

Marco Peruzzi

Intelligenza artificiale e lavoro

Uno studio su poteri datoriali e tecniche di tutela



Giappichelli

INTRODUZIONE

Nel romanzo *L'uomo senza qualità* di Robert Müsil, si racconta di una coppia, un uomo e una donna, che mentre camminano per strada si fermano all'improvviso vedendo davanti a sé un assembramento.

«Come api intorno al buco dell'arnia la gente s'era accalcata lasciando un vuoto nel mezzo».

Nel mezzo ci sta la scena di un incidente stradale: il camionista sceso dalla cabina, «grigio come carta da pacchi», e il malcapitato per terra, «come morto» sull'orlo del marciapiede.

L'uomo, a quel punto, dice alla donna: «– In questi autocarri pesanti che usano qui da noi il freno ha la corsa troppo lunga». La donna trae da queste parole «un senso di sollievo»: «le bastava che con ciò l'orribile incidente fosse in qualche modo sistemato e diventasse un problema tecnico che non la riguardava più da vicino». L'uomo aggiunge: «– Secondo le statistiche americane, [...] negli Stati Uniti centonovantamila persone all'anno rimangono uccise e quattrocentocinquantomila ferite in incidenti automobilistici. – Crede che sia morto?», chiede la donna, ancora con «l'ingiustificata sensazione di aver vissuto una vicenda eccezionale»¹.

Come direbbe Evelyn Wang, la protagonista del film *Everything Everywhere All at Once*: «É solo una ineluttabilità statistica: nulla di speciale»².

O come osserverebbe Ulrich, il matematico protagonista del romanzo di Müsil: «[I]a regolarità delle serie numeriche statistiche è talvolta uguale a quella delle leggi. [...] [S]i è tentato di spiegarla in modo logico e formale, come qualcosa che va da sé; e al contrario s'è anche affermato che la regolarità di fenomeni non legati da una causa comune non può affatto essere spiegata con gli abituali metodi di ragionamento; [...] s'è anche fatta l'ipotesi che non si tratti soltanto di fatti singoli, ma anche di ignote leggi universali. [...] [S]enza dubbio sarebbe molto importante per me personalmente sapere se dietro a tutto ciò non vi siano leggi collettive sconosciute, oppure se per semplice ironia della natura ne derivi il fat-

¹ R. MÜSIL, *L'uomo senza qualità*, volume primo, (trad. it. A. Rho, G. Benedetti, L. Castoldi) Einaudi, Torino, 2014 (prima edizione originale 1930), pp. 7-8.

² D. KWAN-D. SCHEINERT, *Everything Everywhere All at Once*, Eagle pictures, Usa, 2022.

to speciale che non accade mai nulla di speciale, e che il senso supremo viene ad essere qualcosa che si può attingere con la media della più profonda mancanza di senso»³.

L'idea che metriche, numeri, quantificazioni e schemi di regolarità ci consegnino lo strumento per capire il funzionamento dei fenomeni, codificare l'ambiente in cui ci muoviamo, risolvere i problemi che si pongono; l'idea che la sensazione di trovarsi di fronte a qualcosa di eccezionale sia invero «ingiustificata» se si tratta del mero segnale restituito da una «ineluttabilità statistica»; l'idea che tale regolarità sia uguale alle leggi ma non possa essere spiegata con il ragionamento e possa essere compresa solo «con la media della più profonda mancanza di senso».

Sono idee che il tema dell'intelligenza artificiale (IA) alimenta, aggiorna, amplifica.

L'ambito in cui si declinerà la riflessione è quello dell'incidenza dell'utilizzo dell'IA sulle tutele giuridiche del lavoratore.

Come già si è fatto in queste prime righe, l'intenzione è quella di costruire il ragionamento seguendo una linea di comunicazione e continuità con il passato e il presente, incardinandolo in un'analisi sistematica degli strumenti giuridici vigenti e nella valorizzazione delle categorie giuridiche tradizionali proprie della disciplina.

In tal senso, si coglie l'indicazione di chi avverte il rischio di «novismo tecnologico»⁴, «di privilegiare l'approccio empirico» anziché «sistemico», quindi «di sopravvalutare la presenza di vuoti normativi, e sottovalutare invece la capacità di mediazione giuridica delle norme esistenti»⁵. Ovvero anche, sullo specifico versante lavoristico, l'indicazione di chi individua «*the way forward*» nella «*legal evolution, not deregulation*»⁶ e nella necessità che il diritto del lavoro non accetti «passivamente la sirena del determinismo tecnologico», ma con «una valutazione più problematica e articolata» sappia aprirsi alle «nuove sollecitazioni», «senza rinchiudersi entro la cornice regolativa tradizionale»⁷.

³ R. MÜSIL, *L'uomo senza qualità*, cit., pp. 554-555.

⁴ G. FINOCCHIARO-O. POLLICINO, *La Costituzione ci tutela dalle novità della digitalizzazione*, in *Il Sole 24 Ore*, 6 maggio 2023, p. 13.

⁵ U. RUFFOLO, *Artificial intelligence e responsabilità. «Persona elettronica» e teoria dell'illecito*, in A. PAJNO-F. DONATI-A. PERRUCCI (a cura di), *Intelligenza artificiale e diritto: una rivoluzione? Amministrazione, responsabilità, giurisdizione*, vol. II, il Mulino, Bologna, 2022, pp. 231-232. L'autore evidenzia il cambiamento che in tal senso si è registrato nella posizione del Parlamento europeo, dalla Risoluzione del 16 febbraio 2017, in materia di norme di diritto civile sulla robotica, in cui veniva tracciata la via del rinnovamento delle norme e delle categorie giuridiche, alle Risoluzioni del 12 febbraio 2019 sulla politica industriale europea globale e del 20 ottobre 2020 in tema di responsabilità civile per l'intelligenza artificiale, in cui non si ritiene necessaria una revisione completa del sistema, ma solo alcuni adeguamenti specifici e coordinati.

⁶ S. DEAKIN-C. MARKOU, *The law-technology cycle and the future of work*, Centre for Business Research, University of Cambridge, Working Paper No. 504/2018, p. 17.

⁷ A. PERULLI-V. SPEZIALE, *Dieci tesi sul diritto del lavoro*, il Mulino, Bologna, 2022, p. 143.

In quest'ottica di sistema, dove il diritto viene interpellato sia come «strumento ordinatore», sia come «oggetto di cambiamento»⁸, si individueranno i termini di innesto dei contenuti normativi che *de iure condendo* si stanno prospettando a livello Ue, nonché i possibili versanti su cui può essere incentivata l'evoluzione della materia, nel tentativo di contribuire a una costruzione *future-proof* dell'architettura regolativa di tutela.

⁸ A. PAJNO, *Prefazione. La costruzione dell'infosfera e le conseguenze sul diritto*, in A. PAJNO-F. DONATI-A. PERRUCCI (a cura di), *Intelligenza artificiale e diritto: una rivoluzione? Amministrazione, responsabilità, giurisdizione*, vol. II, il Mulino, Bologna, 2022, p. 25.

CAPITOLO PRIMO

IA E TUTELA DEL LAVORATORE: INQUADRAMENTO DEL PROBLEMA E QUESTIONI DEFINITORIE

SOMMARIO: 1. Obiettivo e struttura dell'indagine. – 2. IA, algoritmo, procedimento automatizzato: definizioni e distinzioni.

1. *Obiettivo e struttura dell'indagine*

Un sistema di intelligenza artificiale (IA) può costituire uno strumento che veicola o informa le scelte e l'esercizio delle prerogative datoriali¹.

Nel filtrare i curricula inviati per una posizione lavorativa, sulla base dei dati testuali ivi contenuti, o nell'interagire con i candidati attraverso *chatbots/voicebots* o avatar digitali, il sistema assiste nelle varie fasi del processo di assunzione, elaborando selezioni, classificazioni e graduatorie, nonché raccogliendo e sistematizzando informazioni utili per il reclutamento².

¹ L'incidenza dell'uso dell'IA nel luogo di lavoro può essere osservata anche su un versante diverso e opposto: si pensi all'utilizzo che può essere fatto di ChatGpt da parte dei dipendenti, con o senza l'approvazione del datore di lavoro, per essere supportati nella realizzazione di un compito (ad es. l'ingegnere che carica un codice chiedendo a ChatGpt di trovare il difetto e ottimizzare il *software*) o preparare rapporti (ad es. il verbale di una riunione, con registrazione della discussione e dei partecipanti). A fronte del rischio che questi processi possano ledere segreti aziendali, Samsung, SK Hynix, LG e molte altre aziende si stanno affrettando a vietare o a redigere linee guida per l'uso di ChatGpt e di altri servizi simili sul posto di lavoro. Cfr. P. CLARKE, *ChatGPT leaking Samsung chip secrets is iceberg's tip*, in www.eenewseurope.com, 4 aprile 2023. La problematicità dell'uso di questa *chatbot* è nota per l'ordinamento italiano, considerato il provvedimento del 30 marzo 2023 con cui il Garante per la protezione dei dati personali ha bloccato l'applicazione disponendo, in via d'urgenza, nei confronti di OpenAI, società statunitense sviluppatrice e gestrice del *software*, la misura della limitazione provvisoria del trattamento dei dati personali degli interessati stabiliti nel territorio italiano. Il 28 aprile 2023, il Garante ha successivamente comunicato la riapertura della piattaforma in Italia a fronte della nota inviata da OpenAI e delle maggiori garanzie introdotte.

² Cfr. J. ADAMS-PRASSL, *Regulating algorithms at work: Lessons for a 'European approach to artificial intelligence'*, in *European Labour Law Journal*, 2022, p. 30 ss.; V. DE STEFANO-M. WOUTERS, *AI and digital tools in workplace management and evaluation: An assessment of the EU's le-*

Nel prevedere e scandire le coordinate spazio-temporali della prestazione, sulla base di una stima probabilistica del tempo o del percorso necessari, il sistema plasma, specifica, conforma il contenuto dell'obbligazione lavorativa, identificando lo standard di diligenza richiesto.

Nell'alimentarsi dei dati necessari per il processo computazionale, attraverso sensori montati su entità robotiche, mezzi di trasporto o dispositivi ovvero distribuiti nell'ambiente circostante, il sistema abilita un potenziale controllo sulla prestazione e/o una profilazione del lavoratore³, a cui può essere agganciata o seguire una valutazione sulla conformità dei risultati della prestazione con lo standard di diligenza richiesto, al fine di verificare la necessità o meno di fornire percorsi formativi mirati, contestare un'infrazione con avvio di un procedimento disciplinare oppure attribuire/negare un premio o un avanzamento di carriera⁴.

gal framework, European Parliamentary Research Service (EPRS), PE 729.516 – May 2022; H. PARVIAINEN, *Can algorithmic recruitment systems lawfully utilise automated decision-making in the EU?*, in *European Labour Law Journal*, 13(2), 2022, p. 225 ss.; I. AJUNWA, *An Auditing Imperative for Automated Hiring Systems*, in *Harvard Journal of Law & Technology*, 34(2), 2021, p. 622 ss.; A. TODOLI-SIGNES, *Algorithms, artificial intelligence and automated decisions concerning workers and the risks of discrimination: the necessary collective governance*, in *Transfer*, 2019, 25(4), p. 465 ss.; R. VEDAPRADHA-R. HARIHARAN-R. SHIVAKAMI, *Artificial Intelligence: A Technological Prototype in Recruitment*, in *Journal of Service Science and Management*, 3(12), 2019, p. 382 ss.. Per un riferimento alle tecniche di *recruitment* digitale v. anche la circolare del Ministero del lavoro n. 19 del 20 settembre 2022 in merito all'art. 1-bis d.lgs. n. 152/1997 (v. cap. 2, § 6).

³ Cfr. A. ALOISI-E. GRAMANO, *Artificial Intelligence Is Watching You at Work: Digital Surveillance, Employee Monitoring, and Regulatory Issues in the EU Context*, in *Comparative Labor Law & Policy Journal*, 41, 2019, p. 95 ss.; I. AJUNWA-K. CRAWFORD-J. SCHULTZ, *Limitless worker surveillance*, in *California Law Review*, 105(3), 2017, p. 735 ss. Il tema del controllo a distanza e della profilazione è particolarmente approfondito dalla dottrina italiana, nell'esame del rapporto coordinato tra gli artt. 4 e 8 St. Lav. e il Regolamento Ue n. 2016/679 (GDPR). Cfr. M. AIMO, *Dalle schedature dei lavoratori alla profilazione tramite algoritmi: serve ancora l'art. 8 dello Statuto dei lavoratori?*, in *Lavoro e diritto*, 2021, p. 585 ss.; A. SARTORI, *Il controllo tecnologico sui lavoratori. La nuova disciplina italiana tra vincoli sovranazionali e modelli comparati*, Giappichelli, Torino, 2020; E. DAGNINO, *Dalla fisica all'algoritmo: una prospettiva di analisi giuslavoristica*, Adapt University Press, Bergamo, 2019; F. COSTANTINI, *Profilazione e "automated decision making" in ambito lavorativo nella giurisprudenza italiana*, in *Il lavoro nella giurisprudenza*, 2019, p. 984 ss.; A. TOPO, *Automatic management, reputazione del lavoratore e tutela della riservatezza*, in *Lavoro e diritto*, 2018, p. 453 ss.; A. INGRAO, *Il controllo a distanza sui lavoratori e la nuova disciplina privacy: una lettura integrata*, Cacucci, Bari, 2018; V. NUZZO, *La protezione del lavoratore dai controlli impersonali*, Ed. Scientifica, Napoli, 2018; 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, Torino, 2017, p. 3 ss.; A. DONINI, *Profilazione reputazionale e tutela del lavoratore: la parola al Garante della Privacy*, in *Labour & Law Issues*, 3(1), 2017, p. R 37 ss.; M. MARAZZA, *Dei poteri (del datore di lavoro), dei controlli (a distanza) e del trattamento dei dati (del lavoratore)*, in *WP CSDLE "Massimo D'Antona".IT*, n. 300, 2016 e *Argomenti di diritto del lavoro*, 2016, p. 483 ss.

⁴ Cfr. A. ALOISI, *Boss ex Machina: employer powers in workplaces governed by algorithms and artificial intelligence*, in A. LO FARO (a cura di), *New Technology and Labour Law. Selected topics*, Giappichelli, Torino, 2023, p. 15 ss.; L. ZAPPALÀ, *Informatizzazione dei processi decisionali e diritto*

Nel definire la distribuzione dei carichi di lavoro e la configurazione dell'assetto aziendale, a partire da una previsione della domanda di beni e servizi e della scelta tecnico-organizzativa più efficiente (c.d. *workflow algorithms*), il sistema può consegnare una decisione sull'estensione o collocazione spaziale della prestazione ovvero sulla collocabilità del lavoratore, che può tradursi in un atto datoriale di modifica dell'oggetto, tempo o luogo della prestazione ovvero anche, da ultimo, in un atto di recesso⁵.

L'inserimento dell'algoritmo nell'architettura del processo produttivo veicola e informa le prerogative datoriali, ma al contempo può innescare una trasformazione: i poteri «si mimetizzano e si fanno “diffusi”, transitando dalla dimensione dell'imporre e del correggere a quella dell'indurre e del prevenire»⁶; si diluisce la linea di confine tra conformazione della prestazione e monitoraggio dell'attività lavorativa: «il potere direttivo cambia pelle e si contamina con il potere di controllo assumendo le sembianze di un “potere di controllo direttivo”»⁷; l'ambiente digitale «precostituisce di per sé» un sistema di «comportamenti vincolanti a carico del lavoratore a prescindere dall'esercizio di prerogative formali» e diventa in questo più ardua «la scommessa sull'efficacia discrezionale attribuita alle manifestazioni del potere nell'impresa»⁸.

L'uso di sistemi di IA da parte del datore nell'organizzazione della propria attività pone per il diritto del lavoro principalmente due problematiche.

to del lavoro: algoritmi, poteri datoriali e responsabilità del prestatore nell'era dell'intelligenza artificiale, in S. ALEO (a cura di), *Evoluzione scientifica e profili di responsabilità*, Pacini editore, Pisa, 2021, p. 363 ss.; V. MAIO, *Il diritto del lavoro e le nuove sfide della rivoluzione robotica*, in *Argomenti di diritto del lavoro*, 2018, p. 1414 ss.; A. TODOLI-SIGNES, *Algorithms, artificial intelligence*, cit.; cfr. altresì le relazioni di G. DE SIMONE-M. MARAZZA-A. OCCHINO, al *Séminaire méditerranéen de droit du travail 2018*, dedicato al tema *Intelligence artificielle et droits du travail*, 21-22 giugno 2018, Avignone.

⁵ Cfr. V. DOELLGAST-I. WAGNER-S. O'BRADY, *Negotiating limits on algorithmic management in digitalised services: cases from Germany and Norway*, in *Transfer*, 29(1), 2022, p. 105 ss.; A.J. WOOD, *Algorithmic management: Consequences for work organisation and working conditions*, Report no. JRC124874, Seville, European Commission, 2021; M. OTTO, *Workforce analytics v fundamental rights protection in the EU in the age of big data*, in *Comparative Labor Law & Policy Journal*, 40(3), 2018, p. 389 ss.; D. ANGRAVE-A. CHARLWOOD-I. KIRKPATRICK *et al.*, *HR and analytics: Why HR is set to fail the big data challenge*, in *Human Resource Management Journal*, 26(1), 2016, p. 1 ss.; M.K. LEE-D. KUSBIT-E. METSKY *et al.*, *Working with machines: The impact of algorithmic and data-driven management on human workers*, in *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, Association for Computing Machinery, New York, 2015, p. 1603 ss.

⁶ M. NOVELLA, *Impresa*, in M. NOVELLA-P. TULLINI, *Lavoro digitale*, Giappichelli, Torino, 2022, p. 23.

⁷ L. TEBANO, *Controlli digitali*, in S. BORELLI-V. BRINO-C. FALERI-L. LAZZERONI-L. TEBANO-L. ZAPPALÀ, *Lavoro e tecnologie*, Giappichelli, Torino, 2022, p. 49.

⁸ P. TULLINI, *La questione del potere nell'impresa. Una retrospettiva lunga mezzo secolo*, in *Lavoro e diritto*, 3-4, 2021, p. 442.

Da un lato, aumenta il rischio di violazione dei diritti riconosciuti al lavoratore a tutela dei suoi interessi e dei correlativi comandi e divieti posti al datore nell'esercizio dei suoi poteri.

Dall'altro, può rendere difficile individuare e dimostrare detta violazione, inficiando l'effettività della normativa e, quindi, la possibilità per il lavoratore di ricorrere all'autotutela nel rapporto di lavoro e/o alla tutela giudiziale.

Alla radice di queste problematiche si collocano tre caratteristiche che possono connotare, in diverso grado, un sistema di IA: la complessità, l'autonomia e la dipendenza dai dati⁹. Le stesse possono, infatti, implicare opacità (ossia carenza di conoscibilità e/o spiegabilità) del funzionamento del sistema e/o della decisione dallo stesso consegnata, imprevedibilità dell'*output*, nonché rischi di *bias*, distorsioni causate da pregiudizi, connessi alla stretta dipendenza della razionalità del processo dai dati di *training* e/o di *input* e dalla possibile riproduzione rigida, meccanica ed esponenziale di discriminazioni insite negli stessi¹⁰.

⁹ Al riguardo, cfr. Commission Staff Working Document, *Impact Assessment accompanying the Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative Acts*, Brussels, 21 aprile 2021, SWD(2021) 84 final. In tale sede, in particolare, si individuano come caratteristiche proprie di un sistema di IA: opacità, complessità, continuo adattamento e imprevedibilità, comportamento autonomo e dipendenza dai dati (Table 3, p. 28). Si ritiene di poter ricostruire queste caratteristiche nei termini di sintesi riportati nel testo: l'opacità, ossia la carenza di trasparenza, è descritta come conseguenza della complessità del sistema (quale «molteplicità di componenti e processi di un sistema di IA e loro interconnessioni»); l'adattività, intesa come «capacità funzionale di alcuni sistemi di IA di “apprendere” e “adattarsi” continuamente durante il loro funzionamento, portando talvolta a risultati imprevedibili», da un lato, non costituisce un elemento necessario ai fini della qualificazione di IA, dall'altro, può individuarsi come specifica declinazione dei caratteri della complessità e autonomia (ossia della «capacità funzionale di alcuni sistemi di IA di generare *output* come “decisioni” con un intervento umano limitato o nullo», p. 28). L'*Impact Assessment* prosegue evidenziando che «queste caratteristiche specifiche dei sistemi di IA possono creare nuovi rischi per la sicurezza e i diritti fondamentali e accelerare la probabilità o l'intensità dei rischi esistenti, nonché rendere difficile [...] verificare il rispetto e l'applicazione delle norme esistenti» (p. 29). Ancora, si afferma che «l'IA utilizzata per sostituire o supportare il processo decisionale umano o per altre attività come la sorveglianza può violare i diritti individuali. Non si tratta di un difetto della tecnologia *in sé*, ma della responsabilità degli esseri umani che la progettano e la utilizzano e che devono assicurarsi che queste violazioni non avvengano». Inoltre, «qualora si verificino violazioni dei diritti fondamentali, queste possono anche essere molto difficili da individuare e dimostrare, soprattutto quando il sistema non è trasparente. Ciò mette a dura prova l'effettiva applicazione della legislazione europea esistente volta a salvaguardare i diritti fondamentali» (p. 16); «una delle principali minacce al diritto a un ricorso effettivo è la mancanza di trasparenza nell'uso e nel funzionamento dei sistemi di IA» (p. 21). Rispetto alle due problematiche rilevate con riguardo all'applicazione dei diritti, ossia non solo il rischio di una loro violazione ma anche la maggior difficoltà di individuarla e dimostrarla, si evidenziano correlativamente, nell'elenco dei requisiti indicati per rendere affidabile (*trustworthy*) il sistema di IA, la trasparenza e il controllo umano (p. 41, v. cap. 2, § 1.1).

¹⁰ 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*, 25, 2022, p. 189 ss.

La portata di queste caratteristiche e la loro incidenza sulle menzionate problematiche variano a seconda del tipo o dei tipi di algoritmi utilizzati, potendo passare, come si vedrà oltre nella trattazione, dal grado minimo riscontrabile in caso di algoritmo deterministico a quello massimo in ipotesi di algoritmo di apprendimento automatico (c.d. *machine learning*) associato all'uso di reti neurali profonde (c.d. *deep learning*).

D'altra parte, seppur diversamente declinate in relazione alla portata delle citate variabili, le soluzioni che si individuano a risposta del problema, nell'approccio regolativo di livello Ue, possono essere ricondotte a due principali requisiti di garanzia, da rispettare nella progettazione e nell'uso del sistema: la sorveglianza umana, quale garanzia per la prevenzione o minimizzazione del rischio di violazione dei diritti; la trasparenza, per il rafforzamento della loro effettività. Accanto a questa scelta di contenuto, si può individuare anche una precisa opzione di metodo, nel riferimento alla necessità di adottare un modello basato sulla valutazione e gestione del rischio.

L'obiettivo della presente analisi è quello di contribuire allo sviluppo del ragionamento giuridico sul tema, dal punto di vista lavoristico, proponendo una lettura di sistema, utile a una ricostruzione e applicazione integrata delle regole coinvolte.

Il percorso si comporrà di tre momenti.

In una prima fase, l'indagine sarà volta a inquadrare le citate risposte di contenuto e di metodo prospettate nell'approccio regolativo Ue all'IA. Si evidenzierà, al riguardo, l'importanza di declinare la garanzia della sorveglianza umana sul duplice piano dell'intervento nel processo automatizzato e del governo complessivo del sistema e la stretta interconnessione con la garanzia della trasparenza. Quest'ultima si colloca, infatti, tra i presupposti di efficacia dell'intervento dell'operatore nel ciclo di funzionamento del sistema, ma altresì tra gli obiettivi di un governo del rischio finalizzato ad assicurare l'affidabilità dell'IA. Si procederà, quindi, a individuare, *de iure condito e condendo*, lungo tali direttrici, l'architettura regolativa di matrice Ue funzionale alla tutela del lavoratore nei confronti di poteri datoriali esercitati con l'uso di sistemi algoritmici. Trattasi di fonti che, in molti casi, non sono caratterizzate da una matrice giuslavoristica e che cionondimeno presentano tale funzionalità, sia attraverso la diretta predisposizione di protezioni, sia perché strumentali all'operatività ed effettività delle tecniche di controllo dei poteri datoriali, previste nell'ambito delle fonti giuslavoristiche. Al riguardo, si evidenzieranno il ruolo centrale, quasi egemonico, ricoperto dal Regolamento Ue 2016/679 in materia di trattamento dei dati personali (GDPR) e la rilevanza che potrà assumere, una volta adottato, il Regolamento sull'IA, soprattutto con l'innesto di alcuni suoi contenuti all'interno del quadro normativo del GDPR, attraverso lo specifico canale offerto dallo strumento della valutazione di impatto. Si sottolineerà, quindi, l'importanza di coordinare la riflessione così sviluppata con fonti di più stretta matrice lavoristica. Anzitutto, a completamento della ricostruzione *de iure condendo*, si effettuerà un confronto con le soluzioni

riscontrabili, per il relativo ambito di applicazione, nella proposta di direttiva sul lavoro mediante piattaforme digitali; in termini più generali e complessivi, specifica attenzione sarà dedicata al necessario raccordo e integrazione con la normativa Ue in materia di salute e sicurezza dei lavoratori e, nell'approfondimento della proiezione collettiva della tutela, con quella in tema di coinvolgimento dei lavoratori nell'impresa. L'analisi consentirà, da ultimo, di valutare la tenuta di alcune scelte regolative adottate a livello interno, in particolare dal c.d. decreto trasparenza, d.lgs. n. 104/2022, così come modificato dal d.l. n. 48/2023, c.d. decreto lavoro.

Proseguendo l'indagine, si rifletterà su come le garanzie di sorveglianza umana e trasparenza, su cui si fonda l'architettura normativa individuata, integrino il regolamento negoziale del rapporto di lavoro e operino nella logica del contratto: l'una, quale misura dell'adempimento degli obblighi attribuiti al datore nell'esercizio delle proprie prerogative, quindi elemento di determinazione della diligenza richiesta e criterio che consente di stabilire la colpa in un eventuale giudizio di responsabilità contrattuale soggettiva; l'altra, quale obbligo di protezione collocabile nello spettro d'azione delle clausole generali di correttezza e buona fede. Il ragionamento sarà funzionale anche a una verifica dei termini con cui il principio di responsabilizzazione *ex ante*, sotteso all'obbligo di governo del rischio e quindi alla garanzia di sorveglianza umana nella sua duplice declinazione, si traduce sul piano dell'allocazione delle responsabilità, tenendo in considerazione la complessa pluralità di soggetti che popolano la catena della fornitura e il ciclo di vita di un sistema di IA.

Infine, muovendo dalle riflessioni maturate, l'analisi evidenzierà le criticità del sistema di tutela del lavoratore e le possibili prospettive di miglioramento. Si segnaleranno anzitutto le problematiche che possono incidere sull'effettività della garanzia di trasparenza: dalla cautela che deve guidare il confronto con la razionalità algoritmica, laddove si tratti di vagliare il rispetto dei limiti esterni al potere datoriale, alle difficoltà che si possono realisticamente riscontrare in giudizio, nel controllo sulla veridicità delle affermazioni divulgate dal datore, anche a fronte di un ordine di *disclosure*, fino al necessario bilanciamento con eventuali diritti di privacy collegati al sistema di IA. Particolare attenzione sarà dedicata alla questione discriminatoria, connotata da un imprescindibile legame con il tema della trasparenza e sistematicamente richiamata dalla pluralità di fonti coinvolte nell'indagine. Al riguardo, si valuterà l'utilità di alcune soluzioni apprestate dalla recente direttiva 2023/970 in materia di trasparenza retributiva di genere.

Rispetto alla garanzia di sorveglianza umana, l'analisi evidenzierà come la prospettiva di lettura integrata delle fonti, elaborata nel corso della trattazione, trovi corrispondenza, se non importante consolidamento, all'interno di alcune proposte finalizzate a valorizzare il linguaggio dei diritti fondamentali sia *de iure condito* nel contesto normativo del GDPR, sia *de iure condendo* nel Regolamento

sull'IA. A tal ultimo proposito, si segnalerà la linea di emendamento approvata dalle commissioni IMCO-LIBE del Parlamento europeo, in vista della votazione in plenaria prevista per giugno 2023. Segnatamente, si dispone l'introduzione in capo all'utente (ossia, per quanto ci concerne, il datore di lavoro) di un obbligo di *Fundamental Rights Impact Assessment* (FRIA), da condurre congiuntamente alla valutazione d'impatto sulla protezione dei dati di cui al GDPR.

Prodromica al discorso è la questione definitoria delle fattispecie di riferimento, anzitutto quella di IA, di algoritmo, di procedimento automatizzato. L'analisi e il confronto tra queste nozioni sono necessari per capire sia i diversi gradi in cui si possono presentare le problematiche relative alla tutela del lavoratore e le correlate esigenze di garanzia in termini di sorveglianza umana e trasparenza; sia le risposte regolative che *de iure condito* e *condendo* vi si ricollegano, i canali di comunicazione e integrazione tra i diversi plessi regolativi, i punti di accesso per un approccio condiviso e partecipativo.

2. IA, algoritmo, procedimento automatizzato: definizioni e distinzioni

L'algoritmo è una sequenza finita di istruzioni non ambigue, tali da poter essere eseguite meccanicamente; il codice sorgente è un algoritmo espresso in un linguaggio di programmazione, mentre un sistema algoritmico rappresenta un processo automatizzato, ossia un processo basato sull'esecuzione di un algoritmo (*melius*, di un codice oggetto¹¹).

Non tutti i sistemi algoritmici possono, tuttavia, considerarsi sistemi di IA.

Quali siano le caratteristiche che devono necessariamente sussistere affinché vi possa essere una qualificazione in tal senso è questione tutt'altro che scontata a livello sia di riflessione scientifica sia di dibattito istituzionale¹². Non è un caso che questo profilo costituisca uno dei più controversi all'interno della discussione

¹¹ Il codice oggetto è la traduzione del codice sorgente nel linguaggio binario della macchina, in grado come tale di essere eseguito dalla stessa. La traduzione è generata da un programma informatico, detto compilatore. L'operazione inversa (*reverse engineering*), ossia la ricostruzione del codice sorgente dal codice oggetto, è detta de-compilazione e, come si vedrà oltre, è consentita dalla normativa di tutela del diritto d'autore sul *software* solo in alcune ipotesi (dir. 2009/24/CE, l. n. 633/41, v. cap. 4, § 2.2). Il *software* è l'insieme di procedure e istruzioni (algoritmi tradotti in linguaggio di programmazione) in un sistema di elaborazione dati.

¹² Cfr. 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, il Mulino, Bologna, 2022, p. 43 ss., spec. p. 48. A livello Ue, si può richiamare il rapporto S. SAMOILI-M. LOPEZ COBO-E. GOMEZ GUTIERREZ-G. DE PRATO-F. MARTINEZ-PLUMED-D. DELIPETREV, *AI WATCH. Defining Artificial Intelligence*, Publications Office of the European Union, Luxembourg, 2020.

sulla proposta di Regolamento sull'IA che si sta svolgendo in seno alle istituzioni europee.

Ai fini del ragionamento che si intende sviluppare e in coerenza con l'impostazione antropocentrica che connota la via europea alla regolazione in materia, si partirà dall'assunto secondo cui un sistema di IA può al momento considerarsi non tanto come possibile agente quanto come uno strumento dell'azione umana, «una forma di esercizio», un «registro» del potere¹³. Si farà, inoltre, riferimento alla definizione di IA adottata nella proposta di Regolamento europeo¹⁴, una definizione ad ampio raggio¹⁵ che, rinviando a una vasta lista di approcci e tecniche, consente di includere nella categoria l'uso di algoritmi sia deterministici, sia non deterministici, sia di apprendimento automatico. Questa distinzione, riconducibile per lo più alla tripartizione contenuta nell'allegato I della citata proposta, è necessaria e utile per comprendere le differenti caratterizzazioni dei sistemi rispetto ai requisiti di prevedibilità del risultato, comprensibilità del modello e spiegabilità della decisione.

Proprio per questo, si ritiene significativo proporla in questo contesto di indagine nonostante nell'orientamento generale adottato dal Consiglio dell'Ue a dicembre 2022 si prevedano l'eliminazione dell'allegato I, al fine di sottrarre alla Commissione margini di intervento sul suo contenuto, e lo spostamento nei considerando dei riferimenti di dettaglio alle tecniche di IA, con una bipartizio-

¹³ F. CABITZA, *Deus in machina? L'uso umano delle nuove macchine, tra dipendenza e responsabilità*, in L. FLORIDI-F. CABITZA (a cura di), *Intelligenza artificiale. L'uso delle nuove macchine*, Bompiani, Milano, Firenze, 2021, pp. 67-68; K. CRAWFORD, *The Atlas of AI. Power, Politics, and the Planetary Costs of Artificial Intelligence*, Yale University Press, New Haven, 2021. Rispetto al discusso tema della soggettività giuridica dell'agente digitale e a quello connesso della responsabilità da algoritmo cfr., *ex multis*, G. TEUBNER, *Soggetti giuridici digitali? Sullo status privatistico degli agenti software autonomi*, Edizioni Scientifiche Italiane, Napoli, 2018; U. RUFFOLO, *Le responsabilità da produzione e gestione di intelligenza artificiale self-learning*, in U. RUFFOLO (a cura di), *XXVI Lezioni di Diritto dell'Intelligenza artificiale*, Giappichelli, Torino, 2021, p. 131 ss.; M. COSTANZA, *Impresa robotizzata e responsabilità*, in U. RUFFOLO (a cura di), *Intelligenza artificiale e responsabilità*, Giuffrè, Milano, 2017, p. 107 ss. Rispetto all'imputabilità della decisione algoritmica al datore di lavoro, cfr. G. ZAMPINI, *Intelligenza artificiale e decisione datoriale algoritmica. Problemi e prospettive*, in *Argomenti di diritto del lavoro*, 2022, p. 467 ss.; P. LAMBERTUCCI, *Intelligenza artificiale e tutela dei lavoratori: prime riflessioni*, in *Argomenti di diritto del lavoro*, 2022, p. 897 ss.; A. ROTA, *Rapporto di lavoro e big data analytics: profili critici e risposte possibili*, in *Labour & Law Issues*, 2017, p. 1.42 ss.

¹⁴ Commissione europea, *Proposta di Regolamento che stabilisce regole armonizzate sull'intelligenza artificiale*, Bruxelles, 21 aprile 2021, COM(2021)206 final. Il Consiglio ha già approvato, il 6 dicembre 2022, l'orientamento generale con cui si presenterà ai triloghi (*General approach*, 14954/22); per quanto riguarda il Parlamento europeo, invece, come in parte anticipato, è attesa per la metà di giugno 2023 la votazione in plenaria sul testo di compromesso adottato congiuntamente l'11 maggio 2023 dalle commissioni per il mercato interno e le libertà civili (IMCO-LIBE, Rapporteurs: B. Benifei e I. Tudorache).

¹⁵ Cfr. L. ZAPPALÀ, *Intelligenza artificiale*, in AA.VV., *Lavoro e tecnologie*, Giappichelli, Torino, 2022, p. 133 ss.

ne tra *machine learning* e approcci basati sulla logica e conoscenza (cons. n. 6-*bis* e 6-*ter*; bipartizione ripresa anche nella definizione di sistema di IA di cui all'art. 3, punto 1). In questa prospettiva, si precisa che sono esclusi i sistemi che utilizzino regole definite solamente da persone fisiche per eseguire automaticamente le operazioni (cons. n. 6), mentre le tecniche non deterministiche prima indicate nella lett. *c* dell'allegato sono citate e declinate nel contesto del *machine learning*¹⁶.

Si tornerà per un commento su questi profili dell'orientamento generale del Consiglio dopo alcune precisazioni sulla menzionata tripartizione.

Per meglio comprendere le differenze, al suo interno, tra le categorie di algoritmi, è importante anzitutto distinguere la prevedibilità del risultato, la comprensibilità di un modello, intesa come «la proprietà di un sistema di essere compreso nel suo funzionamento (*how*) da un essere umano», e ancora la spiegabilità del risultato, quale capacità del modello di fornire una spiegazione del perché abbia prodotto un determinato *output*, anche rispetto a uno specifico caso concreto¹⁷. La spiegabilità, oltre a poter essere diretta (se caratterizzante in modo nativo il modello, c.d. *transparency by design*) o indiretta (se investigabile *ex post* attraverso altre tecniche esterne, anch'esse potenzialmente algoritmiche, c.d. metodi di *Explainable AI – XAI*), è un concetto che deve essere declinato in relazione alle capacità cognitive e competenze del destinatario¹⁸, nonché alle finalità dell'accesso alla spiegazione. In tal senso, un modello che di per sé fornisce un risultato direttamente spiegabile per un soggetto esperto (da qui l'importanza di tale figura nei processi di informazione e consultazione dei lavoratori e dei loro rappresentanti: v. cap. 2, § 5), potrebbe richiedere un intervento successivo volto a rendere tale risultato spiegabile o a facilitarne la comprensione per un soggetto non esperto (ad es. un algoritmo di apprendimento automatico supervisionato, basato sulla regressione lineare: v. *infra*)¹⁹.

Gli algoritmi deterministici seguono un approccio basato sulla logica e regolano un processo decisionale non soltanto comprensibile e spiegabile, ma anche dal

¹⁶ Si segnala che anche nel testo di compromesso approvato dalle commissioni congiunte IMCO-LIBE del Parlamento europeo si prevede l'abrogazione dell'allegato I, ma in sede di considerando non si effettua la bipartizione di categorie indicata nell'orientamento generale del Consiglio e una declinazione nel dettaglio delle tecniche è riservata agli approcci di apprendimento automatico, individuati come quelli spesso presenti nei sistemi di IA (cons. n. 6).

¹⁷ M. PALMIRANI, *Interpretabilità, conoscibilità, spiegabilità dei processi decisionali automatizzati*, in U. RUFFOLO (a cura di), *XXVI Lezioni di Diritto dell'Intelligenza artificiale*, Giappichelli, Torino, 2021, p. 70.

¹⁸ Parla, in tal senso, di «informazione “umanizzata di terza generazione”» L. ZAPPALÀ, *Appunti su linguaggio, complessità e comprensibilità del lavoro 4.0: verso una nuova proceduralizzazione dei poteri datoriali*, in WP CSDLE “Massimo D’Antona”.IT, n. 462, 2022, p. 29.

¹⁹ A.B. ARRIETA-N. DÍAZ-RODRÍGUEZ *et al.*, *Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI*, in *Information Fusion*, n. 58, 2019, p. 82 ss.

risultato prevedibile *ex ante*. La deduzione automatizzata di una certa conclusione, a partire dall'informazione raccolta di cui il sistema dispone, segue infatti una sequenza di regole di inferenza che, a parità di dati d'ingresso, può condurre a un solo possibile percorso e a un solo possibile risultato²⁰. Proprio per questo, si è detto che tali algoritmi «*keep the human in the loop*»²¹. A questa categoria può essere ascritto, ad esempio, un algoritmo che traduce e applica i criteri di selezione dei lavoratori licenziandi previsti da un accordo sindacale sottoscritto *ex art.* 4, l. 223/91²². Ancora, può rientrarvi l'algoritmo Frank del noto caso Deliveroo, quantomeno per il segmento legato all'assegnazione degli slot di prenotazione²³.

²⁰ G. AVANZINI, *Decisioni amministrative e algoritmi informatici*, Ed. Scientifica, Napoli, 2019, p. 7. Cfr. la posizione di M. MARAZZA-F. D'AVERSA, *Dialoghi sulla fattispecie dei "sistemi decisionali o di monitoraggio automatizzati" nel rapporto di lavoro (a partire dal Decreto trasparenza)*, in *giustiziacivile.com*, 8 novembre 2022, secondo cui «se il sistema, all'esito di uno specifico *input*, può, in via automatizzata, scegliere tra varie azioni possibili (A+B= C? o C? o C?) si è in presenza di un sistema decisionale automatizzato» (rilevante anche ai fini dell'applicazione dell'art. 1-bis d.lgs. n. 152/1997, v. cap. 2, § 6). «Diversamente, se, all'esito di uno specifico *input*, il sistema può, in via automatizzata, compiere, in forza di una rigida regola, un'unica azione predeterminata (A+B= unicamente a C) si è in presenza di un sistema, sì automatizzato, ma non decisionale» (p. 7).

²¹ P. BOUCHER, *How artificial intelligence works*, in *European Parliamentary Research Service*, marzo 2019.

²² Nel caso deciso da App. Reggio Calabria, 29 gennaio 2021, n. 66, ad esempio, l'algoritmo aveva tradotto e implementato quanto previsto dall'accordo sindacale, quindi la distinzione dei lavoratori da licenziare in tre gruppi (carrellisti, gruisti e mezzi bassi) e la loro selezione sulla base dell'applicazione, in concorso tra loro, dei criteri dei carichi di famiglia e di anzianità aziendale, con attribuzione al primo di un valore del 34% e al secondo di un valore del 66%: le graduatorie erano state così determinate dall'algoritmo sulla base del minor punteggio riportato dai lavoratori nell'ambito di ciascun profilo professionale.

²³ Si consenta sul punto un rinvio a M. PERUZZI, *Il diritto antidiscriminatorio al test di intelligenza artificiale*, in *Labour & Law Issues*, 7(1), 2021, p. 148 ss. Come noto, l'algoritmo Frank, utilizzato da parte di Deliveroo, è stato al centro di un caso deciso dal Tribunale di Bologna con ordinanza del 31 dicembre 2020. Per quanto attiene alla qualificazione della tipologia dell'algoritmo adottata nel testo, si consenta di evidenziare che è sì vero che Deliveroo presentava, 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, nella citata ordinanza, osservava 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 ha reputato 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. In tal senso, nel giudizio, ai fini dell'accertamento della discriminazione,

Nello specifico, sulla base di un ragionamento meccanico, 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. Alla categoria in esame può essere ascritto anche l'algoritmo utilizzato dal Ministero dell'istruzione italiano per gestire l'assegnazione delle sedi di servizio degli insegnanti²⁴. Tarato sui parametri previsti dalla l. n. 107/2015 (c.d. Buona Scuola), tale algoritmo è stato oggetto di approfondito interesse da parte della giurisprudenza amministrativa²⁵, particolarmente significativa rispetto al tema dell'obbligo e/o necessità di ostensione del codice sorgente ai fini della tutela dell'interesse alla trasparenza e annessa (in)opponibilità di eventuali diritti di proprietà intellettuale²⁶ (v. cap. 4, § 2.2), nonché con riguardo alla questione della qualificabilità della decisione automatizzata in termini di atto giuridico, ivi ammessa nei limiti in cui la sussistenza dell'elemento della volontà umana sia garantita dal fatto che l'algoritmo che guida il processo decisionale sia stato completamente scritto dall'uomo in sede di programmazione²⁷ (v. *infra*, invece, gli algoritmi di apprendimento automatico).

era isolato e analizzato, all'interno del più ampio processo di funzionamento del sistema digitale utilizzato da Deliveroo, uno specifico segmento, quello che gestiva la sequenza decisionale che partiva dal computo delle passate cancellazioni tardive e arrivava 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 può affermare che il sistema operasse sulla base di un algoritmo deterministico.

²⁴ Cfr. G. F. ITALIANO-S. CIVITARESE MATTEUCCI-A. PERRUCCI, *L'intelligenza artificiale*, cit., spec. pp. 63-64.

²⁵ T.a.r. Lazio-Roma, sez. III-*bis*, 10 settembre 2018, n. 9227; Cons. Stato, sez. VI, 4 febbraio 2020, n. 881. Cfr. al riguardo, I. PICCININI-M. ISCERI, *IA e datori di lavoro: vero una e-leadership?*, in *Lavoro Diritti Europa*, 2, 2021.

²⁶ Cfr. T.a.r. Lazio-Roma, sez. III-*bis*, 14 febbraio 2017, n. 3769; T.a.r. Lazio-Roma, sez. III, 6 giugno 2019, n. 7333; T.a.r. Lazio-Roma, sez. III, 1° luglio 2020, n. 7526. D'altra parte, nella pronuncia Cons. Stato, sez. VI, 2 gennaio 2020, n. 30 si evidenzia che «a prescindere dal merito della controversia – e, quindi, dalla possibilità sia di qualificare l'algoritmo quale documento amministrativo suscettibile di accesso, sia di accordare prevalenza alle esigenze di trasparenza amministrativa rispetto a quelle di riservatezza della parte controinteressata –, emerge, in via pregiudiziale, la necessità di qualificare l'ideatore di un algoritmo oggetto di istanza di accesso come parte controinteressata all'ostensione, potendo questi, in caso di esibizione, vedere compromesso il diritto a mantenere segreta la regola tecnica in cui si sostanzia la propria creazione. [...] Nel caso di specie non risulta che l'Amministrazione, richiesta dell'accesso all'algoritmo per cui è controversia, abbia individuato un soggetto, controinteressato sul piano sostanziale, inviandogli apposita comunicazione al fine di consentire la formulazione di una motivata opposizione ai sensi dell'art. 3, co. 2, d.P.R. n. 184/06».

²⁷ Nello specifico, nella pronuncia Cons. Stato, sez. VI, 8 aprile 2019, n. 2270, si afferma che «la regola tecnica che governa ciascun algoritmo resta pur sempre una regola amministrativa generale, costruita dall'uomo e non dalla macchina, per essere poi (solo) applicata da quest'ultima, anche se ciò avviene in via esclusiva. Questa regola algoritmica, quindi: possiede una piena valenza giuridica

Diversamente dalla prima categoria analizzata, gli algoritmi non deterministici applicano un metodo statistico/probabilistico, per effettuare una proiezione e individuare così «lo stato più probabile verso cui il mondo evolverà a partire da un certo stato osservato nel presente»²⁸. Proprio per questo, il processo decisionale dagli stessi regolato, per quanto tracciabile e ricostruibile *ex post* se le scelte non deterministiche sono memorizzate, non è prevedibile *ex ante*, essendo almeno una delle istruzioni di cui tale processo si compone aperta a una pluralità di possibili passaggi successivi, «ognuno dei quali ha una certa probabilità di essere scelto»²⁹. Tale potrebbe essere un algoritmo che calcola, per una determinata consegna da effettuare, una stima del tempo necessario o che fornisca una previsione in tempo reale della domanda di beni e servizi con conseguente determinazione della ripartizione di carichi di lavoro in una catena di produzione o della distribuzione delle merci all'interno di un magazzino. A un algoritmo non deterministico può essere agganciato, in successione, un algoritmo deterministico, laddove in ipotesi sia previsto un processo decisionale automatizzato che assegni un determinato punteggio ai lavoratori in base alla puntualità della consegna rispetto al tempo stimato. È importante evidenziare, laddove la previsione diventi prescrizione, contenuto del potere conformativo datoriale e quindi parametro della diligenza richiesta al lavoratore per poter essere considerato adempiente, come la conclusione raggiunta attraverso un'inferenza probabilistica anzitutto sia influenzata dalla composizione del campione osservato e possa, pertanto, condizionare la decisione su un lavoratore, partendo da dati relativi al comportamento di altre persone o altri gruppi di persone che non ne rappresen-

e amministrativa, anche se viene declinata in forma matematica, e come tale, come si è detto, deve soggiacere ai principi generali dell'attività amministrativa, quali quelli di pubblicità e trasparenza (art. 1 l. 241/90), di ragionevolezza, di proporzionalità, etc.». Cfr., al riguardo, N. ABRIANI-G. SCHNEIDER, *Diritto delle imprese e intelligenza artificiale. Dalla Fintech alla Corptech*, il Mulino, Bologna, 2021, spec. pp. 54-57; G. ZAMPINI, *Intelligenza artificiale*, cit. Nell'elaborazione della giurisprudenza amministrativa si evidenzia la peculiarità delle funzioni assolte dalle garanzie di trasparenza e supervisione umana richieste ai fini dell'utilizzo dell'algoritmo nello svolgimento dell'attività amministrativa: il ricorso a tale «modulo organizzativo» non può, infatti, comportare l'elusione delle «guarentigie» e delle «verifiche tipiche» di ogni procedimento amministrativo, in particolare dei canoni di trasparenza, partecipazione, collaboratività, nonché dell'obbligo di motivazione (cfr. *ex multis* T.a.r. Campania-Napoli, sez. III, 14 novembre 2022, n. 7003; Cons. Stato, sez. VI, 13 dicembre 2019, n. 8472). Da qui la necessità che, da un lato, il funzionamento dell'algoritmo sia comprensibile e l'informazione al riguardo accessibile, dall'altro, che i responsabili del procedimento e del provvedimento garantiscano il «presidio umano sull'azione della macchina» anche nel caso si abbia «la certezza del buon funzionamento dell'algoritmo» (D. MARONGIU, *Algoritmo e procedimento amministrativo: una ricostruzione*, in *Giurisprudenza italiana*, 2022, p. 1521).

²⁸ G. D'ACQUISTO, *Intelligenza artificiale. Elementi*, Giappichelli, Torino, 2021, p. 83.

²⁹ G. AVANZINI, *Decisioni amministrative*, cit., p. 7; M. PALMIRANI, *Big Data e conoscenza*, in *Rivista di filosofia del diritto*, 2020, p. 73 ss.; S.J. RUSSELL-P. NORVIG, *Artificial Intelligence: A Modern Approach (AIMA)*, Pearson, New Jersey, 2009.

tano adeguatamente i reali tratti personali³⁰. Ancora, e il tema assumerà implicazioni tanto più rilevanti in caso di algoritmi di apprendimento automatico, la soluzione individuata dal sistema si appoggia sulla presenza di una dipendenza probabilistica tra le variabili, che non deriva necessariamente dalla presenza di un nesso di causalità, ma soltanto da una concomitanza molto elevata di eventi³¹. In linea generale, la categoria degli algoritmi non deterministici (se non inseriti all'interno di un modello di apprendimento automatico) può essere annoverata tra i sistemi comprensibili e spiegabili, dotati quindi di un certo grado di *c.d. transparency by design*.

Gli algoritmi di apprendimento automatico si differenziano dalle categorie precedenti in quanto il modello che governa il processo decisionale non è definito, in questo caso, da regole predeterminate in fase di programmazione dell'algoritmo stesso, ma è generato durante la fase di addestramento. Il funzionamento del sistema si basa certamente sull'esecuzione di un «algoritmo addestratore», scritto in linguaggio di programmazione, in sé intellegibile, che consegna alla macchina un «metodo di apprendimento»³². Dall'analisi automatizzata dei *c.d. training data*, ossia delle informazioni immesse o raccolte durante la fase di allenamento, il sistema estrae, seguendo un processo induttivo, l'«algoritmo addestrato», lo adegua e ottimizza progressivamente. «Impara» così la regola, il modello matematico-numerico, che garantisce in termini probabilistici la migliore utilità attesa e che andrà a governare poi, nella fase di esecuzione vera e propria, lo svolgimento di un certo compito, ossia la procedura di determinazione dell'*output*³³. È proprio all'interno di questa categoria che secondo alcuni la capacità del sistema algoritmico di eseguire un processo decisionale senza l'intervento umano passa da una dimensione di mera automazione, dove l'operato del sistema rimane sia predefinito e vincolato da regole codificate in sede di programmazione sia circoscritto ai confini della rappresentazione formale della conoscenza fornita³⁴, a proprietà maggiormente associabili al concetto di «autonomia»³⁵.

³⁰ Sul concetto di profilazione e sui limiti che può presentare il sistema di garanzie del GDPR, in particolare il diritto di accesso, rispetto all'uso di dati collettivi anonimizzati v. cap. 2, § 2.1.1.

³¹ G. D'ACQUISTO, *Intelligenza artificiale*, cit., p. 76.

³² G. CARULLO, *Decisione algoritmica e intelligenza artificiale*, in *Il diritto dell'informazione e dell'informatica*, 2021, p. 439; G. SARTOR-F. LAGIOIA, *Le decisioni algoritmiche tra etica e diritto*, in RUFFOLO (a cura di), *Intelligenza artificiale. Il diritto, i diritti, l'etica*, Giuffrè, Milano, 2020, p. 69 ss.; A. MERONE, *Fondamenti di machine learning e applicazioni giuridiche*, in R. GIORDANO-A. PANZAROLA et al. (a cura di), *Il diritto nell'era digitale*, Giuffrè, Milano, 2022, p. 1045 ss.

³³ Cfr. N. ABRIANI-G. SCHNEIDER, *Diritto delle imprese*, cit., p. 37.

³⁴ G. SARTOR-F. LAGIOIA, *Le decisioni algoritmiche*, cit., p. 68.

³⁵ M. GABBRIELLI, *Dalla logica al deep learning: una breve riflessione sull'intelligenza artificiale*, in U. RUFFOLO (a cura di), *XXVI Lezioni di Diritto dell'Intelligenza artificiale*, Giappichelli, Torino, 2021, p. 29.

Le modalità di apprendimento automatico possono essere di tre tipologie.

Nell'apprendimento con supervisione, l'algoritmo è addestrato con un insieme di dati precedentemente etichettati, un insieme che contiene quindi tutti i dati che si ritiene caratterizzino il fenomeno da analizzare, già distinti in variabili di ingresso e valori di uscita. Classico esempio è quello dell'algoritmo addestrato su immagini di cani e di gatti – già etichettate con l'esplicita informazione della categoria a cui appartengono – che apprende in autonomia le caratteristiche comuni di ciascun insieme di immagini, costruendo un modello per riconoscere, classificare e distinguere eventuali nuovi dati in ingresso³⁶. In tal senso, si tratta di un apprendimento «che può far leva sulla conoscenza delle risposte corrette»³⁷: «l'algoritmo viene [infatti] allenato a riconoscere e ad associare ai dati di *input* una etichetta/classe/categoria di *output* che in fase di allenamento è conosciuta e quindi verificabile»³⁸. L'algoritmo è chiamato ad “apprendere”, individuare la legge, la funzione matematica, che lega gli *input* alla classe di *output* di appartenenza, in modo da poter stimare, predire il valore di una variabile target al ricorrere di nuove variabili in ingresso (algoritmo di regressione lineare) ovvero la classe di appartenenza di un nuovo elemento osservato, a fronte della specifica combinazione di caratteristiche che lo connotano (algoritmo di classificazione, come nell'esempio). La tecnica dell'apprendimento supervisionato trova una specifica applicazione nel campo della manutenzione predittiva, consentendo ad esempio di prevedere il tempo entro il quale si potrebbe verificare un guasto (se algoritmo di regressione lineare) ovvero la possibilità che un guasto si verifichi in una certa finestra temporale (se algoritmo di classificazione). Parimenti può risultare funzionale nel più ampio contesto della simulazione predittiva, un processo realizzabile anche attraverso i c.d. *digital twin* (e nel metaverso³⁹), che permette di costruire e

³⁶ Cfr. G.F. ITALIANO-S. CIVITARESE MATTEUCCI-A. PERRUCCI, *L'intelligenza artificiale*, cit., spec. pp. 50-51.

³⁷ R. ROVATTI, *Il processo di apprendimento algoritmico e le applicazioni nel settore legale*, in U. RUFFOLO (a cura di), *XXVI Lezioni di Diritto dell'Intelligenza artificiale*, Giappichelli, Torino, 2021, p. 37.

³⁸ A. LOREGGIA-G. SARTOR, *L'Intelligenza Artificiale nella moderazione del digitale*, in *Sistemi intelligenti*, 2022, p. 55.

³⁹ M. BALL definisce il Metaverso come «una rete di massima scalabilità e interoperabile di mondi virtuali 3D renderizzati in tempo reale, che possono essere vissuti in modo sincrono e persistente da un numero effettivamente illimitato di utenti con un senso individuale di presenza al loro interno, e che garantiscono la continuità dei dati relativi a identità, storia, diritti, oggetti, comunicazioni e pagamenti» (M. BALL, *Metaverso*, Garzanti, Milano, 2022, p. 55). Elementi chiave sono individuati nella simulazione di un ambiente, generato dal computer in 3D/3D immersivo, completamente persistente laddove tutto ciò che accade al suo interno permanga nel tempo e non venga invece resettato per ciascun utente, visualizzabile nelle sue diverse componenti, a 360 gradi e secondo molteplici dimensioni di movimento, in risposta agli *input* degli utenti, vivibile e condivisibile da questi in modo sincronico e simultaneo, con un senso di presenza dato da un'esperienza immersiva che può coinvolgere la vista e il tatto (e in alcuni casi anche l'olfatto).

ottimizzare, sulla base dei dati raccolti in tempo reale anche durante la fase di funzionamento dell'oggetto fisico reale (ad es. i rumori, le vibrazioni, i cambi di temperatura), un modello di proiezione digitale della sua prossima evoluzione. Alcune tecniche di approccio supervisionato sono considerate *transparent by design*, quantomeno per il soggetto esperto; altre richiedono l'applicazione *ex post* di specifiche tecniche esterne di *XAI* per poter estrarre la spiegazione⁴⁰.

Nell'apprendimento non supervisionato, i dati in ingresso non sono etichettati. L'algoritmo in questo caso individua sottoinsiemi di elementi omogenei, creando i raggruppamenti tra i dati, i c.d. *cluster*, sulla base dell'estrazione di ricorrenze, pattern comuni, similarità (c.d. *clustering*), prevedendo poi l'appartenenza o prossimità dei nuovi dati di ingresso a un certo insieme. In un secondo momento, una volta effettuata la suddivisione, l'esperto umano è chiamato a interpretare, nominare gli insiemi, i grappoli di dati trovati, partendo dal riconoscimento delle specifiche caratteristiche che li connotano. Sono algoritmi utilizzati, ad esempio, per effettuare la profilazione e targetizzazione della clientela.

Nell'apprendimento per rinforzo, l'algoritmo elabora una legge di selezione delle azioni da intraprendere, che massimizzi la funzione obiettivo assegnata e garantisca progressivamente la migliore utilità attesa, a partire da un'esplorazione e interazione dinamica e continua con l'ambiente circostante, nella quale a ogni tentativo corrispondono rinforzi positivi (ricompense) o negativi (penalità) e, quindi, un punteggio che misura la qualità dell'azione. Sono algoritmi utilizzati, ad esempio, per l'addestramento della navigazione dei robot.

Come evidenzia la dottrina, in tutte le tecniche di apprendimento automatico «decisiva è la rilevanza dei dati», siano essi i *training data*, «con cui viene sviluppato l'algoritmo», gli *input data*, «forniti all'algoritmo una volta sviluppato per compiere una specifica scelta e dunque emettere le relative previsioni», ovvero i *feedback data*, «immessi nel sistema per perfezionarlo nel corso della sua operatività». Il sistema di *machine learning* riesce a «qualificare autonomamente informazioni non ricomprese nel *dataset* di *training* iniziale e inserite nel sistema successivamente», e questo «in base non già ai criteri codificati dagli sviluppatori dell'algoritmo, bensì criteri che la stessa macchina elabora autonomamente»⁴¹.

Lo strumento più efficace per il funzionamento dei meccanismi di apprendimento automatico, all'aumentare della complessità del fenomeno da analizzare e

Se un *digital twin* può essere utilizzato anche al di fuori del metaverso, quest'ultimo necessita, invece, inevitabilmente di tale rappresentazione digitale. Proseguendo il discorso nel testo, questo strumento può permettere simulazioni preventive e test (c.d. *virtual commissioning*), che possono evitare lesioni alla salute nel mondo reale, così come eliminare l'esposizione a determinati rischi attraverso il mero contatto con le copie virtuali.

⁴⁰ A.B. ARRIETA-N. DÍAZ-RODRÍGUEZ *et al.*, *Explainable Artificial Intelligence*, cit., p. 94.

⁴¹ N. ABRIANI-G. SCHNEIDER, *Diritto delle imprese*, cit., pp. 25-26.