



Numero 3 / 2024

Marco PERUZZI

L'impatto dell'IA nella selezione del personale, negli annunci di lavoro mirati a filtrare le candidature, nella determinazione della "reputazione" della persona che lavora e nell'assegnazione dei compiti

L'impatto dell'IA nella selezione del personale, negli annunci di lavoro mirati a filtrare le candidature, nella determinazione della "reputazione" della persona che lavora e nell'assegnazione dei compiti

Marco PERUZZI

Associato di Diritto del Lavoro, Università di Verona

Abstract

Il contributo analizza il sistema di garanzie applicabili all'uso di sistemi di IA volti a selezionare i candidati all'assunzione, determinare la reputazione della persona che lavora e l'assegnazione dei compiti. L'indagine si confronta con un quadro regolativo in divenire. In particolare, partendo dal ruolo del GDPR nella ricostruzione delle tutele, cercherà di individuare l'impatto dell'AI Act e della direttiva sul lavoro su piattaforma digitale.

The essay analyses the system of safeguards applicable to the use of AI systems to select job candidates, determine the reputation of the person at work and the assignment of tasks. The investigation addresses an evolving regulatory framework. In particular, starting with the role of the GDPR in the reconstruction of protections, it explores the impact of the AI Act and the Platform Work Directive.

Premessa. Esempi di sistemi di IA e distinzioni.

Un sistema di intelligenza artificiale (IA) può costituire uno strumento che veicola o informa le scelte e l'esercizio di prerogative datoriali.

Può anzitutto assistere nelle varie fasi del processo di assunzione, aiutando a formulare gli annunci di lavoro, elaborando selezioni, classificazioni e graduatorie, raccogliendo e sistematizzando informazioni utili per il reclutamento, nonché supportando il percorso di inserimento dei neoassunti (c.d. *onboarding*).

Qureos, ad esempio, prevede tra le sue funzioni la generazione di *job description* customizzate e di domande *ad hoc* per il colloquio. Prevede, altresì, un sistema di *machine learning* per filtrare i curricula inviati per una posizione lavorativa o i profili professionali caricati su bacheche digitali come LinkedIn¹, sulla base degli elementi testuali ivi contenuti: il sistema si evolve sulla base dei

¹ Cfr. il *Codice di condotta per le Agenzie per il lavoro* (APL) di Assolavoro, approvato dal Garante con provvedimento dell'11 gennaio 2024 e pubblicato in gazzetta ufficiale a inizio marzo 2024. Tra le varie disposizioni, si prevede, ad esempio, che le agenzie possano raccogliere e trattare i dati personali inerenti al profilo presenti sui social di natura professionale relativi ai candidati, dovendo cancellare, in linea di principio, i dati raccolti durante il processo di selezione non appena sia evidente che non verrà fatta alcuna offerta di impiego o nel caso in cui l'offerta non venga accettata dal candidato. Inoltre, i dati relativi ai candidati trattati nell'ambito delle attività dell'agenzia, con esclusione della somministrazione di lavoro, possono essere conservati per un periodo di massimo quarantotto mesi, fatta salva la possibilità dell'interessato di poter chiedere in ogni

dati accumulati nel corso del processo di recruitment e adatta conseguentemente le proprie raccomandazioni. Lo stesso LinkedIn sta lanciando la piattaforma Recruiter, basata su approcci di IA generativa, per aiutare le imprese a vagliare e a trovare, all'interno del portale, i profili professionali che più rispondono alle proprie esigenze².

La piattaforma di interviste digitali Hire Vue permette di automatizzare il primo giro di colloqui con interviste video unidirezionali: il selezionatore imposta le domande in anticipo e i candidati registrano le risposte. In questo modo, Hilton dice di essere riuscita a ridurre il tempo del processo di assunzione da 43 giorni a soli 5 giorni. A settembre 2023 Hire Vue si è preoccupata di sfatare falsi miti che circolavano sul suo conto. «La verità?» - si legge sul sito web dell'azienda - «HireVue non utilizza il riconoscimento facciale. Le valutazioni video sono effettuate esclusivamente sul linguaggio del candidato [quindi, *voice* e *speech analytics*], in particolare sul modo in cui parla delle esperienze passate relative alle competenze rilevanti per la posizione lavorativa in questione».

Paradox fornisce una applicazione per pubblicare annunci di lavoro sulle varie bacheche digitali di rilievo presenti sul web e una chatbot, che interagisce con i candidati per gestire il processo di *recruitment*: lo stesso candidato può iniziare, interrompere e riprendere il processo quando vuole.

Funzionando come una piattaforma di incontro tra domanda e offerta di lavoro, la start-up italiana Joinrs (già Tutored) raccoglie le preferenze di giovani in cerca di stage/lavoro e le offerte delle aziende, ricostruendo per l'utente gli annunci che potrebbero interessargli. Per ogni annuncio sono generate due percentuali: una relativa all'affinità con le preferenze inserite, l'altra alla compatibilità del profilo con quello ricercato dall'azienda.

Il sistema Cloud Talent Solution, prodotto da Google, è stato utilizzato da Johnson & Johnson per automatizzare, migliorare e ridurre nei tempi una parte del processo di assunzione. In particolare, ha migliorato la fase di matching tra chi ricerca lavoro e gli annunci presenti sul sito dell'azienda³. Sempre Johnson & Johnson utilizza il sistema Textio per localizzare ed eliminare i bias presenti nei suoi annunci di lavoro: «abbiamo riscontrato che molti dei nostri annunci di lavoro si orientavano al maschile» - riporta l'azienda sul proprio sito - «ma quando abbiamo iniziato a modificare le descrizioni con Textio, abbiamo visto un aumento delle candidature femminili del 9% ». Il sistema Textio aiuta anche a tracciare i *bias* nelle valutazioni scritte (*performance feedback*) dei dipendenti effettuate dal management.

Il profilo da ultimo richiamato sposta invero il piano dell'osservazione dall'uso dell'IA per automatizzare un processo all'uso dell'IA per analizzare e verificare le modalità di funzionamento e gli *outcome* del processo stesso, mapparne i rischi e intervenire con misure di riduzione o mitigazione degli stessi.

momento la cancellazione. I dati trattati nell'ambito dei rapporti di somministrazione di lavoro possono essere, invece, conservati per un periodo di undici anni dalla cessazione del rapporto di lavoro.

² Cfr. L. Valente, *Così l'intelligenza artificiale rivoluziona le politiche attive del lavoro*, su *lavoce.it*, 21 maggio 2024, con riguardo a quanto previsto dall'art. 26, d.l. 7 maggio 2024, n. 60. In merito al funzionamento del Sistema informativo per l'inclusione sociale e lavorativa (SIISL), il c.d. "decreto coesione" stabilisce, in particolare, che i datori di lavoro possano pubblicare sul Sistema le posizioni vacanti all'interno dei loro organici e che all'interno dello stesso siano inserite anche le posizioni vacanti pubblicate dai datori di lavoro su piattaforme pubbliche nazionali e internazionali. Al fine di favorire l'incontro tra domanda e offerta di lavoro, si prevede che il Sistema utilizzi, nei limiti consentiti dalle disposizioni vigenti, gli strumenti di intelligenza artificiale per l'abbinamento ottimale delle offerte e delle domande di lavoro ivi inserite.

³ Cfr. B. Marr, *Data-driven HR*, KoganPage, London, 2024.

I sistemi di IA possono essere in tal senso utilizzati in termini più ampi e complessivi per valutare la *fairness* e/o la conformità regolatoria dei processi automatizzati.

Un esempio significativo è Credo AI, azienda che sul proprio sito si presenta pronta a fornire strumenti di IA a supporto della *compliance* alle regole dell'AI Act europeo, ma che già fornisce policy pack per il proprio software di governance ritagliati su alcuni quadri regolativi in vigore, come ad esempio quello della NYC Local Law No. 144/2011, che sarà analizzato nel corso della trattazione.

Proseguendo, invece, nella ricostruzione del possibile uso dell'IA nell'esercizio di prerogative datoriali, si può evidenziare l'ipotesi di processi automatizzati volti ad attribuire un punteggio alla persona che lavora, con determinazione di un profilo reputazionale e conseguenze sul piano dello svolgimento del rapporto di lavoro.

Sono noti, per l'ordinamento italiano, i casi degli algoritmi utilizzati dalle piattaforme digitali di food delivery Deliveroo e Foodhino per regolare il processo di prenotazione degli slot di lavoro da parte dei rider (ossia la gestione del loro accesso al lavoro). Per semplificare, l'algoritmo deduceva dall'informazione relativa alle sessioni di lavoro prenotate e non cancellate dal rider nel termine previsto dal contratto un determinato punteggio e una connessa limitazione della possibilità di accesso al sistema per la prenotazione di ulteriori sessioni di lavoro.

Betterworks, Leapsome, Quantive sono aziende che forniscono c.d. OKR software (Objectives + Key Results), piattaforme che consentono di individuare degli obiettivi, tracciare lo stato di avanzamento degli stessi, rendere visibile il contributo di ciascun dipendente, in alcuni casi con approcci di *gamification*.

Bob è un sistema che permette di misurare il tasso di abbandono dei dipendenti. La presentazione del software chiarisce che l'indicatore si basa su trend statistici e, quindi, non assicura di rappresentare l'effettivo pensiero del lavoratore. La presentazione individua anche alcuni dei fattori che possono incidere sul tasso di abbandono: *work-life balance*, il coinvolgimento (il c.d. *engagement*), le opportunità di crescita. Si specifica, a titolo esemplificativo, che i dipendenti con più di 5 richieste di *time-off*/permesso/congedo negli ultimi 90 giorni presentano un alto rischio di abbandono, a prescindere dal numero di giorni di ogni richiesta.

Un'ulteriore possibilità di utilizzo dell'IA da parte del datore riguarda la gestione del *workflow*. Nel definire la distribuzione dei carichi di lavoro e la configurazione dell'assetto aziendale, anche a partire da una previsione della domanda di beni e servizi e della scelta tecnico-organizzativa più efficiente, il sistema può consegnare una decisione sull'estensione o collocazione spaziale della prestazione ovvero sulla collocabilità del lavoratore.

Autodesk, ad esempio, è una delle tante aziende che fornisce software di *Manufacturing Execution System* (MES), ossia software, oggi integrati con approcci di *machine learning*, che digitalizzano la catena di produzione, con gestione e monitoraggio del reparto. Sul sito dell'azienda, si presentano le varie funzioni di Prodsmart: «definisce le priorità ed evita i sovraccarichi», «ottimizza il flusso di lavoro», «monitora le prestazioni di ciascun operaio in tempo reale», «specifica l'allocazione delle risorse del team, nonché le attività che devono svolgersi, dove, quando e come», «allinea la gestione dei turni con l'attività del reparto di produzione».

Si consideri, da ultimo, la possibilità di traslare integralmente l'esercizio delle prerogative datoriali all'interno di una dimensione di realtà virtuale, come il Metaverso. Si può richiamare l'esempio di

Accenture, che per ovviare ai limiti posti durante la fase pandemica ha cominciato già nel 2022 a gestire i propri processi di *recruitment* all'interno della dimensione virtuale.

I sistemi di IA descritti utilizzano approcci algoritmici molteplici e differenti. Ciò che è importante sottolineare, ai fini dell'analisi che ci occupa, è la possibilità di distinguere tra approcci *logic-based* e approcci di *machine learning*. Entrambi rientrano all'interno dell'ampia definizione di IA prevista dal Regolamento sull'IA (v. art. 3 e Cons. n. 12). La distinzione è rilevante soprattutto per capire i termini e i diversi gradi con cui si possono configurare, all'interno delle due categorie, i rischi per i diritti dei lavoratori - con riguardo anzitutto all'opacità del processo e alla determinazione di *bias* discriminatori - e quindi le necessità che si pongono nella costruzione delle correlate garanzie di trasparenza e sorveglianza umana.

Per comprendere la differenza, può essere utile porre a confronto due noti casi di discriminazione algoritmica.

Il primo è quello 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 dai Tribunali di Bologna e di Palermo, l'applicazione indifferenziata dei parametri di valutazione da parte dell'algoritmo non esclude 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 disabilità⁴. Nel caso descritto, il sistema è governato da un algoritmo *logic-based*, ossia da un algoritmo che 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 implicati⁵, il sistema rimane di carattere deterministico (quindi 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 è annidata nella regola "apparentemente neutra" tradotta in linguaggio di programmazione.

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.

A essere utilizzato, in questo caso, 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

⁴ Trib. Bologna, 31 dicembre 2020, in *RIDL* 2021, II, 188 con nota di G. Gaudio; Trib. Palermo, 17 novembre 2023, in *RIDL*, 2023, II, 763 con nota di V. Nuzzo.

⁵ Lo dimostra Trib. Torino, 5 agosto 2023, laddove affronta il problema della trasparenza dell'algoritmo utilizzato da Foodinho.

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)⁶, quanto la possibilità che le correlazioni (anche imprevedibili) che l'algoritmo di *machine learning* individua all'interno dei dataset consegnati nascondano degli impatti discriminatori. Come evidenza la dottrina, «il punto fondamentale è che il modello creato da un algoritmo di *deep learning* non è facilmente dissezionabile, non è decomponibile come le linee di codice di algoritmi tradizionali. [...] Possiamo provare a osservare dall'esterno, a fare delle domande, interrogare l'algoritmo e analizzare le sue risposte»⁷. Entra in gioco, in questo caso, l'investigazione *ex post* attraverso i c.d. metodi di *Explainable AI – XAI*, o anche *FairXAI*, ossia altre tecniche, anch'esse algoritmiche, volte a spiegare l'*outcome*, ma anche ad esaminare la *fairness* del sistema. Sulle prospettive che apre quest'esigenza di osservazione dei risultati sia in termini di modelli di valutazione/gestione dei rischi, obblighi di proceduralizzazione e trasparenza, nonché applicazione di modelli algoritmici a supporto delle garanzie di tutela si tornerà oltre nella trattazione.

La (ri)costruzione delle garanzie nel quadro normativo in divenire

Per verificare possibilità e limiti di utilizzo di questi e altri sistemi di intelligenza artificiale sul lavoro è necessario confrontarsi con una pluralità di fonti, anche di matrice non lavoristica. Tra queste si inseriranno a breve l'AI Act e la proposta di direttiva in materia di lavoro su piattaforma digitale. In particolare, 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. Questo secondo passaggio è tanto più necessario se si considera il carattere orizzontale del Regolamento sull'IA, 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.

Per quanto riguarda il primo profilo d'analisi indicato, 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 rischio e l'individuazione della trasparenza e supervisione umana quali garanzie imprescindibili di una prospettiva antropocentrica, ossia di una prospettiva che esclude il riconoscimento allo strumento tecnologico di una autonoma capacità giuridica e richiede, invece, all'uomo di rimanere responsabile.

⁶ Parla al riguardo di «algoritmo "specchio"» A. Lo Faro, *Algorithmic Decision Making e gestione dei rapporti di lavoro: cosa abbiamo imparato dalle piattaforme*, in *federalismi.it*, n. 25, 2022, 189 ss.

⁷ G.F. Italiano-S. Civitarese Matteucci-A. Perrucci, *L'intelligenza artificiale: dalla ricerca scientifica alle sue applicazioni. Una introduzione di contesto*, in A. Pajno-F. Donati-A. Perrucci (a cura di), *Intelligenza artificiale e diritto: una rivoluzione? Diritti fondamentali, dati personali e regolazione*, vol. I, Bologna, 2022, 54.

Sono scelte che trovano specifica declinazione nell'AI Act⁸.

L'implementazione del modello *risk-based* all'interno del Regolamento, volto appunto a «introdurre un insieme proporzionato ed efficace di regole vincolanti» (Cons. n. 26), prevede una prima categorizzazione tra rischio inaccettabile, alto e non-alto. A tali distinzioni si connettono rispettivamente la proibizione di immissione nel mercato, messa in servizio e/o utilizzo del sistema (art. 5), la previsione di una serie di requisiti e obblighi da rispettare (Capo III), l'adozione modulata e meramente volontaria delle garanzie attraverso codici di condotta (art. 95). 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 (ad es. le chatbot) e per quelli a carattere generativo, compresi i sistemi di AI per finalità generali (general purpose AI – GPAI, come ad es. ChatGpt), a tutela di chi sia esposto ai c.d. *deep fakes* (Capo IV, Art. 50).

Per quanto attiene alla categoria del rischio inaccettabile, nella prospettiva lavoristica può essere interessante segnalare il divieto di 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 sicurezza⁹ (si pensi ai sistemi di tracciamento oculare di rilevamento della fatica e distrazione al volante, forniti ad es. da Caterpillar), ovvero il divieto di 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.

Come detto, è alla categoria dei sistemi ad alto rischio che si rivolgono i requisiti e gli obblighi previsti dal Regolamento. La classificazione del sistema come “ad alto rischio” non deriva soltanto – come era invece previsto nella proposta iniziale – dalla sua inclusione all'interno di un determinato elenco di aree critiche e casi d'uso (art. 6, par. 2), 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 ci occupa, è importante evidenziare due profili.

Da un lato, i sistemi di IA destinati a essere utilizzati per il riconoscimento delle emozioni (nella misura in cui il pertinente diritto Ue o nazionale ne permetta l'uso: punto 1, Allegato III) e quelli 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. Per quanto attiene a quest'ultimi, trattasi, segnatamente, dei sistemi utilizzati «per l'assunzione o la selezione di persone fisiche, in particolare per pubblicare annunci di lavoro mirati, analizzare o filtrare le candidature e valutare i candidati»; nonché «per adottare decisioni riguardanti le condizioni dei rapporti di lavoro, la promozione o cessazione dei rapporti contrattuali di lavoro, per assegnare compiti sulla base del comportamento individuale o dei tratti e delle caratteristiche personali o per monitorare e valutare le prestazioni e il comportamento delle persone nell'ambito di tali rapporti di lavoro».

⁸ Cfr. P. Loi, *Il rischio proporzionato nella proposta di regolamento sull'IA e i suoi effetti nel rapporto di lavoro*, n. 29, 2023, 171 ss.; M. Barbera, “*La nave deve navigare*”. *Rischio e responsabilità al tempo dell'impresa digitale*, in *Labour & Law Issues*, n. 9(2), 2023, 1 ss.; A. 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.

⁹ La deroga è stata introdotta nel testo finale di compromesso raggiunto in sede di negoziato interistituzionale. Non era, invece, presente nelle proposte di emendamento indicate dal Pe nella Risoluzione legislativa del 14 giugno 2023. Cfr., con riferimento a quest'ultima, L. Tebano, *Poteri datoriali e dati biometrici nel contesto dell'AI Act*, in *federalismi.it*, n. 25, 2023, 198 ss.

Dall'altro lato, a norma dell'art. 6, par. 3, del Regolamento, il suddetto rischio significativo di danno è escluso laddove il sistema non influenzi materialmente il risultato del processo decisionale¹⁰, mentre è da ritenersi sempre presente qualora il sistema effettui profilazione di persone fisiche. Considerato che quest'ultima condizione è molto probabile che si verifichi in caso di utilizzo di processi automatizzati di monitoraggio o decisionali sul lavoro, può ritenersi 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 (come la maggior parte di quelli descritti in premessa)¹¹.

Nel quadro di regole dell'AI Act, è il fornitore a essere gravato, prima, dall'obbligo di verificare se il sistema debba essere classificato come "ad alto rischio", quindi, in tal caso, dall'obbligo di mappare e valutare i rischi e di predisporre un sistema di misure per la loro gestione (art. 9)¹². Rispetto alla qualità dei dati di addestramento, convalida e prova (molto importante, come visto, nel caso di *machine learning*), 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*, in particolare quelli che possono «comportare discriminazioni vietate dal diritto dell'Unione», nonché misure volte a localizzarli, prevenirli e mitigarli (art. 10). 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 (ex dati sensibili) 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.

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*¹³, ma anche

¹⁰ È quanto può accedere laddove il sistema si limiti a svolgere un compito procedurale ristretto oppure a migliorare il risultato di un'attività umana precedentemente completata, ad esempio classificando documenti, rilevando duplicati o migliorando lo stile del linguaggio ivi utilizzato; ovvero laddove si limiti a rilevare modelli decisionali o deviazioni da modelli decisionali precedenti e non sia destinato a sostituire o influenzare la valutazione umana precedentemente completata, senza un'adeguata revisione umana (ad esempio effettua una mera verifica *ex post* della coerenza della valutazione operata da un docente rispetto un certo modello predefinito); ovvero laddove svolga un compito soltanto preparatorio per una successiva valutazione, come la gestione dei file, con funzioni di indicizzazione, ricerca ed elaborazione del testo (Cons. n. 53).

¹¹ Cfr. S. Marassi, *Intelligenza artificiale e sicurezza sul lavoro*, in M. Biasi (a cura di), *Diritto del lavoro e intelligenza artificiale*, Giuffrè, 2024, p. 207 ss. L'autrice significativamente evidenzia che l'elenco delle aree critiche collegate all'ambiente di lavoro fa riferimento a specifiche finalità d'uso, elemento tutt'altro che secondario nel contesto regolativo dell'AI Act, come segnalato dal Cons. n. 52. La mancanza di un esplicito riferimento a finalità di tutela della salute e sicurezza interroga sui termini in base ai quali sistemi utilizzati a tale scopo possano essere attratti nella classificazione ad alto rischio. Si può ragionare sulla importanza che potrebbe assumere, ai fini della classificazione, la stretta connessione che si dovesse instaurare tra la finalità di sicurezza, per cui è adottato il sistema, e alcune funzioni indicate nell'area critiche dell'allegato III, come le decisioni sulle condizioni dei rapporti di lavoro o l'assegnazione dei compiti. Rimane ferma la possibilità di classificare il sistema come ad alto rischio a norma dell'art. 6, par. 2-3, laddove si tratti di sistema di riconoscimento delle emozioni (ad es. la fatica, la distrazione, il dolore, v. *supra* nel testo) ovvero anche in base all'art. 6, par. 1, qualora sia componente di sicurezza di prodotti o esso stesso prodotto e in forza di una delle fonti Ue elencate nell'all. I, tra cui rientra anche il nuovo Regolamento macchine 2023/1230, sia sottoposto a una procedura di valutazione di conformità da parte di organismi terzi.

¹² Trattasi, in particolare, dei rischi noti e ragionevolmente prevedibili in condizioni di corretto utilizzo del sistema di IA ovvero di quelli che potrebbero emergere in caso di un suo uso improprio ragionevolmente prevedibile. Si specifica, d'altra parte, che i rischi interessati sono «solo quelli che possono essere ragionevolmente attenuati o eliminati attraverso lo sviluppo o la progettazione del sistema di IA ad alto rischio o la fornitura di informazioni tecniche adeguate» (art. 9, par. 3).

¹³ U. 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 A. Pajno-F. Donati-A. Perrucci (a cura di), *Intelligenza artificiale e*

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 stesso¹⁴.

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 consegna¹⁵. 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'uso¹⁶ (art. 26) ed è soggetto, a sua volta, a un obbligo di valutazione dell'impatto del sistema sui diritti fondamentali (art. 27). Quest'ultimo 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 (requisito molto importante in caso di adattabilità e apprendimento continuo del sistema), 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

diritto: una rivoluzione? Amministrazione, responsabilità, giurisdizione, vol. II, il Mulino, Bologna, 2022, 247). Nella proposta di nuova direttiva sulla responsabilità per danno da prodotti difettosi, approvata nel testo di compromesso interistituzionale dal Pe il 12 marzo 2024, si specifica al Cons. n. 13 che «il produttore o lo sviluppatore di software, compreso il fornitore di sistemi di IA ai sensi del regolamento [sull'IA], dovrebbe essere considerato un fabbricante». Il sistema normativo ivi offerto costituisce, ad oggi, l'unica risposta regolativa Ue alla questione della responsabilità del fornitore di sistema di IA per danni causati dallo stesso: il percorso di approvazione della parallela proposta di direttiva in tema di IA e responsabilità civile si è, infatti, arenato. Questa seconda direttiva sarebbe peraltro intervenuta anche sul tema della responsabilità del *deployer*, lasciato, quindi, per ora, agli schemi normativi di responsabilità ricavabili da altre fonti Ue o interne. Può essere significativo segnalare come all'interno della proposta di direttiva in materia di *product liability*, tra le modifiche volte ad agevolare il danneggiato sotto il profilo probatorio, l'art. 10, par. 4, introduce «l'elemento di maggiore innovazione ma certamente destinato a generare profonde incertezze e numerose discrepanze a livello applicativo»; la disposizione permette, infatti, al giudice di presumere il carattere difettoso del prodotto, il nesso causale con il danno o entrambi, «in ragione della "complessità tecnica o scientifica" del prodotto stesso» (A. Bertolini, *La responsabilità civile derivante dall'utilizzo di sistemi di intelligenza artificiale: il quadro europeo*, in M. Biasi (a cura di), *Diritto del lavoro e intelligenza artificiale*, Giuffrè, 2024, 55 ss., spec. 71).

¹⁴ L. 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.; A. Agrawal, J. Gans, A. Goldfarb, *Macchine predittive*, Franco Angeli, 2018.

¹⁵ Cfr. S. Ciucciovino, *La disciplina nazionale sulla utilizzazione della intelligenza artificiale nel rapporto di lavoro*, in *Lavoro Diritti Europa*, 2024, n. 1, 1 ss.

¹⁶ 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. A norma 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.

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* in forza 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. Il profilo è tanto più rilevante, nella prospettiva d'analisi che ci occupa, considerato il ruolo centrale che quest'ultimo quadro regolativo assume nella ricostruzione delle garanzie di tutela del lavoratore di fronte all'uso datoriale di sistemi automatizzati.

Prima di affrontare questo profilo, pare opportuno ribadire come a caratterizzare il Regolamento non sia soltanto la natura orizzontale e complementare, ma anche il carattere minimo del sistema di garanzie apprestato: al riguardo, è importante segnalare che a norma dell'art. 2, par. 11, il Regolamento 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 AI da parte dei datori di lavoro, o di incoraggiare o consentire l'applicazione di contratti collettivi più favorevoli. Guardando al contesto italiano, si può richiamare, con riferimento all'intervento sul tema da parte della contrattazione collettiva, la recente ipotesi di piattaforma per il rinnovo del Ccnl Industria metalmeccanica e dell'installazione di impianti (in scadenza il 30 giugno 2024). Si prevede, in particolare, la definizione di un'apposita normativa in materia di intelligenza artificiale e tutela delle lavoratrici e dei lavoratori, « per evitare un utilizzo distorto che determini forme di abusi o violazioni normative/contrattuali e favorire la partecipazione alla definizione della governance dei dati e del loro utilizzo (ad esempio, protezione, conservazione dei dati, diritto informazione, ecc.) e degli algoritmi per limitare il loro impatto negativo e condividere i benefici dell'intelligenza artificiale ». Può essere significativo richiamare anche 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.

Il ruolo del GDPR

Rinviando 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 a norma dell'art. 35 GDPR sia inevitabile se

il datore di lavoro utilizza sistemi decisionali o di monitoraggio automatizzati¹⁷. 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 coinvolte¹⁸. 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), a quelle previste dalla proposta di direttiva sul lavoro su piattaforma digitale (v. *infra*), ovvero a quelle stabilite a livello interno dall'art. 1-bis d.lgs. 152/97 (per i processi decisionali e di monitoraggio integralmente automatizzati), che da tali fonti e dal Regolamento sull'IA trae ispirazione¹⁹. Peraltro, per rispondere a un'attenta riflessione maturata in dottrina sul punto, può ritenersi che proprio perché identificate all'esito di una valutazione di impatto su tutti i diritti e le libertà dei lavoratori interessati, le garanzie di tutela non potrebbero fermarsi a una mera informativa sul modo di operare dell'IA, ma dovrebbero assicurare anche la «spiegabilità degli effetti e dell'impatto», il «*cosa fa*» la IA²⁰. È una prospettiva che sarà agevolata e rafforzata dal prossimo raccordo con l'*AI Act* considerate, come visto, la garanzia di interpretabilità dell'*output* da parte del *deployer* nonché l'esplicita previsione del diritto di qualsiasi persona esposta a decisioni filtrate da sistemi di IA di ottenere una spiegazione sul ruolo di tali sistemi nella procedura decisionale e sugli elementi principali della decisione presa (art. 86)²¹. Particolarmente

¹⁷ 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.

¹⁸ Risulta, al riguardo, particolarmente significativa la segnalazione di F. Pizzetti, *La protezione dei dati personali e la sfida dell'Intelligenza Artificiale*, in F. Pizzetti (a cura di), *Intelligenza artificiale, protezione dei dati personali e regolazione*, Giappichelli, 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).

¹⁹ Si rilevi che le misure da adottare a seguito di DPIA sono poste a tutela dei soggetti a cui si riferiscono i dati personali trattati, a prescindere quindi dalla tipologia del rapporto di lavoro; le caratteristiche del rapporto di lavoro possono rilevare, d'altra parte, ai fini della configurazione dell'obbligo di DPIA e della modulazione delle misure, in particolare la vulnerabilità dell'interessato, individuata nelle linee guida in capo ai dipendenti, e la presenza di uno squilibrio di potere. Come si vedrà oltre nella trattazione, anche le garanzie di trasparenza e sorveglianza umana/riesame umano previste dalla direttiva sul lavoro su piattaforma digitale sono estese ai lavoratori autonomi. L'estensione a (alcuni) lavoratori autonomi è presente, altresì, a livello interno, nell'art. 1-bis d.lgs. 152/97. Il comma 7 della disposizione prevede, infatti, 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. 81/2015. A norma 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. Il richiamato accordo collettivo nazionale sugli shopper prevede, all'art. 5, obblighi di informazione a tutela dei lavoratori e delle rappresentanze sindacali in caso di utilizzo di sistemi (non solo integralmente) automatizzati. Per un'analisi dell'art. 1-bis cfr., *ex multis*, E. Dagnino, *Il diritto interno: i sistemi decisionali e di monitoraggio (integralmente) automatizzati tra trasparenza e coinvolgimento*, in M. Biasi (a cura di), *Diritto del lavoro e intelligenza artificiale*, Giuffrè, 2024, 147 ss.; A. Donini, *Informazione sui sistemi decisionali e di monitoraggio automatizzati: poteri datoriali e assetti organizzativi*, in *DLM*, 2023, 85; A. Zilli, *La via italiana per condizioni di lavoro trasparenti e prevedibili*, in *DRI*, 2023, 1, p. 25.

²⁰ S. Ciucciovino, *La disciplina nazionale sulla utilizzazione della intelligenza artificiale*, cit., pag. 15.

²¹ Pare utile ricordare, sempre in seno all'*AI Act*, l'obbligo per i *deployer* che utilizzano sistemi che prendono decisioni o assistono nel prendere decisioni relative a persone fisiche, di informare le stesse del fatto che sono soggette all'uso di tali sistemi (Art. 26, par. 11, fatti salvi gli ulteriori obblighi di trasparenza previsti dall'art. 50 per i sistemi interattivi e di IA

rilevante risulta, rispetto a questa fase della DPIA, anche l'innesto *de iure condendo* dei contenuti di cui all'art. 13, a norma del quale la documentazione pertinente e le istruzioni per l'uso consegnate al datore, in qualità di *deployer*, devono 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.

Ragionare di valutazione di impatto nella prospettiva di trattamenti automatizzati effettuati attraverso sistemi di IA consente di evidenziare quanto le garanzie di trasparenza non bastino, soprattutto in caso di approcci di apprendimento automatico profondo. Devono essere affiancate da garanzie di sorveglianza umana, non solo nel senso del coinvolgimento dell'operatore umano nel ciclo di vita del sistema, ma anche sul piano più complessivo della valutazione e gestione dei rischi, a livello di verifica periodica di impatto dei processi decisionali (integralmente o parzialmente) automatizzati sui diritti e sulle libertà dei soggetti esposti a dette decisioni, nonché a livello di qualità di dati (anche derivati e inferiti). Anche questa prospettiva uscirà rafforzata dal raccordo con l'*AI Act*: come detto, tutti i *deployer* di sistemi ad alto rischio sono, infatti, 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); soprattutto, sono chiamati 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).

Si consenta, infine, di ricordare che la DPIA include «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 i diritti e le libertà coinvolti²². 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 – c.d. *FairXAI*),

generativa), nonché l'obbligo per i datori di lavoro di informare i rappresentanti dei lavoratori e i lavoratori interessati che saranno soggetti al sistema (art. 26, par. 7). La previsione di un obbligo di informazione delle rappresentanze, per quanto limitato nel contenuto, apre il varco al percorso consentito a livello interno dall'art. 1-*bis* nell'individuazione di un diritto di informazione del sindacato, distinto e ulteriore rispetto a quello dei lavoratori. Si noti che il d.d.l. di iniziativa governativa in materia di IA prevede lo stesso obbligo informativo datoriale contenuto nell'*AI Act*, disponendo però una inspiegabile (e peraltro irrilevante, vista la futura applicazione diretta del Regolamento) limitazione dello stesso ai sistemi di cui all'art. 1-*bis*, ossia quelli integralmente automatizzati (art. 10, A.S. n. 1146).

²² Cfr. H. Janssen-M. Seng Ah Lee-J. Singh, *Practical fundamental rights impact assessments*, in *International Journal of Law and Information Technology*, 2022, 30, 200.

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.

Raccordi, norme speciali di dettaglio, possibili prospettive, sguardi oltreoceano.

Come si è potuto comprendere dalla breve analisi svolta nel paragrafo precedente, l'individuazione delle tutele del lavoratore di fronte all'uso datoriale di algoritmi richiede di ricostruire, lungo i meridiani delle garanzie della trasparenza e della sorveglianza umana, la complessa architettura regolativa coinvolta, analizzando i punti di integrazione e raccordo tra le fonti.

Tali raccordi confermano, anzitutto, la natura complementare dell'*AI Act*, necessitando in tal senso di specifica attenzione ai fini dell'individuazione delle posizioni d'obbligo configurabili in capo al datore in qualità di *deployer* di un sistema di IA: è quanto si può evidenziare, ad esempio, partendo dalle osservazioni presentate a dicembre 2023 da parte del garante della privacy italiano rispetto all'utilizzo datoriale di servizi informatici di gestione della posta elettronica forniti da terzi²³.

In tale prospettiva, si evidenzia come il datore, in qualità appunto di *deployer*, non sia un mero destinatario passivo delle istruzioni consegnate dal fornitore, ma debba invece verificare che le impostazioni del sistema gli consentano di operare in conformità alla normativa applicabile - anzitutto il GDPR, ma anche la normativa statutaria ad es. sul divieto di indagine sulle opinioni personali - dovendo se necessario dare a sua volta istruzioni al fornitore affinché la configurazione dello strumento digitale sia in linea con le sue esigenze di *compliance*. Del resto, come visto, è lo stesso Regolamento sull'IA a richiedere al *deployer* di monitorare il funzionamento del sistema e sospenderne eventualmente l'utilizzo (art. 26, v. *supra*).

La ricostruzione del quadro di tutele richiede, altresì, di analizzare – *de iure condito e condendo* - le norme speciali di protezione dei lavoratori sia a livello di diritto Ue sia a livello interno.

Per quanto riguarda il primo livello di osservazione, si possono in questa sede brevemente richiamare alcune disposizioni previste nella proposta di direttiva in materia di piattaforme digitali²⁴, quali regole di dettaglio *ex art. 88 GDPR*, in tema di trasparenza dei processi automatizzati, nonché sorveglianza umana e riesame umano delle decisioni prese o (anche solo) *sostenute* dagli stessi²⁵.

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

²³ Provvedimento del 21 dicembre 2023 - Documento di indirizzo "Programmi e servizi informatici di gestione della posta elettronica nel contesto lavorativo e trattamento dei metadati". L'ampia discussione accesa da questo documento ha portato il Garante ad avviare una consultazione pubblica sulla questione della congruità del termine di conservazione dei metadati ivi indicato e a differire, conseguentemente, l'efficacia del provvedimento (v. il provvedimento 22 febbraio 2024, n. 127 e l'avviso pubblico di avvio della consultazione sul "termine di conservazione dei metadati generati e raccolti automaticamente dai protocolli di trasmissione e smistamento della posta elettronica").

²⁴ Il testo di compromesso, su cui Consiglio e Parlamento hanno faticosamente raggiunto un accordo a marzo 2024, è stato poi approvato dal Pe il 24 aprile.

²⁵ Si noti che, come in forza dell'art. 10, par. 5, della direttiva, possa essere solo sostenuta (e non anche presa) da un sistema automatizzato «qualsiasi decisione di limitare, sospendere o chiudere il rapporto contrattuale o l'account di una persona che svolge un lavoro mediante piattaforme digitali o qualsiasi altra decisione di pregiudizio equivalente».

dei sistemi a rischio inaccettabile: si pensi, in particolare, al divieto di trattare dati relativi allo stato emotivo o psicologico (ma senza, si noti, la deroga per le ragioni mediche o di sicurezza, inserita invece nel testo di compromesso finale dell'*AI Act*) 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 sessuale²⁶. 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), l'art. 9 prevede specifici obblighi di trasparenza (a favore di lavoratori autonomi e subordinati), 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, a norma 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, 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 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 a norma 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)²⁷ e possono esercitare i diritti

²⁶ 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.

²⁷ 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

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, in forza dell'art. 28, alla possibilità di introdurre disposizioni diverse), si ponga l'accento sulla questione dei rischi psicosociali, della possibile pressione determinata dai sistemi automatizzati e della tutela della salute mentale, nonché sulla necessità di garantire la protezione contro violenza e molestie, richiedendosi che, a livello interno, si preveda l'obbligo per le piattaforme di adottare misure preventive, compresi canali di segnalazione efficaci.

A livello interno, si possono richiamare, oltre al già citato art. 1-bis d.lgs. 152/97, anche l'art. 111 bis del d.lgs. 196/03, secondo cui le informazioni di cui all'art. 13 GDPR, «nei casi di ricezione dei curricula spontaneamente trasmessi dagli interessati al fine della instaurazione di un rapporto di lavoro, vengono fornite al momento del primo contatto utile, successivo all'invio del curriculum medesimo [ma rispetto all'impatto della direttiva 2023/970 sulle informazioni dovute in caso di sistemi di *recruitment* automatizzati v. *infra*]. Nei limiti delle finalità di cui all'articolo 6, paragrafo 1, lettera b), del Regolamento [ossia, «il trattamento è necessario all'esecuzione di un contratto di cui l'interessato è parte o all'esecuzione di misure precontrattuali adottate su richiesta dello stesso»], il consenso al trattamento dei dati personali presenti nei curricula non è dovuto».

In una logica di tutela volta anzitutto a verificare i possibili rischi di discriminazione all'interno dei sistemi automatizzati può essere utile richiamare, come possibile prospettiva di sviluppo delle garanzie, alcune soluzioni apprestate dalla direttiva 2023/970 in materia di trasparenza retributiva di genere.

La direttiva invita, infatti, a non concentrarsi soltanto sulla soluzione delle difficoltà che si possono raccogliere, a valle, sul piano della prova della (in)sussistenza di una discriminazione (soluzioni quali l'inversione completa dell'onere probatorio in caso di violazione degli obblighi informativi e 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), 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.

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.

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.

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.

La 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. *FairXAI*.

A tal ultimo riguardo, si è già potuto anticipare in premessa come alcune aziende, ad esempio, Credo AI, forniscano strumenti di IA a supporto della *compliance* regolatoria. Se la prospettiva è quella di garantire anche un modello di governance digitale rispetto alle regole dell'AI Act europeo, l'azienda già fornisce *policy pack* ritagliati su alcuni quadri regolativi in vigore, anzitutto quello, molto interessante ai fini della nostra analisi, previsto dalla legge della Città di New York n. 144/2011, entrata in vigore il 6 luglio 2023²⁸.

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.

Qualora il processo automatizzato sia 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

²⁸ Trattasi di una legge, approvata dal New York City Council, volta a modificare l'*Administrative code of the city of New York*. In particolare, introduce all'interno di tale codice, segnatamente all'interno del Capo V «*Unfair Trade Practices*» del Titolo 20 «*Consumer and Worker Protection*», un sub-chapter dedicato all'uso di sistemi automatizzati nell'ambito del lavoro.

Opportunity Commission (EEOC) ai datori di lavoro con almeno 100 dipendenti o agli appaltatori federali con almeno 50 dipendenti, al ricorrere di determinati criteri²⁹.

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*?

²⁹ Si rinvia, sul punto, al 2023 EEO-1 Component 1 Data Collection Instruction Booklet.