



Sorveglianza umana e diligenza nel controllo dell'algoritmo

LOREDANA ZAPPALÀ

Università degli Studi di Catania
Associata di Diritto del Lavoro
lzappala@lex.unict.it

ABSTRACT

Starting from the anthropocentric principle underpinning the European regulatory framework for artificial intelligence, the article analyses how human oversight reshapes the object of the employment obligation by extending it to the critical assessment of algorithmic systems. It argues that this individual obligation cannot be understood in isolation but must be conceived as a specific component of a broader organisational obligation incumbent upon the employer, consisting in the establishment of the organisational arrangements necessary to make human oversight genuinely effective. The proceduralisation of risk management, training, professional competences and supervisory activities progressively gives substance to the employee's duty of oversight and contributes to defining the applicable standard of care. Consequently, the employee's liability may arise only where the employer has previously demonstrated that all organisational arrangements and technologically feasible measures necessary to ensure effective human oversight have been properly implemented.

Keywords: human oversight; artificial intelligence; employment obligation; high-risk systems; standard of care.

<https://doi.org/10.60923/issn.2421-2695/25824>

Sorveglianza umana e diligenza nel controllo dell'algoritmo

SOMMARIO: 1. Sorveglianza umana e riserva di umanità: la centralità del diritto del lavoro a presidio della affidabilità dei sistemi di IA. – 2. Controllo, governo della tecnica ed estensione degli obblighi del lavoratore addetto alla supervisione dell'IA. – 3. Dominio cognitivo della tecnica e ridefinizione dell'obbligazione lavorativa. – 4. *Standard* di diligenza nella supervisione umana e perimetro dell'obbligazione esigibile. – 5. La responsabilizzazione degli addetti alla supervisione umana: un rischio già attuale. – 6. La proceduralizzazione della supervisione umana nell'*AI Act* e nella direttiva piattaforme. – 7. La filiera organizzativa della supervisione umana nella normazione tecnica UNI. – 8. Sorveglianza umana e organizzazione del lavoro: verso un nuovo riparto delle responsabilità.

1. Sorveglianza umana e riserva di umanità: la centralità del diritto del lavoro a presidio della affidabilità dei sistemi di IA

La centralità del controllo umano sull'intelligenza artificiale costituisce oggi uno dei principali punti di convergenza tra riflessione filosofica⁽¹⁾, elaborazione etica⁽²⁾ e regolazione giuridica³. Se, sul piano filosofico, il rapporto tra uomo e tecnica ha da sempre interrogato la capacità dell'essere umano di governare gli strumenti da lui stesso creati, e, sul piano etico, ha alimentato il dibattito sui limiti della delega delle decisioni alle macchine, il diritto europeo ha progressivamente trasformato tale questione in un preciso obiettivo regolatorio: assicurare che lo sviluppo e l'impiego dei sistemi di intelligenza artificiale non determinino uno spostamento del centro decisionale dalla persona alla tecnologia, preservando la "riserva di umanità"⁽⁴⁾, vale a dire la necessità di garantire nel processo

⁽¹⁾ Sul piano filosofico, v. L. Floridi, *La differenza fondamentale. Artificial Agency: una nuova filosofia dell'intelligenza artificiale*, Mondadori, 2025, secondo il quale l'intelligenza artificiale non costituisce una forma di intelligenza assimilabile a quella umana, bensì una peculiare forma di *artificial agency*: un sistema capace di agire e produrre effetti senza possedere coscienza, comprensione o capacità di giudizio. Proprio tale distinzione costituisce uno dei principali fondamenti teorici dell'esigenza di mantenere la responsabilità e il controllo in capo alla persona umana; S. Epifani, *Il teatro delle macchine pensanti*, Digital Trasformation Institute, 2025.

⁽²⁾ Sul piano etico, è significativo come le più recenti riflessioni sul rapporto tra persona e intelligenza artificiale abbiano progressivamente spostato l'attenzione dalla mera affermazione della centralità dell'essere umano alla costruzione di una responsabilità effettiva nei confronti delle decisioni algoritmiche. In questa prospettiva si colloca l'Enciclica *Magnifica Humanitas*, 2026, nella quale la dignità della persona viene assunta quale limite invalicabile di ogni sviluppo tecnologico e la tecnica è considerata autenticamente umana solo nella misura in cui continui a riconoscere nell'uomo il soggetto del giudizio, della responsabilità e del discernimento (punti da 102 a 111). Sulla Enciclica v. *Primi Appunti Sull'enciclica "Magnifica Humanitas" Di Papa Leone Xiv*, LDE, 2026, con i commenti di T. Treu, G.M. Ballistreri, M. Biasi, L. Corazza, G. Mari, A. Tursi, S. Zamagni, S. Caponnetti, M. Ferraresi, D. Fumarola, M. Marchesini, T. Orrù, C. Pisani, A. Sitzia, B. Bignami.

⁽³⁾ Per alcuni riferimenti essenziali T. Treu, *Il controllo umano delle tecnologie: regole e procedure*, WP CSDLE "Massimo D'Antona".IT - 492/2025; M. Peruzzi, *Intelligenza artificiale e lavoro*, Giappichelli, 2024.

⁽⁴⁾ Per una riflessione a partire dall'attività amministrativa v. G. Gallone, *Riserva di umanità e funzioni amministrative*, Wolters Kluwer, 2023; G. Gallone, *Digitalizzazione, amministrazione e persona: per una "riserva di*

decisionale un contributo umano capace di controllare, validare ovvero smentire la decisione automatizzata, con diritto del soggetto controllato di esprimere la propria opinione e contestare la decisione assunta.

Questa prospettiva costituisce il tratto distintivo dell'approccio europeo all'intelligenza artificiale. Sin dal Libro bianco del 2020, la Commissione europea ha individuato nella costruzione di un'intelligenza artificiale antropocentrica e affidabile la condizione per coniugare innovazione tecnologica e tutela dei diritti fondamentali, affermando che la fiducia nelle tecnologie digitali presuppone il rispetto della dignità umana, dello Stato di diritto e dei valori costituzionali comuni. Il principio antropocentrico, dunque, non assume una funzione meramente assiologica, ma rappresenta il criterio ordinatore dell'intero impianto regolatorio europeo, destinato a orientare sia la progettazione dei sistemi di IA, sia le modalità della loro integrazione nei processi decisionali pubblici e privati.

Questa impostazione trova oggi la sua compiuta espressione, a livello europeo, nell'art. 14, Reg. (UE) 2024/1689⁽⁵⁾, che impone ai fornitori e ai *deployer* dei sistemi di IA ad alto rischio di predisporre misure idonee ad assicurare una supervisione umana effettiva, e riceve un'ulteriore e più penetrante declinazione nella direttiva (UE) 2024/2831 sul lavoro mediante piattaforme digitali⁽⁶⁾, che – negli artt. 10 e 11 – declina le modalità della supervisione e del riesame umano. Il principio è stato recepito nell'ordinamento italiano dall'art. 3, co. 3, l. n. 132/2025.

Anche sul terreno delle relazioni collettive, già l'Accordo quadro delle parti sociali europee sulla digitalizzazione del 22 giugno 2020⁽⁷⁾ ha individuato, tra i quattro assi portanti della trasformazione digitale del lavoro, il principio di controllo dell'uomo, affermando come tale il controllo degli esseri umani sulle macchine e sui sistemi di IA de essere garantito nei luoghi di lavoro e che lo stesso debba essere inserito all'interno di un più ampio processo partecipato di progettazione dell'organizzazione del lavoro, fondato sul

umanità" tra spunti codicistici di teoria giuridica dell'automazione, *Pubblica Amministrazione, Persona e Amministrazione*, 2023, n. 1, 329 e ss.

⁽⁵⁾ La letteratura è ormai vastissima, per un riferimento essenziale v. O. Pollicino - F. Domani - G. Finocchiaro - F. Paolucci (a cura di), *La disciplina dell'intelligenza artificiale*, Giuffrè Francis Lefebvre, 2025; sulla lettura lavoristica *Diritto del lavoro e intelligenza artificiale*, a cura di M. Biasi, Giuffrè Francis Lefebvre, 2025; M. Peruzzi, *Intelligenza artificiale e lavoro*, cit.; L. Tebano, *Intelligenza artificiale datore di lavoro: scenari e regole*, *DLM*, 2024, n. 3, 449 ss.; M. T. Carinci - A. Ingrao, *L'impatto dell'AI Act sul diritto del lavoro*, *DLRI*, 2024, 4, 451 ss.

⁽⁶⁾ Cfr. P. Tullini, *La Direttiva piattaforme e i diritti del lavoro digitale*, *LLI*, 2022, n. 1, 43 ss.; A. Alaimo, *Lavoro e piattaforme tra subordinazione e autonomia: la modulazione delle tutele nella proposta della Commissione europea*, *DRI*, 2022, 2, 639 ss.; V. Brino, *La tutela della persona che lavora nell'era dell'IA tra sfide etiche e giuridiche*, *DLM*, 3, 431 ss.; M. Novella - P. Tullini, *Lavoro digitale*, Giappichelli, 2026; nonché S. Borelli - V. Brino - C. Faleri - L. Lazzeroni - L. Tebano - L. Zappalà, *Lavoro e tecnologie. Dizionario del diritto del lavoro che cambia*, Giappichelli, 2022.

⁽⁷⁾ T. Treu, *La digitalizzazione del lavoro: proposte europee e piste di ricerca*, *Federalismi.it*, 2022, 190 ss.; C. Faleri, *Brevi spunti di riflessione sull'evoluzione delle relazioni sindacali nell'economia digitale*, *Ianus DF*, 2021, 24, 93 ss.; A. Rota, *Sull'Accordo quadro europeo in tema di digitalizzazione del lavoro*, *LLI*, 2020, 2, 25 ss.; M. Peruzzi, *Il dialogo sociale di fronte alle sfide della digitalizzazione*, *DRI*, 2020, 1213 ss.; P. Tullini, *L'economia digitale alla prova dell'interesse collettivo*, *LLI*, 2018, 1, 4 ss.

coinvolgimento delle parti sociali, sulla formazione, sulla valutazione dei rischi e sulla trasparenza dei sistemi algoritmici.

La medesima centralità sta trovando, poi, progressiva affermazione anche nella contrattazione collettiva del pubblico impiego. In particolare, l'ipotesi di CCNL del Comparto Funzioni Centrali del 9 giugno 2026, all'art. 14, co. 1, ribadisce come l'utilizzo di sistemi di IA non possa dar luogo a decisioni esclusivamente automatizzate che producano effetti giuridici o incidano in modo significativo sul rapporto di lavoro del dipendente, senza un appropriato e congruo intervento umano, così confermando detto controllo stia progressivamente evolvendo da principio regolatorio europeo a criterio di governo dell'organizzazione del lavoro e delle relazioni industriali *tout court*.

Un'ulteriore conferma della progressiva giuridificazione del principio del necessario coinvolgimento umano proviene dalla giurisprudenza amministrativa. Già da alcuni anni, il giudice amministrativo ha iniziato a delimitare il perimetro di legittimità delle decisioni adottate mediante procedure automatizzate, individuando nella permanenza di un effettivo controllo umano il presupposto imprescindibile perché l'esercizio del potere continui ad essere imputabile all'amministrazione⁽⁸⁾. Il principio della supervisione umana cessa così di essere soltanto un criterio di progettazione dei sistemi di intelligenza artificiale e diviene un parametro di legittimità dell'azione amministrativa, destinato a garantire che l'automazione non si traduca nell'abdicazione della funzione decisionale, ma rimanga uno strumento al servizio di una decisione che conserva natura e responsabilità umane.

Da questo approccio discendono diverse conseguenze di particolare rilievo per il diritto del lavoro. La prima riguarda il ruolo che quest'ultimo è chiamato ad assumere nella concreta attuazione del principio antropocentrico. La regolazione europea individua nella supervisione umana il principale presidio affinché l'intelligenza artificiale rimanga al servizio della persona, definendo solo in parte i contenuti organizzativi e le modalità di esercizio.

È proprio il diritto del lavoro, allora, a divenire il luogo privilegiato nel quale il principio della supervisione umana acquista effettività, attraverso la disciplina dell'organizzazione del lavoro, delle competenze professionali, della formazione, dei poteri di intervento e delle responsabilità dei soggetti chiamati ad esercitarla. Al diritto del lavoro è affidata una funzione che si aggiunge a quelle tradizionalmente assolute nell'ordinamento

⁽⁸⁾ In tale direzione si colloca il consolidato orientamento del Consiglio di Stato, secondo cui il ricorso a procedure algoritmiche è compatibile con i principi dell'azione amministrativa soltanto ove l'Amministrazione conservi un effettivo potere di controllo, verifica e imputazione della decisione, restando esclusa la possibilità di affidare l'esercizio del potere a meccanismi di mera automazione (la vicenda era quella, ben nota, dell'algoritmo del MIUR che ha gestito il trasferimento dei docenti della scuola, v. Cons. St. 4 febbraio 2020, n. 881 che ha consolidando le precedenti decisioni Cons. St., sez. VI, 8 aprile 2019, n. 2270 e Cons. St., sez. VI, 13 dicembre 2019, n. 8472). La più recente giurisprudenza mostra un ulteriore avanzamento di tale impostazione. In particolare, TAR Sicilia, Catania, sez. III, 22 aprile 2026, n. 1157, in materia di contratti pubblici, e TAR Lazio, Roma, sez. III-bis, 2 febbraio 2026, n. 1895, in materia concorsuale, valorizzano il principio della supervisione umana quale criterio interpretativo dell'ordinamento, sino a riconoscere nella "riserva di umanità" il limite oltre il quale l'automazione non può spingersi, anche con riguardo a vicende anteriori alla piena applicazione del Reg. (UE) 2024/1689.

– dalla tutela della posizione contrattuale del prestatore alla conformazione dell'esercizio delle prerogative datoriali, sino alla garanzia del corretto funzionamento del mercato del lavoro – per assumere una diversa e più ampia dimensione di interesse generale. Esso è chiamato, infatti, a dare concreta attuazione a uno dei principi ordinanti della regolazione europea dell'intelligenza artificiale, concorrendo a garantire l'effettività del principio antropocentrico attraverso la costruzione delle condizioni organizzative e la proceduralizzazione dell'esercizio delle prerogative datoriali che rendono possibile un controllo umano effettivo delle decisioni algoritmiche.

L'attuazione di tale principio non si esaurisce, tuttavia, come si vedrà, nell'individuazione del soggetto chiamato a esercitare la supervisione, ma investe, più a monte, la costruzione degli assetti organizzativi e professionali che rendono quel controllo concretamente possibile. È proprio lungo questa direttrice che il diritto del lavoro è chiamato a confrontarsi con questioni inedite concernenti il contenuto dell'obbligazione lavorativa, gli *standard* di diligenza e il riparto delle responsabilità nell'utilizzo dei sistemi di intelligenza artificiale.

2. Controllo umano, governo della tecnica ed estensione degli obblighi del lavoratore addetto alla supervisione dell'IA

La nuova funzione affidata al diritto del lavoro apre un secondo e, ancora poco esplorato, terreno di riflessione. Se l'effettività del principio antropocentrico dipende dalla concreta possibilità che una persona eserciti un controllo consapevole sulle decisioni algoritmiche, la supervisione umana cessa di rappresentare una mera garanzia esterna del sistema e diviene essa stessa oggetto di organizzazione. L'obbligo, imposto al datore di lavoro, di assicurare un controllo umano effettivo presuppone infatti l'individuazione di lavoratori chiamati a svolgere tale funzione, la definizione delle competenze richieste, la predisposizione di adeguati percorsi formativi, l'attribuzione di effettivi poteri di intervento e la costruzione di procedure che rendano il controllo realmente esercitabile e non meramente nominale.

Da questa premessa discende una conseguenza di particolare rilievo. Spostando il punto di osservazione dal lato del lavoratore addetto a tale attività emerge che la supervisione dell'algoritmo entra progressivamente a far parte dell'oggetto della prestazione lavorativa, modificando il contenuto delle mansioni affidate ai soggetti chiamati ad esercitarla. Il lavoratore non è più soltanto destinatario, utilizzatore o esecutore di decisioni supportate dall'intelligenza artificiale, ma diviene il titolare di una funzione di garanzia, chiamato ad assicurare che il processo decisionale automatizzato rimanga conforme ai presupposti di liceità, affidabilità e centralità della persona posti dal legislatore europeo. La sua prestazione non si esaurisce, pertanto, nell'esecuzione di un'attività produttiva da eseguire con la diligenza parametrata al tipo di prestazione, ma si estende all'esercizio di un giudizio critico sull'operato dell'algoritmo, alla capacità di riconoscerne i

limiti, di valutarne gli esiti, di individuarne le anomalie e, ove necessario, di sospenderne o correggerne gli effetti, assumendo la responsabilità della decisione finale.

La centralità della supervisione umana ridefinisce, da questa prospettiva, il rapporto tra diritto del lavoro e tecnica⁽⁹⁾. Se nel paradigma tradizionale il lavoratore subordinato è chiamato ad utilizzare la tecnica organizzata e prescelta dal datore di lavoro nell'ambito dell'esercizio del potere direttivo, l'affermazione del principio antropocentrico aggiunge oggi una dimensione ulteriore alla prestazione. Il lavoratore incaricato della supervisione non è più soltanto tenuto a gestire correttamente la tecnologia impiegata nell'organizzazione produttiva, ma è chiamato a esercitare su di essa una vera e propria funzione di dominio cognitivo, consistente nella capacità di comprenderne il funzionamento, valutarne criticamente gli esiti e intervenire ogniqualvolta il sistema si discosti dai presupposti che ne giustificano l'impiego.

Tale funzione trascende il tradizionale controllo sulla corretta esecuzione della prestazione lavorativa. Il supervisore è chiamato, infatti, non soltanto a verificare che il sistema realizzi gli obiettivi organizzativi perseguiti dal datore di lavoro e che continui a soddisfare le esigenze di efficienza per le quali è stato introdotto, ma soprattutto ad assicurare che l'attività algoritmica rimanga compatibile con i limiti posti dall'ordinamento. Il controllo umano diviene così il luogo nel quale la razionalità computazionale viene costantemente confrontata con la razionalità umana e giuridica, imponendo al lavoratore di riconoscere eventuali inferenze irragionevoli, correlazioni prive di fondamento causale, effetti discriminatori derivanti da *bias* nei dati o nei modelli, compressioni sproporzionate dei diritti fondamentali o, più in generale, esiti incompatibili con i principi di liceità, proporzionalità, trasparenza e ragionevolezza che presidiano l'impiego dell'intelligenza artificiale.

Tale funzione di dominio cognitivo attribuisce, pertanto, al lavoratore un ruolo inedito. Egli non è più soltanto il soggetto che opera all'interno dell'organizzazione tecnica predisposta dal datore di lavoro, ma diviene il garante del permanere di quella tecnologia entro i limiti della razionalità umana. L'oggetto della prestazione si estende, così, dall'esecuzione dell'attività produttiva al controllo critico del processo decisionale automatizzato, trasformando il lavoratore in un presidio permanente dell'affidabilità del sistema e della conformità del suo operare ai valori e ai principi che ispirano il modello europeo di intelligenza artificiale.

La supervisione umana rivela così una portata ben più ampia di quella che emerge dalla sola lettura delle disposizioni europee. Essa non introduce soltanto un nuovo obbligo organizzativo gravante sul datore di lavoro, ma determina una trasformazione qualitativa dell'obbligazione lavorativa di coloro ai quali tale funzione viene affidata. Il controllo dell'algoritmo, per i lavoratori la cui prestazione si interseca con sistemi di IA, diviene parte

⁽⁹⁾ Sul rapporto tra tecnica, organizzazione del lavoro e subordinazione rimane imprescindibile G. Vardaro, *Tecnica, tecnologia e ideologia della tecnica nel diritto del lavoro, Politica del diritto*, 1986, il quale mostra come la tecnica non costituisca un fattore esterno al diritto del lavoro, ma uno degli elementi strutturali attraverso i quali si conformano l'organizzazione produttiva e l'esercizio del potere direttivo.

integrante della prestazione dedotta nel contratto e richiede una professionalità nuova, nella quale assumono rilievo non solo le competenze tecniche necessarie a comprendere il funzionamento del sistema, ma anche la capacità di esercitare un'autonoma valutazione critica e di assumere decisioni motivate rispetto agli esiti prodotti dall'intelligenza artificiale.

È proprio lungo questa direttrice che si colloca il presente contributo. Se il diritto europeo individua nella supervisione umana il presidio dell'approccio antropocentrico, resta infatti da comprendere quale sia il contenuto giuridico di tale funzione nel rapporto di lavoro. L'ingresso della supervisione nell'oggetto della prestazione apre questioni del tutto nuove riguardanti la configurazione delle mansioni, gli *standard* di diligenza richiesti al lavoratore, le modalità di verifica dell'adempimento, l'obbligo di motivare l'eventuale discostamento dalla decisione algoritmica e, più in generale, i criteri di imputazione delle responsabilità derivanti dall'esercizio – o dall'omesso esercizio – dell'attività di controllo. Il rischio, diversamente, è che la supervisione umana rimanga una mera clausola di stile o, peggio, si trasformi in un meccanismo di traslazione della responsabilità dall'organizzazione al singolo lavoratore, chiamato a rispondere di un controllo che l'ordinamento pretende effettivo, ma del quale non definisce ancora contenuti, modalità di esercizio e *standard* professionali.

3. Dominio cognitivo della tecnica e ridefinizione dell'obbligazione lavorativa

La supervisione umana apre poi un ulteriore e più complesso terreno di riflessione, che investe direttamente la teoria dell'obbligazione lavorativa. L'attribuzione al lavoratore di una funzione di controllo sull'intelligenza artificiale ripropone, infatti, uno dei temi classici del diritto del lavoro, quello del rapporto tra l'intrinseca elasticità della prestazione e la necessità di oggettivarne il contenuto, ma lo fa entro coordinate profondamente mutate. Non è più soltanto in discussione la misura nella quale l'organizzazione possa incidere sul contenuto della prestazione, bensì la possibilità stessa di ricomprendere nell'oggetto del contratto un'attività di natura cognitiva, consistente nell'esercizio di un giudizio critico sull'operato della tecnica: in questo caso su sistemi di IA strutturalmente “scudati” in una *black box* o che autoapprendono e modificano il loro comportamento in corso d'opera.

La questione si inserisce in un dibattito di lunga tradizione⁽¹⁰⁾. La struttura dell'obbligazione lavorativa è stata infatti costantemente attraversata dalla tensione tra due esigenze solo apparentemente antitetiche. Da un lato, l'inserimento della prestazione nell'organizzazione produttiva rende fisiologicamente dinamico il contenuto del contratto, che non può essere integralmente predeterminato al momento della sua conclusione, ma che è destinato a specificarsi progressivamente attraverso l'esercizio del potere direttivo,

⁽¹⁰⁾ Si rinvia a L. Zappalà, *La colpa del lavoratore*, Giappichelli, 2020.

l'evoluzione delle tecnologie, delle modalità produttive e dell'organizzazione del lavoro⁽¹¹⁾. Dall'altro lato, proprio questa naturale adattabilità ha imposto alla riflessione giuslavoristica un costante sforzo di oggettivazione del comportamento dovuto, volto ad impedire che la flessibilità organizzativa si traduca nell'indeterminatezza del debito di lavoro e nell'illimitata espansione delle pretese creditorie del datore⁽¹²⁾.

L'equilibrio tra queste due esigenze ha storicamente rappresentato uno dei punti di maggiore elaborazione teorica del diritto del lavoro. Se l'elasticità dell'obbligazione costituisce il riflesso necessario dell'inserimento del lavoratore nell'organizzazione dell'impresa, l'oggettivazione del contenuto della prestazione continua ad assolvere la funzione di delimitare il perimetro del comportamento esigibile, rendendo conoscibile, verificabile e controllabile la misura dell'obbligazione assunta dal prestatore.

L'ingresso del dominio cognitivo della tecnica IA modifica, tuttavia, i termini di questo equilibrio. La questione non consiste più soltanto nel determinare fino a che punto il datore possa pretendere un diverso *facere* dal lavoratore, bensì nello stabilire se e in quale misura possa entrare nell'oggetto del contratto un'attività cognitiva rivolta al controllo di sistemi di intelligenza artificiale capaci di elaborare grandi quantità di dati, generare inferenze, produrre correlazioni e creare *algo-created* data destinati ad incidere su persone, diritti e processi decisionali, autoapprendere da se stessa e mutare in autonomia le modalità di assunzione delle decisioni.

La novità non risiede, dunque, nella semplice estensione quantitativa dell'obbligazione, ma nello spostamento del suo baricentro. L'elasticità della prestazione non riguarda più soltanto il contenuto operativo del lavoro, ma investe il processo cognitivo che il lavoratore è chiamato a sviluppare nell'esercizio della funzione di supervisione di un sistema che può cambiare continuamente. Ciò che entra progressivamente nell'oggetto del contratto non è un diverso modo di eseguire la prestazione, bensì un'attività di comprensione, valutazione critica e controllo della razionalità algoritmica.

È proprio su questo punto che il principio della supervisione umana rivela tutta la propria ambivalenza. Una prima lettura, di carattere minimale, considera la supervisione come un mero requisito procedimentale: è sufficiente che una persona sia collocata nel processo decisionale automatizzato e sia formalmente investita di poteri di verifica o di intervento. In questa ricostruzione, il contenuto dell'obbligazione lavorativa non subisce sostanziali trasformazioni, limitandosi ad assorbire nuovi compiti organizzativi funzionali all'utilizzo della tecnologia.

Una diversa lettura, che appare maggiormente coerente con la logica del principio antropocentrico, conduce invece ad esiti profondamente differenti. Se il controllo umano deve essere effettivo, il lavoratore non è chiamato semplicemente a presidiare il

⁽¹¹⁾ M. Persiani, *Contratto di lavoro e organizzazione*, Cedam, 1966.

⁽¹²⁾ M. Napoli, *La stabilità reale del rapporto di lavoro*, Franco Angeli, 1980; A. Viscomi, *Diligenza e prestazione di lavoro*, Giappichelli, 1997; A. Viscomi, *L'adempimento dell'obbligazione di lavoro tra criteri lavoristici e principi civilistici*, DLRI, 2010, 595 ss.

procedimento, ma ad esercitare un autentico dominio cognitivo sull'IA, verificando criticamente gli esiti del sistema, riconoscendo anomalie, *bias*, inferenze irragionevoli, correlazioni prive di fondamento e decisioni suscettibili di comprimere diritti fondamentali. Il comportamento richiesto trascende, così, il piano dell'esecuzione materiale della prestazione per collocarsi su quello del giudizio professionale che, a seconda del tipo di sistema di IA utilizzato, può essere molto complesso.

Le conseguenze delle due ricostruzioni sono radicalmente diverse. Nel primo caso, l'elasticità dell'obbligazione continua ad operare entro la tradizionale logica dell'organizzazione del lavoro, limitandosi ad incorporare nuovi compiti funzionali all'impiego dell'intelligenza artificiale. Nel secondo, invece, essa viene chiamata a soddisfare una finalità ulteriore. L'estensione del comportamento esigibile non risponde più esclusivamente all'interesse dell'impresa al corretto funzionamento dell'organizzazione produttiva, ma è funzionale anche alla realizzazione di interessi che l'ordinamento considera di carattere generale: preservare una riserva di umanità nei processi decisionali automatizzati, garantire il rispetto dei diritti fondamentali, prevenire discriminazioni algoritmiche, assicurare la trasparenza e la verificabilità delle decisioni e, in definitiva, rendere effettivo il principio antropocentrico sul quale si fonda la disciplina europea dell'intelligenza artificiale.

È proprio questo mutamento di funzione a rendere insufficiente il ricorso alle categorie tradizionali dell'obbligazione lavorativa. Quanto più l'ordinamento estende il contenuto della prestazione per affidare al lavoratore compiti di garanzia che trascendono l'interesse del creditore e si proiettano nella tutela di valori generali, tanto più diviene necessario delimitare il perimetro dell'attività cognitiva concretamente esigibile. In difetto, il rischio è duplice: da un lato, trasformare la supervisione umana in una formula meramente simbolica; dall'altro, consentire al datore di lavoro di trasferire impropriamente sul lavoratore, dietro il velo della supervisione, la responsabilità per l'affidabilità di sistemi di intelligenza artificiale la cui progettazione, selezione e governo organizzativo rimangono estranei alla sua sfera di controllo.

4. *Standard di diligenza nella supervisione umana e perimetro dell'obbligazione esigibile*

La riconduzione del dominio cognitivo della tecnica IA all'interno dell'obbligazione lavorativa apre una questione ulteriore, che investe direttamente la teoria della diligenza nell'adempimento. Una volta ammesso che il lavoratore possa essere chiamato ad esercitare un controllo critico sull'operato dell'intelligenza artificiale, occorre stabilire secondo quali criteri tale attività possa essere valutata e, correlativamente, entro quali limiti il suo eventuale difetto possa assumere rilevanza sul piano dell'inadempimento.

La questione presenta profili solo in parte affrontati dalla dottrina⁽¹³⁾. La tradizionale costruzione della diligenza professionale⁽¹⁴⁾ presuppone, infatti, l'esistenza di *standard* tecnici e sociali progressivamente consolidatisi all'interno della comunità professionale di riferimento. La misura dell'obbligazione non viene determinata caso per caso, ma rinvia ad un patrimonio di conoscenze, regole tecniche, prassi operative ed esperienze condivise, che consentono di oggettivare il comportamento esigibile dal prestatore.

È proprio su questo terreno che le categorie tradizionali mostrano i propri limiti. L'attività di supervisione dei sistemi di intelligenza artificiale sfugge, almeno allo stato, a quel processo di sedimentazione professionale che, nella generalità delle prestazioni lavorative, consente di oggettivare il comportamento diligente attraverso regole tecniche, prassi organizzative ed esperienze condivise. Il controllo dell'algoritmo si colloca, infatti, in uno spazio nel quale non esiste ancora una parametro di "normalità" della supervisione, né un patrimonio di *standard* professionali capaci di definire, con sufficiente determinatezza, quale debba essere l'intensità del controllo richiesto, quali verifiche debbano essere effettuate, quali anomalie il lavoratore sia concretamente tenuto a riconoscere e quando il mancato intervento possa assumere rilievo sul piano dell'inadempimento.

Ne deriva che viene meno proprio quel processo di necessaria oggettivazione del comportamento dovuto attraverso il quale la clausola generale della diligenza si traduce, tradizionalmente, in un parametro concretamente verificabile. L'assenza di *standard* professionali consolidati impedisce di risolvere il problema ricorrendo ai tradizionali criteri di valutazione della diligenza. Non è infatti sufficiente richiedere al lavoratore una generica attenzione o una maggiore prudenza nell'utilizzo dell'intelligenza artificiale. Il dominio cognitivo della tecnica implica attività – comprendere il funzionamento del sistema, valutarne criticamente gli esiti, riconoscere inferenze distorsive, individuare *bias*, motivare il dissenso rispetto alla decisione algoritmica – per le quali non esistono ancora modelli professionali generalmente condivisi, né criteri socialmente consolidati di corretto adempimento.

L'assenza di tali parametri pone, tuttavia, un ulteriore problema. In mancanza di criteri oggettivi di valutazione, il giudizio sull'adempimento tende inevitabilmente a rifluire nelle clausole generali⁽¹⁵⁾ di verifica della correttezza dell'adempimento secondo i criteri di correttezza e buona fede⁽¹⁶⁾. Si tratta, però, di un approdo che rischia di ampliare in misura difficilmente controllabile il perimetro dell'obbligazione lavorativa. Se il contenuto del dominio cognitivo rimane indeterminato, addirittura tale da ricomprendere anche il

⁽¹³⁾ Il tema è stato correttamente affrontato da M. Peruzzi, *Intelligenza artificiale e lavoro*, cit., 126 ss.

⁽¹⁴⁾ A. Viscomi, *Diligenza e prestazione di lavoro*, cit.; C. Cester, *La diligenza del lavoratore*, Art. 2104, in C. Cester - M.G. Mattarolo, *Diligenza e obbedienza del prestatore di lavoro*, in *Il Codice Civile. Commentario*, diretto da F. D. Busnelli, fondato da P. Schlesinger, Giuffrè, 2007.

⁽¹⁵⁾ Sulle quali resta fondamentali il riferimento P. Tullini, *Clausole generali e rapporto di lavoro*, Maggioli, 1990; P. Campanella, *Clausole generali e obblighi del prestatore di lavoro*, DLRI, 2015, 89 ss.

⁽¹⁶⁾ L. Montuschi, *L'applicazione giurisprudenziale del principio di correttezza e di buona fede nel rapporto di lavoro*, LD, 1996, 139 ss.; A. Perulli, *La buona fede nel diritto del lavoro*, RGL, 2002, 3 ss.

controllo di una macchina che autoapprende e che modifica i propri comportamenti, sarà il giudizio *ex post* sulla condotta del lavoratore a riempire di significato l'obbligo di supervisione, con il rischio di trasformare la clausola generale in uno strumento di continua espansione dell'area del comportamento esigibile.

Il pericolo è tanto più evidente quanto più si considera che il controllo richiesto al lavoratore ha ad oggetto sistemi algoritmici il cui funzionamento rimane, almeno in parte, opaco o comunque caratterizzato da elevata complessità tecnica. La difficoltà di ricostruire integralmente il percorso logico seguito dall'intelligenza artificiale generativa rischia, così, di tradursi in un paradossale irrigidimento della posizione del supervisore: proprio perché il processo decisionale della macchina non è pienamente intelligibile, potrebbe risultare più agevole imputare al lavoratore di non aver esercitato un controllo sufficientemente critico, sufficientemente penetrante o sufficientemente diligente. In tal modo, l'imperscrutabilità della *black box* finirebbe per riflettersi sulla misura dell'obbligazione lavorativa, ampliando l'area della responsabilità individuale attraverso il ricorso a clausole generali il cui contenuto rimane inevitabilmente elastico e indefinito.

È proprio questo rischio a rendere insufficiente una ricostruzione della supervisione umana affidata esclusivamente alla buona fede o alla diligenza professionale. Quanto più il legislatore attribuisce al lavoratore una funzione di garanzia nell'utilizzo dell'intelligenza artificiale, tanto più diviene necessario sottrarre il contenuto di tale funzione ad una determinazione esclusivamente giudiziale o *ex post*, individuando preventivamente criteri oggettivi che consentano di distinguere ciò che è realmente esigibile da ciò che, invece, appartiene alla responsabilità dell'organizzazione.

5. La responsabilizzazione degli addetti alla supervisione umana: un rischio già attuale

La definizione del perimetro dell'obbligazione esigibile costituisce il necessario antecedente logico della responsabilità del lavoratore. Come è noto, soltanto la preventiva determinazione del comportamento dovuto consente di stabilire quando il mancato o inesatto adempimento possa assumere rilevanza sul piano disciplinare, contrattuale o, nel lavoro pubblico, amministrativo-contabile. Tale relazione, tradizionalmente acquisita nella teoria dell'obbligazione lavorativa, assume una valenza del tutto peculiare nell'ambito della supervisione umana dei sistemi di intelligenza artificiale, poiché l'indeterminatezza del contenuto della prestazione tende inevitabilmente a riflettersi sull'estensione dell'area della responsabilità del lavoratore.

Ed è proprio in questa zona di confine che si manifesta il primo, concreto, rischio evolutivo dell'obbligo di supervisione umana. Prima ancora che siano stati elaborati *standard* professionali idonei a definire il contenuto della prestazione di supervisione diligente, iniziano infatti ad emergere modelli organizzativi che attribuiscono ai lavoratori specifici doveri di controllo, di verifica critica e di presidio dell'affidabilità dei sistemi di intelligenza

artificiale, con il conseguente ampliamento dell'area delle possibili responsabilità derivanti dal loro mancato o inesatto adempimento.

L'attualità del problema emerge con particolare evidenza dai più recenti strumenti di *soft regulation* adottati nel settore pubblico, che mostrano come la responsabilizzazione degli addetti alla supervisione umana costituisca ormai una traiettoria già avviata. Emblematico è, in tal senso, lo schema di Codice etico e di comportamento⁽¹⁷⁾ allegato alle Linee guida AgID per l'adozione di IA nella pubblica amministrazione⁽¹⁸⁾, nel quale il controllo umano tende progressivamente a perdere la sola dimensione organizzativa per assumere quella di un vero e proprio dovere professionale del dipendente.

Sul presupposto che la qualità dell'IA dipenda non soltanto dalla correttezza del modello algoritmico, ma anche dalla capacità dell'amministrazione di comprenderne il funzionamento, interpretarne criticamente i risultati, individuarne eventuali errori o *bias* e mantenere un controllo effettivo sul processo decisionale, l'art. 2 dello schema di Codice etico chiarisce, infatti, come il Codice non si limiti a disciplinare i principi di utilizzo dell'intelligenza artificiale da parte dell'amministrazione, individuando oltre ai principi guida dell'azione amministrativa, anche i comportamenti dei soggetti che operano presso l'amministrazione stessa; declinando, in altre parole, nell'ambiente digitale, i tradizionali doveri di lealtà, imparzialità, diligenza, correttezza e operosità già posti a carico dei dipendenti pubblici dal Codice di comportamento. Significativo è l'art. 13 dello schema, dedicato all'utilizzo dei sistemi di intelligenza artificiale da parte del personale, in cui si specifica come sia richiesto al personale stesso non soltanto di utilizzare correttamente i sistemi messi a disposizione dall'amministrazione, ma di comprenderne il funzionamento, di essere consapevole dei rischi connessi ai *bias*, alle allucinazioni e ai fenomeni di disinformazione propri dell'IA generativa e, soprattutto, di sottoporre gli *output* del sistema ad un continuo processo di verifica, evitando di recepirli acriticamente nello svolgimento della propria attività lavorativa.

Le successive Linee guida AgID sullo sviluppo dei sistemi di IA⁽¹⁹⁾ confermano e enfatizzano il ruolo della supervisione umana, specificando come il rapporto tra amministrazione e intelligenza artificiale possa assumere intensità profondamente diverse, distinguendo tra operatore base, operatore avanzato, operatore esperto e operatore controllore in funzione del diverso livello di autonomia tecnica, delle competenze richieste e della capacità di incidere sul ciclo di vita del sistema. Tale classificazione, pur concepita con finalità prevalentemente organizzative di distinzione delle amministrazioni utilizzatrici,

(17) Sul ruolo dei codici di condotta come strumento di applicazione volontaria delle norme dell'*AI Act*, anche tenendo conto delle migliori pratiche del settore si rinvia a S. Ciucciiovino, *Risorse umane e intelligenza artificiale alla luce del regolamento (UE) 2024/1689, tra norme legali, etica e codici di condotta*, DRI, 2024, 3, 573 ss.

(18) AgID, *Linee guida per l'adozione di IA nella pubblica amministrazione*, adottate con determinazione del D.G. n. 17 del 17 febbraio 2025, sulle quali è stata avviata la consultazione pubblica.

(19) AgID, *Linee guida per lo sviluppo di sistemi di Intelligenza Artificiale nella Pubblica Amministrazione*, adottate con determinazione del D.G. n. 43 del 10 marzo 2026, sulle quali è stata avviata la consultazione pubblica.

offre un utile criterio anche per una lettura giuslavoristica dello *standard* di diligenza richiesto al personale che in essa opera.

Significativa, ai fini della possibile graduazione della diligenza nella supervisione umana, è la classificazione dei ruoli contenuta nelle medesime Linee guida che individua quattro “famiglie di utilizzatori” – operatore base, operatore avanzato, operatore esperto e operatore controllore – precisando che esse non identificano necessariamente amministrazioni diverse, potendo la medesima amministrazione assumere profili differenti in relazione al caso d’uso concretamente considerato.

La classificazione è costruita in funzione del diverso livello di autonomia e controllo ottenibile, nonché delle competenze e delle risorse disponibili: per l’operatore base non sono richieste competenze specialistiche avanzate né particolari risorse umane, finanziarie o strumentali; per l’operatore avanzato sono invece necessarie competenze specialistiche e risorse dedicate; per l’operatore esperto sono richieste competenze specialistiche diffuse e rilevanti risorse dedicate; per l’operatore controllore, infine, competenze avanzate e consistenti risorse umane, finanziarie e strumentali.

Al di là della specifica esperienza maturata nel settore pubblico, tali indicazioni assumono un rilievo sistematico ben più ampio. Esse testimoniano come la supervisione umana stia progressivamente cessando di rappresentare un mero requisito organizzativo dell’impiego dell’intelligenza artificiale per trasformarsi in un contenuto della prestazione lavorativa suscettibile di autonoma valutazione. La graduazione dei livelli di controllo, delle competenze richieste e dell’intensità della supervisione mostra, infatti, che il comportamento del lavoratore che utilizza l’IA o deputato alla sua supervisione tende ad essere sempre più misurato e misurabile non soltanto in relazione alla corretta esecuzione dell’attività affidata, ma anche alla qualità del giudizio critico esercitato sull’operato del sistema algoritmico.

Proprio questa progressiva responsabilizzazione rende, tuttavia, ancora più evidente la necessità di una rigorosa delimitazione dell’obbligazione esigibile. Se il controllo umano è destinato a divenire parametro di valutazione dell’adempimento e, conseguentemente, possibile fondamento di responsabilità disciplinari, contrattuali o amministrativo-contabili, non è sufficiente affermarne in termini generali l’obbligatorietà. Occorre, piuttosto, definire preventivamente che cosa il lavoratore sia concretamente tenuto a controllare, quale livello di approfondimento possa essergli richiesto, entro quali limiti egli possa fare affidamento sull’*output* algoritmico e quando, invece, sorga il dovere di discostarsene o di attivare ulteriori verifiche.

L’esigenza di delimitare lo *standard* di diligenza assume una rilevanza centrale se si considera, come prima accennato, che l’oggetto stesso della supervisione è costituito da sistemi il cui funzionamento rimane, almeno in parte, opaco e che ormai hanno la capacità di autoaggiornarsi imparando da se stessi in modo dinamico. La complessità dei modelli algoritmici, la difficoltà di ricostruire il percorso inferenziale che conduce alla decisione, la continua evoluzione dei sistemi di apprendimento e la stessa asimmetria informativa tra chi progetta il sistema e chi è chiamato a supervisionarne l’utilizzo rendono particolarmente

problematico individuare il confine tra l'errore imputabile al lavoratore e quello riconducibile alle caratteristiche tecnologiche del sistema o alle scelte organizzative del datore di lavoro.

A ciò si aggiunge un'ulteriore criticità. La responsabilità del supervisore non dipende soltanto dalla capacità di individuare eventuali anomalie del sistema, ma anche dalla possibilità concreta di discostarsi dalla decisione dell'algoritmo o dalla semplice datificazione di alcuni passaggi decisionali. Proprio su questo versante si registra, allo stato, una significativa necessità di attenzione e di analisi dei contesti concreti. Mancano, infatti, criteri preventivamente definiti che consentano di stabilire quando il lavoratore sia tenuto a non conformarsi alla decisione della macchina, quali elementi debbano giustificare tale scelta, come essa debba essere motivata, quali procedure interne debbano essere attivate, quali livelli autorizzativi debbano eventualmente essere coinvolti e quali modalità di tracciabilità debbano accompagnare l'esercizio del dissenso rispetto all'algoritmo.

In assenza di tali coordinate, il controllo umano rischia di oscillare tra due esiti opposti ma ugualmente distorsivi: da un lato, un atteggiamento di sostanziale acquiescenza nei confronti dell'*output* algoritmico, indotto dall'incertezza circa le condizioni che legittimano il discostamento; dall'altro, una responsabilizzazione *ex post* del lavoratore per non avere esercitato un potere di intervento il cui contenuto e le cui modalità di esercizio non erano stati previamente determinati dall'organizzazione.

Paradossalmente, proprio l'imperscrutabilità della *black box*, unita all'assenza di regole organizzative sul dissenso rispetto alla decisione algoritmica, rischia di ampliare l'area della responsabilità del supervisore. Quanto meno è trasparente il processo decisionale della macchina e quanto meno sono definiti i percorsi attraverso i quali il lavoratore può correggerne o disattenderne gli esiti, tanto più il giudizio sull'adempimento tende a risolversi in una valutazione *ex post* della sufficienza/insufficienza del controllo esercitato, con un inevitabile ampliamento delle clausole generali di diligenza e buona fede.

È proprio in questa tensione che emerge la funzione essenziale del diritto del lavoro. La costruzione dello *standard* di diligenza non serve soltanto a definire il contenuto dell'obbligazione, ma anche a preservare il corretto riparto del rischio tra organizzazione e prestatore. Diversamente, il principio della supervisione umana rischierebbe di trasformarsi in una clausola di imputazione generale della responsabilità, consentendo di trasferire sul lavoratore le conseguenze di limiti tecnologici, scelte progettuali o carenze organizzative che rimangono integralmente nella disponibilità del datore di lavoro.

6. La proceduralizzazione della supervisione umana nell'*AI Act* e nella direttiva piattaforme

Per fare fronte a questo rischio, l'art. 14 dell'*AI Act* fornisce un preciso modello procedurale finalizzato a garantire la sorveglianza umana; un modello che si inserisce pienamente nella logica di gestione e di mitigazione dei rischi connessi all'utilizzo

dell'IA⁽²⁰⁾. La disposizione non disciplina, infatti, un singolo atto di controllo, ma costruisce una vera e propria procedura organizzativa della supervisione, articolata lungo una pluralità di momenti tra loro funzionalmente collegati e destinati a rendere il controllo umano concretamente esercitabile.

La prima caratteristica che emerge dalla disciplina europea consiste nella anticipazione della supervisione rispetto all'utilizzo del sistema. Il controllo umano non nasce nel momento in cui il lavoratore riceve l'*output* algoritmico da supervisionare, ma costituisce un requisito che deve essere incorporato sin dalla progettazione e dallo sviluppo del sistema. La supervisione viene così trasformata da attività successiva all'elaborazione algoritmica a componente strutturale, imponendo che i sistemi di IA siano concepiti, anche attraverso adeguate interfacce uomo-macchina, in modo da rendere possibile un controllo effettivo durante l'intero periodo di utilizzo.

A tale scelta si accompagna una seconda direttrice di particolare rilievo. Il Regolamento chiarisce che la supervisione non costituisce una misura uniforme, ma un insieme di presidi organizzativi destinati a graduarsi in funzione dei rischi, del livello di autonomia del sistema e del concreto contesto di utilizzo. La procedimentalizzazione del controllo assume, pertanto, una struttura necessariamente differenziata, nella quale l'intensità della supervisione è destinata ad aumentare proporzionalmente alla capacità del sistema di incidere su salute, sicurezza e diritti fondamentali.

La disposizione non individua, tuttavia, direttamente i doveri del lavoratore incaricato del controllo, bensì costruisce un vero e proprio obbligo organizzativo gravante sul *deployer*, imponendo che il sistema sia fornito e organizzato in modo tale da porre le persone incaricate della supervisione nelle condizioni di esercitare effettivamente il controllo loro affidato.

Il legislatore non si limita poi a prescrivere che il supervisore comprenda il funzionamento del sistema, ne interpreti correttamente gli *output* o sia in grado di intervenire sul processo decisionale, ma impone che egli abbia concretamente la possibilità di svolgere tali attività. L'elencazione contenuta nell'art. 14, par. 4, assume, pertanto, una valenza che trascende la mera descrizione delle attività del supervisore. Essa individua, piuttosto, i contenuti minimi che l'organizzazione è chiamata a trasferire al lavoratore affinché la supervisione divenga concretamente possibile. La capacità di comprendere le caratteristiche, le potenzialità e i limiti del sistema; di monitorarne il funzionamento e individuarne anomalie, disfunzioni o prestazioni inattese; di mantenere costante

⁽²⁰⁾ R. Gellert, *The risk-based approach to data protection*, Oxford University Press, 2020; A. Aloisi - V. De Stefano, *Between risk mitigation and labour rights enforcement: Assessing the transatlantic race to govern AI-driven decision-making through a comparative lens*, ELLJ, 2023, vol. 14, 2, 283 ss.; Barbera M., «La nave deve navigare». *Rischio e responsabilità al tempo dell'impresa digitale*, LLI, 2023, 9, 2, 1 ss.; P. Loi, *Il rischio proporzionato nella proposta di Regolamento sull'IA e i suoi effetti nel rapporto di lavoro*, Federalismi.it, 2023, 4, 238 ss.; A. Mantelero - M. Peruzzi, *L'AI e la gestione del rischio nel sistema integrato delle fonti*, RGL, 2024, n. 4, 517 ss.; A. Ingrao, *Hic sunt leones! La piramide del rischio costruita dalla proposta di regolamento sull'intelligenza artificiale emendata*, LPO, 2023, 11-12, 778 ss.; L. Zappalà, *Sistemi di IA ad alto rischio e ruolo del sindacato alla prova del risk-based approach*, LLI, 2024, vol. 10, 1, 52 ss.

consapevolezza del rischio di *automation bias*; di interpretare criticamente le decisioni prodotte prodotti mediante gli strumenti di spiegabilità disponibili; di decidere, ove necessario, di discostarsi dalla decisione algoritmica ovvero di intervenire direttamente sul funzionamento del sistema costituiscono, prima ancora che obblighi del supervisore, contenuti dell'obbligo organizzativo del *deployer*.

Sotto questo profilo, l'art. 14 sembra presupporre una sorta di trasposizione organizzativa della valutazione di impatto effettuata sul sistema di IA che, dal profilo di individuazione del rischio *ex ante*, plana sulle attività di supervisione. Una valutazione di impatto funzionale a far comprendere a chi deve sorvegliare l'impatto stesso della decisioni e rischi. L'attività di analisi dei rischi, dei possibili effetti pregiudizievoli anche rispetto ai diritti fondamentali, delle misure di mitigazione adottate, dei limiti conosciuti del sistema, dei correttivi predisposti e delle condizioni nelle quali il rischio residuo permane non può infatti arrestarsi alla documentazione tecnica della fase preventiva, in una logica di *compliance*.

In questa prospettiva, tale attività di attivazione formativa/informativa nei confronti di chi deve controllare assume una funzione specifica rispetto alla valutazione di impatto di cui all'art. 27 dell'*AI Act*⁽²¹⁾, ove applicabile. Essa diviene lo strumento attraverso il quale il datore di lavoro riempie di contenuto l'obbligazione lavorativa di supervisione sul piano individuale, traducendo i rischi individuati sul piano tecnologico in contenuti della prestazione, in competenze professionali, in istruzioni operative e in criteri di esercizio del controllo umano. La supervisione cessa così di essere affidata ad una generica capacità critica del lavoratore e viene progressivamente oggettivata attraverso la preventiva esplicitazione dei rischi da presidiare, delle anomalie da ricercare, delle misure di mitigazione già implementate e delle situazioni che richiedono un intervento correttivo.

Emerge, dunque, con chiarezza come la supervisione umana presupponga un preventivo processo di esplicitazione e trasferimento della conoscenza. I risultati dell'analisi dei rischi, della valutazione d'impatto, delle misure di mitigazione adottate, dei limiti conosciuti del sistema, dei possibili *bias* e delle situazioni nelle quali la decisione algoritmica potrebbe risultare inattendibile non possono rimanere confinati nella fase di progettazione o nella documentazione tecnica del sistema, ma devono tradursi in un patrimonio di conoscenze effettivamente accessibile al lavoratore chiamato ad esercitare la supervisione. Solo attraverso tale trasferimento cognitivo la funzione di controllo diviene concretamente esercitabile e, conseguentemente, giuridicamente esigibile.

⁽²¹⁾ L'art. 27, *AI Act* impone infatti al datore di lavoro o committente che si configuri come *deployer* di condurre una valutazione d'impatto sui diritti fondamentali (cd. *Fria*) per i sistemi di intelligenza artificiale (IA) ad alto rischio di cui all'art. 6 *AI Act*, tra cui rientrano, in generale, quelli di *management* algoritmico. Tale obbligo era stato proposto dal Parlamento europeo con un ambito di applicazione soggettivo ad ampio spettro, applicandosi a tutti i *deployer*. Tuttavia, la versione finale dell'*AI Act* lo ha circoscritto ai soli *deployer* che sono organismi di diritto pubblico o enti privati che forniscono servizi pubblici, nonché ai *deployer* di sistemi utilizzati per valutazioni sulle persone fisiche nel settore creditizio e assicurativo. Sul ruolo della *Fria* v. G. Gaudio, *Valutazione di impatto e management algoritmico*, RGL, 2024, 4, 538 ss.

Ne consegue che la formazione richiesta dall'art. 26, *AI Act* ⁽²²⁾ non può essere ridotta ad un generico obbligo di *AI literacy*. Essa costituisce il naturale completamento dell'obbligo di organizzazione delineato dall'art. 14 e assolve ad una funzione ben più specifica: rendere il lavoratore concretamente in grado di esercitare quel controllo che il legislatore europeo considera il presidio ultimo dell'affidabilità dei sistemi di intelligenza artificiale. In questa prospettiva, formazione, competenze, conoscenza dei rischi e capacità di interpretazione non rappresentano misure autonome, ma segmenti di un unico processo organizzativo destinato a rendere effettiva — e quindi anche esigibile — la supervisione umana.

L'impianto delineato dall'art. 14, *AI Act* trova un significativo sviluppo nella direttiva (UE) 2024/2831 sul lavoro mediante piattaforme digitali⁽²³⁾, la quale non si limita a ribadire la necessità della supervisione umana, ma ne amplia sensibilmente la dimensione organizzativa. La direttiva, in particolare, introduce un modello di governo permanente della supervisione ancora più incisivo, nel quale il controllo sull'intelligenza artificiale non costituisce più un momento statico del processo decisionale, ma un'attività continuativa di monitoraggio, verifica e revisione degli effetti prodotti dai sistemi algoritmici.

La direttiva impone, infatti, alle piattaforme di effettuare, con cadenza periodica e con il coinvolgimento dei rappresentanti dei lavoratori, una valutazione dell'impatto delle decisioni algoritmiche sulle condizioni di lavoro e sulla parità di trattamento. La valutazione di impatto cessa così di rappresentare un'attività concentrata nella fase di progettazione o di messa in esercizio del sistema e diviene un processo continuo di verifica degli effetti concretamente prodotti dall'algoritmo nel corso del suo utilizzo. Valutazione di impatto che, come si è detto, prevede il preventivo coinvolgimento a monte delle risorse dedicate alla supervisione che, di fatto, dovranno essere formate sui rischi, sugli impatti, oltre che sulle misure correttive da adottare.

Mentre l'*AI Act* richiede che il *deployer* individui persone dotate delle competenze, della formazione e dell'autorità necessarie, la direttiva piattaforme impone un obbligo ulteriore, e molto più incisivo, prevedendo che siano assicurate risorse umane sufficienti per svolgere efficacemente le attività di supervisione e di valutazione. Il legislatore europeo, dunque, non si limita più a prescrivere l'idoneità del supervisore, ma richiede che l'organizzazione disponga di un assetto strutturale adeguato a garantire, anche in termini quantitativi, la continuità e l'effettività del controllo umano.

Particolarmente significativa è, inoltre, la tutela riconosciuta ai soggetti incaricati della supervisione. La direttiva prevede espressamente che essi non possano essere destinatari di licenziamenti, misure disciplinari o altri trattamenti sfavorevoli conseguenti all'esercizio delle proprie funzioni. Si tratta di una previsione assente nell'*AI Act*, che rivela come il legislatore europeo abbia piena consapevolezza del possibile conflitto tra le

⁽²²⁾ Secondo la quale i *deployer* affidano la sorveglianza umana a persone fisiche che dispongono della competenza, della formazione e dell'autorità necessarie nonché del sostegno necessario.

⁽²³⁾ M. P. Aimò, *Il management algoritmico nel lavoro mediante piattaforme: osservazioni sulle prime regole di portata europea*, *Federalismi.it*, 2024, 25, 216 ss.

esigenze di efficienza perseguite dall'organizzazione e l'esercizio di un controllo umano realmente critico sugli esiti dell'attività algoritmica. L'effettività della supervisione richiede, infatti, non soltanto competenze e autorità, ma anche una posizione di autonomia idonea a consentire al supervisore di discostarsi dalla decisione del sistema senza il timore di conseguenze pregiudizievoli.

Un ulteriore elemento di novità riguarda la procedimentalizzazione del dissenso rispetto alla decisione algoritmica. Se l'*AI Act* riconosce al supervisore il potere di ignorare, annullare o ribaltare l'*output* del sistema, la direttiva costruisce una vera e propria procedura di riesame umano, fondata sul diritto del lavoratore ad ottenere una spiegazione intelligibile della decisione automatizzata, sulla designazione di un referente umano competente con il quale confrontarsi, sull'obbligo di motivazione delle decisioni che incidono sugli aspetti essenziali del rapporto di lavoro e sulla previsione di un procedimento di revisione, rettifica e, ove necessario, di modifica del sistema algoritmico. Il controllo umano viene così inserito all'interno di una procedura formalizzata che disciplina non soltanto il momento della decisione, ma anche quello della sua contestazione e della successiva revisione.

Infine, la direttiva attribuisce alla supervisione una funzione che trascende il controllo del singolo provvedimento. L'accertamento di decisioni discriminatorie o lesive dei diritti del lavoratore non determina soltanto la revisione del caso concreto, ma impone alla piattaforma di modificare il sistema decisionale automatizzato o, se necessario, di cessarne l'utilizzo. La supervisione si basa, pertanto, nella direttiva piattaforma, su una costante procedura di apprendimento organizzativo: gli errori riscontrati nel controllo umano alimentano un processo di continuo aggiornamento della valutazione dei rischi e di progressivo miglioramento del sistema algoritmico.

7. La filiera organizzativa della supervisione umana nella normazione tecnica UNI

Un'ulteriore conferma della necessità di una preventiva costruzione organizzativa della supervisione umana proviene dalla UNI CEI ISO/IEC 42001⁽²⁴⁾, la quale, pur collocandosi sul piano della normazione volontaria, individua un modello di gestione dell'intelligenza artificiale che potenzialmente presenta rilevanti ricadute sul contenuto dell'obbligazione lavorativa del soggetto incaricato della supervisione stessa. La norma, infatti, non si limita a richiedere l'adozione di misure di gestione del rischio, ma costruisce una sequenza procedimentale nella quale ogni fase è funzionale a rendere concretamente esercitabile il controllo umano.

Di particolare interesse è, anzitutto, la disciplina della valutazione del rischio dell'IA. La norma impone all'organizzazione di definire un processo formalizzato e partecipato di

⁽²⁴⁾ UNI CEI ISO/IEC 42001, *Tecnologie informatiche – Intelligenza artificiale – Sistema di gestione*, versione italiana del gennaio 2025.

risk assessment idoneo a produrre risultati coerenti, validi e comparabili, attraverso l'identificazione dei rischi che possono compromettere gli obiettivi perseguiti dal sistema, la valutazione delle conseguenze che tali rischi possono produrre per l'organizzazione, per gli individui e per la società, la determinazione del relativo livello di rischio e la successiva prioritizzazione ai fini del trattamento⁽²⁵⁾. Non si tratta di un'attività meramente ricognitiva, ma di un processo documentato che deve accompagnare l'intero sistema di gestione dell'IA, soprattutto tenendo conto che la stessa può nel tempo cambiare comportamento, diventando inspiegabile⁽²⁶⁾.

Un passaggio fondamentale della norma è la previsione e l'identificazione dei ruoli che all'interno dell'organizzazione possono essere previsti per produrre e gestire l'IA⁽²⁷⁾: la pluralità di figure – con competenze disciplinari diversificate – chiamate valutare, monitorare e supervisionare l'IA suggerisce una riflessione che appare fondamentale, vale a dire che la sorveglianza umana è un'attività complessa che, ove l'impresa decida di introdurre sistemi di IA, non potrà mai essere svolta da una sola persona, richiedendo la sorveglianza stessa una pluralità di competenze interdisciplinari (umanistiche, informatiche, sociali, di valutazione del rischio, ecc.) che necessariamente fanno capo non ad un singolo, ma a una organizzazione più complessa. La centralità dei ruoli passa anche per l'assicurazione che le responsabilità e le autorità relative ai ruoli rilevanti siano assegnate e comunicate all'interno dell'organizzazione (punto 5.3.): il ruolo assegnato deve innanzitutto essere compreso, evidentemente con una chiara individuazione della propria sfera di intervento e delle proprie responsabilità.

L'attività di *risk assessment*, il trattamento del rischio e la valutazione di impatto costituiscono, infatti, il momento nel quale l'organizzazione individua preventivamente

⁽²⁵⁾ Cfr. p. 2 ove si prevede che il sistema di gestione dell'IA dovrebbe essere integrato con i processi dell'organizzazione e con la struttura gestionale complessiva. Nella progettazione di processi, sistemi informativi e controlli si dovrebbero prendere in considerazione problemi specifici correlati all'IA. Esempi fondamentali di tali processi di gestione sono:

- determinazione di obiettivi organizzativi, coinvolgimento delle parti interessate e politica dell'organizzazione;
- gestione dei rischi e delle opportunità;
- processi per la gestione delle problematiche legate all'affidabilità (*trustworthiness*) dei sistemi di intelligenza artificiale, quali sicurezza (*security*), salvaguardia delle persone/incolumità (*safety*), correttezza (*fairness*), trasparenza (*transparency*), qualità dei dati e qualità dei sistemi di IA durante il loro ciclo di vita.

⁽²⁶⁾ La pianificazione del rischio si articola in diverse fasi; in particolare, l'organizzazione deve stabilire e mantenere criteri di rischio dell'IA che supportino:

- la distinzione tra rischi accettabili e rischi non accettabili;
- l'esecuzione (*performing*) di valutazioni del rischio dell'IA (*AI risk assessments*);
- la conduzione (*conducting*) del trattamento del rischio dell'IA (*AI risk treatment*);
- la valutazione (*assessing*) degli impatti del rischio dell'IA (*AI risk impacts*).

⁽²⁷⁾ Si suggerisce di distinguere Produttori di IA (*AI producers*), compresi sviluppatori di IA (*AI developers*), progettisti di IA (*AI designers*), operatori di IA (*AI operators*), tecnici di laboratorio e valutatori di IA (*AI testers and evaluators*), implementatori di IA (*AI deployers*), esperti delle scienze sociali e umane applicate all'AI (*AI human factor professionals*), esperti di dominio (*domain experts*), valutatori dell'impatto dell'IA (*AI impact assessors*), committenti (*procurers*), professionisti della *governance* e della supervisione dell'IA (*AI governance and oversight professionals*), v. 7.

quali rischi il supervisore – come diversamente declinato nei vari ruoli – sarà chiamato a presidiare, quali effetti sui diritti fondamentali sono prevedibili, quali misure di mitigazione sono già state adottate, quali controlli risultano necessari e quali eventi devono attivare l'intervento umano. La procedura di valutazione non esaurisce, pertanto, la propria funzione nella dimostrazione della conformità del sistema, ma diviene il primo atto di conformazione dell'obbligazione lavorativa di supervisione, poiché individua e rende conoscibile l'oggetto stesso del controllo richiesto al lavoratore, dettaglia i tipi di controllo, assumendosi la responsabilità della scelta e del disegno complessivo di gestione del rischio⁽²⁸⁾.

Nella medesima direzione si colloca la disciplina del supporto organizzativo prevista dal punto 7 della norma. L'organizzazione è chiamata non soltanto a mettere a disposizione risorse adeguate, ma anche ad individuare preventivamente le competenze⁽²⁹⁾ necessarie per i soggetti le cui attività incidono sul sistema di IA, a verificarne il possesso, a colmare eventuali carenze formative, nonché a garantire la consapevolezza del personale circa il proprio ruolo nel sistema di gestione dell'IA⁽³⁰⁾ e circa le conseguenze derivanti dal mancato rispetto delle procedure predisposte, nonché a strutturare adeguati sistemi di comunicazione e di gestione delle informazioni documentate.

Particolarmente significativa è proprio la disciplina delle informazioni documentate. La norma richiede che criteri di valutazione del rischio, procedure operative, risultati delle valutazioni di impatto, controlli adottati e misure di trattamento del rischio siano formalizzati, aggiornati, controllati e resi disponibili all'interno dell'organizzazione. Anche tale previsione trascende la dimensione della mera *compliance* documentale. Essa costituisce il presupposto affinché il lavoratore incaricato della supervisione possa conoscere i limiti

⁽²⁸⁾ Cfr. p. 6.2., ove si prevede che nel pianificare come raggiungere i propri obiettivi di IA, l'organizzazione deve determinare:

- cosa viene fatto;
- quali risorse vengono richieste;
- chi ne è responsabile;
- quando viene completato;
- come sono valutati i risultati.

⁽²⁹⁾ Cfr. p. 7.2. dedicato alla competenze, ove si prevede che l'organizzazione deve:

- stabilire le competenze necessarie per la(e) persona(e) che svolge/svolgono attività lavorative sotto il suo controllo e che influenzano le sue prestazioni di IA;
- assicurare che queste persone siano competenti sulla base di istruzione, formazione, o esperienza appropriate;
- ove applicabile, intraprendere azioni per acquisire le necessarie competenze e valutare l'efficacia delle azioni intraprese.

Devono essere disponibili informazioni documentate appropriate quale evidenza delle competenze.

⁽³⁰⁾ Cfr. p. 7.3. ove si prevede che le persone che svolgono attività lavorative sotto il controllo dell'organizzazione devono essere consapevoli:

- della politica di IA;
- del proprio contributo all'efficacia del sistema di gestione dell'IA, compresi i benefici derivanti dal miglioramento delle prestazioni di IA;
- delle implicazioni derivanti dalla non conformità ai requisiti del sistema di gestione dell'IA.

del sistema, i rischi già individuati dall'organizzazione, gli indicatori di anomalia, le soglie di attenzione e le modalità di intervento che gli vengono richieste.

Fondamentale, in tale contesto, è poi il ruolo degli *audit* interni, il riesame interno di direzione, monitoraggio e miglioramento.

La norma tecnica mostra così con particolare evidenza come il contenuto della supervisione umana venga progressivamente costruito attraverso una preventiva attività di specificazione organizzativa nella quale il datore di lavoro traduce la valutazione del rischio in competenze, procedure, informazioni documentate, criteri di intervento e ruoli specifici e differenziati per taglio del potenziale intervento.

In altre parole, la UNI CEI ISO/IEC 42001 costruisce una vera e propria filiera organizzativa della supervisione umana, nella quale il controllo esercitato dal lavoratore rappresenta soltanto il momento conclusivo di un processo molto più ampio. La supervisione umana individuale non costituisce il punto di partenza del sistema, bensì il suo possibile approdo finale, fermo restando che nella organizzazioni di grandi dimensioni la sorveglianza sarà necessariamente attribuita a *equipe* multidisciplinari.

Osservata da questa angolazione, la supervisione umana cessa di identificarsi con il comportamento del singolo lavoratore e diviene il risultato di un'attività organizzativa in continua evoluzione. L'obbligazione lavorativa del supervisore è progressivamente conformata dalla filiera organizzativa costruita dal datore di lavoro. È quest'ultimo che, attraverso la valutazione del rischio, la valutazione di impatto e la definizione delle misure di trattamento, individua preventivamente quali siano le anomalie da intercettare, i rischi da presidiare, le possibili interferenze con i diritti fondamentali, i correttivi disponibili, le soglie di intervento e le modalità attraverso le quali il controllo umano dovrà essere esercitato.

La procedura assume così una funzione ben diversa da quella meramente certificativa o documentale. Essa rappresenta il meccanismo attraverso il quale il patrimonio conoscitivo elaborato nella fase di gestione del rischio viene progressivamente trasferito al lavoratore o al gruppo di lavoratori deputati alla supervisione, trasformandosi nel contenuto concreto delle eventuali obbligazione individuali. In altri termini, è proprio la filiera organizzativa a riempire di contenuto l'obbligazione lavorativa, rendendo determinabile ciò che, in assenza di tale processo, resterebbe affidato ad un generico dovere di vigilanza o ad un indeterminato obbligo di pensiero critico.

Questa ricostruzione consente, infine, di cogliere il significato sistematico della normazione tecnica. La UNI non si limita a suggerire buone pratiche organizzative, ma offre il paradigma attraverso il quale il principio della supervisione umana può essere tradotto in un'obbligazione lavorativa concretamente esigibile. Ed è proprio perché la prestazione del supervisore è il punto terminale di questa filiera organizzativa che la verifica del suo eventuale inadempimento non potrà mai prescindere dall'accertamento della correttezza e della completezza delle attività organizzative che la precedono. È qui che si colloca il tema, affrontato nel paragrafo successivo, del nuovo riparto delle responsabilità tra organizzazione e lavoratore.

8. Sorveglianza umana e organizzazione del lavoro: verso un nuovo riparto delle responsabilità

L'analisi congiunta dell'*AI Act*, della direttiva sulle piattaforme e della più recente normazione tecnica consente di cogliere un dato sistematico che trascende il contenuto delle singole disposizioni. L'obbligazione di supervisione non può essere isolata dal contesto organizzativo nel quale essa prende forma⁽³¹⁾. L'intero impianto dell'*AI Act* mostra come il controllo umano costituisca soltanto l'ultimo segmento di una più ampia architettura organizzativa, costruita attraverso obblighi di progettazione, gestione del rischio, documentazione, formazione, monitoraggio e definizione delle procedure di intervento. La supervisione non nasce, dunque, come una capacità individuale del lavoratore, ma come il risultato di una scelta organizzativa imputabile al *deployer*.

In questa prospettiva, il tema della responsabilità del supervisore si intreccia inevitabilmente con quello, più generale, della responsabilità dell'impresa nella predisposizione di assetti organizzativi adeguati.

Se, come è stato osservato in dottrina con riferimento all'art. 2086 c.c., l'adeguatezza organizzativa costituisce oggi un autonomo *standard* giuridico destinato a conformare l'esercizio dei poteri imprenditoriali⁽³²⁾, lo stesso paradigma sembra destinato a proiettarsi nella *governance* dei sistemi di intelligenza artificiale.

Soltanto un'impresa che abbia assolto agli obblighi di progettazione, valutazione dei rischi, formazione, informazione, definizione delle procedure decisionali e individuazione dei limiti dell'intervento umano può dirsi aver trasformato il generico obbligo di supervisione in una prestazione concretamente esigibile.

Solo a questo punto diviene possibile individuare il frammento di prestazione effettivamente rimesso al singolo lavoratore e, conseguentemente, ricostruire lo *standard* di diligenza richiesto nell'esercizio della supervisione. Tale *standard*, pertanto, non può essere desunto in via astratta dall'art. 2104 c.c., ma prende forma attraverso l'organizzazione predisposta dall'impresa, che delimita ciò che il supervisore conosce, può conoscere, è posto nelle condizioni di verificare e, soprattutto, è concretamente in grado di correggere.

La conseguenza sistematica è rilevante. In assenza di assetti organizzativi adeguati, di procedure affidabili, di formazione effettiva, di informazioni intelligibili e di strumenti tecnologicamente idonei, difetta il presupposto stesso per l'individuazione della condotta diligente esigibile dal lavoratore.

La proceduralizzazione della supervisione umana non risponde allora soltanto all'esigenza di rendere più efficace il controllo sull'intelligenza artificiale, ma svolge una funzione essenziale di delimitazione delle responsabilità. Essa oggettiva il frammento individuale dell'obbligazione di supervisione, inserendolo all'interno della più ampia

⁽³¹⁾ Per riflessioni simili v. pure M. Peruzzi, *Intelligenza artificiale e lavoro*, cit., 126 ss.

⁽³²⁾ P. Tullini, *Assetti organizzativi dell'impresa e poteri datoriali. La riforma dell'art. 2086 c.c.: prima lettura*, RIDL, 2020, 135 ss.; M. Peruzzi, *Intelligenza artificiale e lavoro*, cit., 126 ss.

architettura organizzativa predisposta dal datore di lavoro e definendo preventivamente ciò che il singolo supervisore è concretamente tenuto a conoscere, verificare, valutare e, se del caso, correggere.

Ne consegue che la responsabilità del supervisore non può costituire il punto di partenza dell'accertamento, ma il suo approdo finale. Solo ove il datore di lavoro dimostri di avere predisposto tutti gli assetti organizzativi e tutte le misure tecnicamente e organizzativamente esigibili per rendere la supervisione effettiva – assicurando il massimo livello di sicurezza tecnologicamente ragionevole nella gestione del sistema – diviene possibile individuare il frammento di prestazione imputabile al singolo lavoratore e valutare se quest'ultimo abbia effettivamente violato lo *standard* di diligenza richiesto dalla natura della specifica prestazione. Fino a quel momento, la supervisione umana rimane, prima ancora che un dovere del lavoratore, una responsabilità dell'organizzazione.

Diversamente opinando, il principio antropocentrico rischierebbe di essere paradossalmente rovesciato nel suo contrario: da garanzia di governo umano dell'intelligenza artificiale a strumento di deresponsabilizzazione dell'impresa mediante la traslazione sul supervisore dei rischi derivanti da un'organizzazione inadeguata dell'uso dell'IA.

Bibliografia

- Aimo M.P., *Il management algoritmico nel lavoro mediante piattaforma: osservazioni sulle prime regole di portata europea*, in *Federalismi.it*, 2024, 25, 216 ss.
- Alaimo A., *Lavoro e piattaforme tra subordinazione e autonomia: la modulazione delle tutele nella proposta della Commissione europea*, in *DRI*, 2022, 2, 639 ss.
- Aloisi A. - De Stefano V., *Between Risk Mitigation and Labour Rights Enforcement: Assessing the Transatlantic Race to Govern AI-Driven Decision-Making through a Comparative Lens*, in *ELLJ*, 2023, vol. 14, 2, 283 ss.
- Barbera M., «*La nave deve navigare*». *Rischio e responsabilità al tempo dell'impresa digitale*, in *LLI*, 2023, vol. 9, 2, 1 ss.
- Biasi M. (a cura di), *Diritto del lavoro e intelligenza artificiale*, Giuffrè Francis Lefebvre, 2025.
- Borelli S. - Brino V. - Faleri C. - Lazzeroni L. - Tebano L. - Zappalà L., *Lavoro e tecnologie. Dizionario del diritto del lavoro che cambia*, Giappichelli, 2022.
- Brino V., *La tutela della persona che lavora nell'era dell'IA tra sfide etiche e giuridiche*, in *DLM*, 2024, n. 3, 431 ss.
- Campanella P., *Clausole generali e obblighi del prestatore di lavoro*, in *DLRI*, 2015, 89 ss.
- Carinci M.T. - Ingrao A., *L'impatto dell'AI Act sul diritto del lavoro*, in *DLRI*, 2024, 451 ss.
- Cester C., *La diligenza del lavoratore. Art. 2104*, in C. Cester - M.G. Mattarolo, *Diligenza e obbedienza del prestatore di lavoro*, in F.D. Busnelli (diretto da), *Il Codice civile. Commentario*, Giuffrè, 2007.
- Ciucciovino S., *Risorse umane e intelligenza artificiale alla luce del regolamento (UE) 2024/1689, tra norme legali, etica e codici di condotta*, in *DRI*, 2024, n. 3, 573 ss.
- Epifani S., *Il teatro delle macchine pensanti*, Digital Trasformation Institute, 2025.
- Faleri C., *Brevi spunti di riflessione sull'evoluzione delle relazioni sindacali nell'economia digitale*, in *Ianus DF*, 2021, 24, 93 ss.
- Floridi L., *La differenza fondamentale. Artificial Agency: una nuova filosofia dell'intelligenza artificiale*, Mondadori, 2025.
- Gallone G., *Digitalizzazione, amministrazione e persona: per una «riserva di umanità» tra spunti codicistici di teoria giuridica dell'automazione*, in *Pubblica Amministrazione, Persona e Amministrazione*, 2023, 1, 329 ss.
- Gallone G., *Riserva di umanità e funzioni amministrative*, Wolters Kluwer, 2023.
- Gaudio G., *Valutazione di impatto e management algoritmico*, in *RGL*, 2024, 4, 538 ss.
- Gellert R., *The Risk-Based Approach to Data Protection*, Oxford University Press, Oxford, 2020.
- Ingrao A., *Hic sunt leones! La piramide del rischio costruita dalla proposta di regolamento sull'intelligenza artificiale emendata*, in *LPO*, 2023, 11-12, 778 ss.
- Loi P., *Il rischio proporzionato nella proposta di Regolamento sull'IA e i suoi effetti nel rapporto di lavoro*, in *Federalismi.it*, 2023, 4, 238 ss.
- Mantelero A. - Peruzzi M., *L'AI e la gestione del rischio nel sistema integrato delle fonti*, in *RGL*, 2024, 4, 517 ss.
- Montuschi L., *L'applicazione giurisprudenziale del principio di correttezza e di buona fede nel rapporto di lavoro*, in *LD*, 1996, 139 ss.
- Napoli M., *La stabilità reale del rapporto di lavoro*, Franco Angeli, 1980.
- Novella M., Tullini P., *Lavoro digitale*, Giappichelli, 2026.
- Perulli A., *La buona fede nel diritto del lavoro*, in *RGL*, 2002, 3 ss.
- Peruzzi M., *Il dialogo sociale di fronte alle sfide della digitalizzazione*, in *DRI*, 2020, 1213 ss.
- Peruzzi M., *Intelligenza artificiale e lavoro*, Giappichelli, 2024.
- Persiani M., *Contratto di lavoro e organizzazione*, Cedam, 1966.
- Pollicino O. - Domani F. - Finocchiaro G. - Paolucci F. (a cura di), *La disciplina dell'intelligenza artificiale*, Giuffrè Francis Lefebvre, 2025.
- Rota A., *Sull'Accordo quadro europeo in tema di digitalizzazione del lavoro*, in *LLI*, 2020, 2, 25 ss.
- Tebano L., *Intelligenza artificiale datore di lavoro: scenari e regole*, in *DLM*, 2024, 3, 449 ss.
- Treu T., *Il controllo umano delle tecnologie: regole e procedure*, in *WP CSDLE «Massimo D'Antona».IT*, 2025, n. 492.

- Treu T., *La digitalizzazione del lavoro: proposte europee e piste di ricerca*, in *Federalismi.it*, 2022, 190 ss.
- Tullini P., *Assetti organizzativi dell'impresa e poteri datoriali. La riforma dell'art. 2086 c.c.: prima lettura*, in *RIDL*, 2020, 135 ss.
- Tullini P., *Clausole generali e rapporto di lavoro*, Maggioli, 1990.
- Tullini P., *L'economia digitale alla prova dell'interesse collettivo*, in *LLI*, 2018, 1, 4 ss.
- Tullini P., *La Direttiva piattaforme e i diritti del lavoro digitale*, in *LLI*, 2022, 1, 43 ss.
- Vardaro G., *Tecnica, tecnologia e ideologia della tecnica nel diritto del lavoro*, in *Politica del diritto*, 1986.
- Viscomi A., *Diligenza e prestazione di lavoro*, Giappichelli, 1997.
- Viscomi A., *L'adempimento dell'obbligazione di lavoro tra criteri lavoristici e principi civilistici*, in *DLRI*, 2010, 595 ss.
- Zappalà L., *La colpa del lavoratore*, Giappichelli, 2020.
- Zappalà L., *Sistemi di IA ad alto rischio e ruolo del sindacato alla prova del risk-based approach*, in *LLI*, 2024, vol. 10, 1, 52 ss.