatizzate*, in *Diritto dell'informazione e dell'informatica*, 2020, 4, 693 ss.
- Caruso B. - Zappalà L., *Un diritto del lavoro "tridimensionale": valori e tecniche di fronte ai mutamenti dei luoghi di lavoro*, in *WP CSDL*, It. n. 439/2021.
- Centamore G. - Ratti L., *Oltre il dilemma qualificatorio: potenzialità e limiti del diritto antidiscriminatorio nella protezione del lavoratore on-demand*, in *Lavoro tramite piattaforma e divieti di discriminazione nell'UE*, in C. Alessi - M. Barbera - L. Guaglianone (a cura di), *Impresa, lavoro e non lavoro nell'economia digitale*, Cacucci, 2019, 663 ss.
- Comandè G., *Intelligenza artificiale e responsabilità tra liability e accountability. Il carattere trasformativo dell'IA e il problema della responsabilità*, in *Analisi giuridica dell'economia*, 2019, 172 ss.
- Costantini F., *Profilazione e "automated decision making" in ambito lavorativo nella giurisprudenza italiana*, in *LG*, 2019, 11, 984 ss.
- D'Arcangelo L., *L'obbligo di protezione dei dati del lavoratore: adempimento e sanzioni*, in *DLM*, 2020, 1, 75 ss.
- Dagnino E. - I. Armaroli I., *A Seat at the Table: Negotiating Data Processing in the Workplace: A National Case Study and Comparative Insights*, in *CLL&PJ*, 2019, 41, 1, 173 ss.
- Dagnino E., *Dalla fisica all' algoritmo: una prospettiva di analisi giuslavoristica*, AUP, 2019.
- Dagnino E., *People Analytics: lavoro e tutele al tempo del management tramite big data*, in *LLI*, 2017, 3, I.2. ss.
- De Stefano V., *"Negotiating the Algorithm": Automation, Artificial Intelligence, and Labor Protection*, *CLL&PJ*, 2019, 41, 1, 15 ss.

- Del Punta R., *Innovazioni tecnologiche e diritto del lavoro*, in *Rivista degli infortuni e delle malattie professionali*, 2019, 106, fasc. 2/3, 261 ss.
- Donini A., *Tecniche avanzate di analisi dei dati e protezione dei lavoratori*, in *DRI*, 2018, 222 ss.
- Faioli M., *Discriminazioni digitali e tutela giudiziaria su iniziativa delle organizzazioni sindacali*, in *DRI*, 2021, 204 ss.
- Falletti E., *Decisioni automatizzate e diritto alla spiegazione: alcune riflessioni comparatistiche*, in *Diritto dell'informazione e dell'informatica*, 2020, 2, 169 ss.
- Flor R., *Cybersecurity ed il contrasto ai cyber-attacks a livello europeo: dalla CIA-Triad Protection ai più recenti sviluppi*, in *Diritto di internet*, 2019, 3, 453 ss.
- Gaudio G., *Algorithmic management, poteri datoriali e oneri della prova: alla ricerca della verità materiale che si cela dietro l'algoritmo*, in *LLI*, 2020, 2, 21 ss.
- Giacomelli L., *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.
- Gottardi D., *Giustizia retributiva e trasparenza*, in *Studi in Memoria di Massimo Roccella*, ESI, 2021, in corso di pubblicazione.
- Granieri M., *Il trattamento di categorie particolari di dati personali nel regolamento UE 2016/679*, in *NLCC*, 2017, 165 ss.
- Gruppo indipendente di esperti ad alto livello sull'intelligenza artificiale, *Una definizione di IA: principali capacità e discipline*, Bruxelles, 2019.
- Hacker P., *Teaching fairness to artificial intelligence: existing and novel strategies against discrimination under EU law*, in *CMLR*, 2018, 1143 ss.
- Hendrickx F., *Privacy 4.0 at Work: Regulating Employment, Technology and Automation*, in *CLL&PJ*, 2019, 41, 1, 147 ss.
- Ingrao A., *Data-Driven management e strategie collettive di coinvolgimento dei lavoratori per la tutela della privacy*, in *LLI*, 2019, 2, 129 ss.
- King A. G. - Mrkonich M., *"Big Data" and the Risk of Employment Discrimination*, in *Oklahoma Law Review*, 2016, 68, 3, 557 ss.
- Kullman M., *Platform Work, Algorithm Decision-Making, and EU Gender Equality Law*, in *IJCLLR*, 2018, 34, 11 ss.
- Maio V., *Il diritto del lavoro e le nuove sfide della rivoluzione robotica*, in *ADL*, 2018, 6, 1414 ss.
- Militello M. - Strazzari D., *I fattori di discriminazione*, in M. Barbera - A. Guariso (a cura di), *La tutela antidiscriminatoria*, Giappichelli, 2019, 85 ss.
- Monaco M. P., *Controlli a distanza sui lavoratori: evoluzione, riforme e privacy*, in *Labor*, 2021, 2, 155 ss.
- Numerico T., *Social network e algoritmi di machine learning: problemi cognitivi e propagazione dei pregiudizi*, in *Sistemi intelligenti*, 2019, 469 ss.
- Ogrisek C., *GDPR and Personal Data Protection in the Employment Context*, in *LLI*, 2017, 2, R-1 ss.
- Pacella G., *La nozione euro-unitaria di lavoratore dipendente alla prova della gig-economy: si pronuncia la Corte di giustizia europea*, in *LLI*, 1, R.17 ss.
- Peck D., *They're watching you at work*, *theatlantic.com*, dec. 2013.
- Perulli A., *La "soggettivazione regolativa" nel diritto del lavoro*, in *DRI*, 2019, 1, 111 ss.
- Perulli A., *La discriminazione algoritmica: brevi note introduttive a margine dell'Ordinanza del Tribunale di Bologna*, in *LDE*, 2021, n. 1.
- Peruzzi M., *La prova del licenziamento ingiustificato e discriminatorio*, Giappichelli, 2017.

- Rota A., *Rapporto di lavoro e big data analytics: profili critici e risposte possibili*, in *LLI*, 2017, 3(1), I, 32 ss.
- Ruffolo U., *Intelligenza artificiale, machine learning e responsabilità da algoritmo*, in *GI*, 2019, 7, 1657 ss.
- Salento A., *Digitalizzazione delle imprese e trasformazione delle competenze. Quadro analitico e riscontri empirici*, in *Labor*, 2019, 2, 131 ss.
- Sanguinetti G., *Machine Learning: accuratezza, interpretabilità e incertezza*, in *Ithaca: Viaggio nella Scienza*, 2020, 16, 71 ss.
- Santagata De Castro R., *Anti-discrimination Law in the Italian Courts: the new frontiers of the topic in the age of algorithms*, in *WP CSDL E*, It. n. 440/2021.
- Sartori A., *Il controllo tecnologico sui lavoratori*, Giappichelli, 2020.
- Schildt H., *The Data Imperative*, OUP, 2020.
- Selbst A. D. - Powles J., *Meaningful information and the right to explanation*, in *International Data Privacy Law*, 2017, 233 ss.
- Sitzia A. - Crafa S., *Impronte digitali, algoritmo e trattamento di dati personali: questioni di "law and technology"*, in *LG*, 2019, 3, 245 ss.
- Tebano L., *Tutela della privacy e potere di controllo del datore di lavoro tra l'ordinamento italiano e le fonti europee*, in *LLI*, 2017, 2, C-1 ss.
- Tullini P. (a cura di), *Controlli a distanza e tutela dei dati personali del lavoratore*, Giappichelli, 2017.
- Tullini P., *Divieto di indagini sulle opinioni e trattamenti discriminatori*, in M. Pedrazzoli (coordinato da) *Il nuovo mercato del lavoro*, Zanichelli, 2004, 145 ss.
- Tullini P., *La digitalizzazione del lavoro, la produzione intelligente e il controllo tecnologico nell'impresa*, in P. Tullini (a cura di), *Web e lavoro. Profili evolutivi e di tutela*, Giappichelli, 2017, 3 ss.
- Tullini P., *La salvaguardia dei diritti fondamentali della persona che lavora nella gig-economy*, in *costituzionalismo.it*, 2020, 1, 39 ss.
- Wachter S. - Mittelstadt B. - Floridi L., *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*, in *IDPL*, 2017, 7, 2 ss.
- Wachter S. - Mittelstadt B. - Floridi L., *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*, in *International Data Privacy Law*, 2017, 76 ss.