



Numero 2 / 2026

Tiziana ORRÙ

**Intelligenza artificiale, dignità della persona e
responsabilità umana nella governance tecnologica**

Intelligenza artificiale, dignità della persona e responsabilità umana nella governance tecnologica

Tiziana ORRÙ

S. Procuratore generale presso la Corte di cassazione.

Introduzione

L'affermazione dell'intelligenza artificiale quale infrastruttura dei processi economici, sociali e decisionali impone una riflessione non riducibile alla dimensione tecnica, poiché i sistemi algoritmici, attraverso previsioni, raccomandazioni e decisioni, incidono sull'accesso a beni, servizi e opportunità, trasformando il rapporto tra potere, conoscenza e responsabilità e prospettando il rischio di una riduzione della persona a dato o funzione calcolabile. In tale contesto, il contributo adotta un metodo integrato che, muovendo dalla natura dell'IA quale ambiente decisionale in cui norme, architetture tecniche e visioni dell'umano si intrecciano, sviluppa l'analisi su tre livelli tra loro complementari — normativo, rappresentato dal Regolamento (UE) 2024/1689; operativo, espresso dal NIST AI Risk Management Framework; ed etico-antropologico, richiamato dalla enciclica Magnifica Humanitas — evidenziando come il primo vincoli, il secondo traduca in procedure e il terzo orienti e legittimi, così da mostrare che il governo dell'IA non possa esaurirsi in un apparato di regole, ma richieda una cultura dell'imputazione consapevole capace di integrare legalità, affidabilità e dignità della persona, preservando il primato del giudizio umano e della responsabilità morale quale criterio ultimo dell'innovazione.

1. Il paradigma antropocentrico dell'intelligenza artificiale

La centralità della persona costituisce il principio assiologico fondamentale nel dibattito etico-giuridico sull'intelligenza artificiale. La prospettiva *human centric* non esprime una semplice formula di indirizzo, ma delinea un criterio di legittimazione dell'innovazione tecnologica: i sistemi di IA devono essere progettati e utilizzati come strumenti al servizio della persona, rispettosi della dignità umana, dell'autonomia individuale e dei diritti fondamentali.

In tale prospettiva, l'antropocentrismo richiede di interrogarsi non solo sulla funzionalità tecnica degli algoritmi, ma sulle condizioni della loro integrazione nei processi decisionali, poiché sistemi efficienti sul piano computazionale possono risultare inadeguati quando producano opacità, asimmetrie informative o trasferimenti occulti dell'imputazione decisionale.

La prospettiva *human centric* richiede che il controllo umano sia previsto già nella fase di progettazione. Occorre predisporre sistemi comprensibili, documentabili e verificabili, che consentano agli operatori di cogliere i limiti dell'output algoritmico e di intervenire quando l'automazione produca effetti sproporzionati o ingiustificati. La supervisione non coincide con la presenza formale di un soggetto umano, ma con una reale capacità di valutazione, correzione e decisione responsabile.

Tale esigenza assume particolare rilievo nei sistemi ad alto rischio, nei quali la sorveglianza umana deve costituire un presidio reale e non solo procedurale, idoneo a garantire la comprensione degli output e la loro eventuale revisione quando incidano su diritti e condizioni di vita. Il rischio di *automation bias* impone, in questa prospettiva, adeguata formazione e assetti organizzativi che rendano il controllo umano effettivo e consapevole.

Nel contesto lavoristico, tale esigenza assume particolare rilievo. I sistemi impiegati per il reclutamento, la gestione del personale, la valutazione delle prestazioni, l'assegnazione di mansioni o la profilazione dei lavoratori incidono su ambiti nei quali l'automazione può generare effetti discriminatori, opacità decisionale e nuove forme di asimmetria organizzativa.

La dimensione lavoristica mostra con particolare evidenza la tensione tra razionalizzazione organizzativa e tutela della persona. Gli strumenti algoritmici promettono efficienza, uniformità valutativa e capacità predittiva, ma possono anche cristallizzare disuguaglianze pregresse, rendere meno trasparente il potere datoriale e trasformare la valutazione del lavoratore in una sequenza di indicatori quantitativi. La protezione della dignità richiede allora che il dato non sostituisca la persona e che la misurazione non divenga l'unico criterio di riconoscimento professionale.

Da ciò discende la necessità di ripensare la governance organizzativa. Accanto alla figura del supervisore umano, l'AI Act valorizza funzioni specialistiche dedicate alla conformità, alla gestione della qualità, alla valutazione etica e alla prevenzione dell'eccessivo affidamento sugli output algoritmici. L'AI Compliance Officer, il responsabile del sistema di gestione della qualità e il comitato etico assumono, in tale prospettiva, un ruolo complementare: non sostituiscono il dovere dell'organizzazione, ma lo rendono più strutturato, documentabile e consapevole.

In questa prospettiva, la governance antropocentrica non è affidata a un singolo presidio, ma a un insieme coordinato di compiti. Il supervisore umano presidia l'uso concreto del sistema; la funzione di compliance verifica la conformità agli obblighi normativi; il responsabile della qualità assicura procedure di controllo, tracciabilità e aggiornamento; il comitato etico consente una valutazione più ampia degli impatti sociali e organizzativi. La loro efficacia dipende tuttavia dalla capacità dell'ente di integrare tali funzioni in un processo decisionale unitario, evitando che esse restino adempimenti separati e meramente formali.

Il paradigma antropocentrico assume così una duplice funzione: da un lato limita l'automazione, impedendo che essa si sostituisca integralmente al giudizio umano; dall'altro orienta l'innovazione, richiedendo che i sistemi siano progettati per accrescere le capacità delle persone e non per ridurle a oggetto di calcolo. La questione non è dunque se l'IA debba essere utilizzata, ma a quali condizioni il suo impiego possa dirsi compatibile con la dignità, la libertà e la responsabilità che fondano l'ordine giuridico democratico.

2. Governance dell'intelligenza artificiale tra AI Act, NIST e Magnifica Humanitas

La governance dell'intelligenza artificiale si configura oggi come sistema multilivello. L'AI Act definisce obblighi, divieti, classificazioni di rischio e presidi giuridicamente vincolanti; il NIST AI Risk Management Framework fornisce un modello flessibile e volontario per mappare, misurare, governare e gestire i rischi lungo l'intero ciclo di vita dei sistemi; Magnifica Humanitas introduce un criterio etico-antropologico che interroga le finalità dell'innovazione e gli assetti di potere che essa produce.

La distinzione tra questi piani è essenziale. Il diritto stabilisce soglie e responsabilità; la tecnica offre strumenti di implementazione e controllo; l'etica domanda se l'uso della tecnologia sia orientato al bene della persona e della comunità. Confondere questi livelli impoverirebbe l'analisi: l'enciclica non è una fonte di regole, così come il NIST non ha norme vincolanti; tuttavia, entrambi contribuiscono a rendere più concreta e controllabile l'attuazione dei principi che il diritto positivo traduce in obblighi.

Il valore di Magnifica Humanitas risiede soprattutto nella capacità di ricondurre la questione tecnologica alla dimensione del potere. L'innovazione digitale non è guidata soltanto dagli Stati, ma anche da soggetti privati transnazionali dotati di risorse, dati e infrastrutture capaci di incidere sulle condizioni reali della vita sociale. In tale scenario, la regolazione giuridica deve confrontarsi con poteri opachi, diffusi e difficilmente riconducibili ai tradizionali schemi di imputazione pubblica.

Per questo l'argomento morale non sostituisce quello giuridico, ma lo integra. Esso ricorda che la conformità formale non esaurisce il dovere di controllo: un sistema può rispettare requisiti procedurali e tuttavia produrre effetti socialmente regressivi, rafforzare disuguaglianze, comprimere l'autonomia personale o favorire forme di dipendenza tecnologica. Il governo dell'IA richiede quindi una valutazione che tenga insieme legalità, affidabilità tecnica e legittimazione etica.

In termini organizzativi, ciò implica che il governo dell'IA non possa essere ridotto alla mera adozione di policy interne o alla predisposizione di modelli astratti di conformità. Esso richiede, piuttosto, un sistema documentabile di decisioni, controlli e doveri, capace di collegare la fase di progettazione con quella di impiego e monitoraggio. La valutazione del rischio deve essere aggiornata nel tempo, poiché l'impatto di un sistema algoritmico può mutare in relazione ai dati utilizzati, al contesto applicativo, agli utenti coinvolti e agli effetti sociali prodotti. Solo un modello dinamico consente di evitare che l'adempimento formale si traduca in deresponsabilizzazione sostanziale.

Da questo punto di vista, il raccordo tra AI Act, NIST e Magnifica Humanitas permette di configurare una responsabilità non solo reattiva, ma preventiva e riflessiva. La regola giuridica individua il perimetro dell'obbligo; il modello di rischio ne favorisce la traduzione in procedure verificabili; la prospettiva etico-antropologica impedisce che tale gestione sia assorbita da una logica puramente efficientistica. La qualità del governo tecnologico si misura, dunque, nella capacità di rendere l'innovazione trasparente, contestabile e orientata alla persona.

Questa impostazione multilivello trova riscontro anche nel recente quadro nazionale. L'evoluzione dell'impianto normativo italiano, con la legge n. 132/2025 e con gli schemi di decreto legislativo esaminati dal Consiglio dei ministri il 10 giugno 2026, ribadisce l'impostazione umanocentrica del governo dei sistemi di IA. Il divieto di decisioni integralmente automatizzate in ambiti sensibili conferma che il presidio umano non costituisce un elemento accessorio, ma una condizione sostanziale di legittimità della decisione automatizzata.

3. Il rischio antropologico dell'intelligenza artificiale

Il rischio più profondo dell'intelligenza artificiale non consiste soltanto nell'errore, nel *bias* o nella discriminazione, ma nella progressiva riconfigurazione dell'umano secondo le categorie del calcolo. L'individuo viene descritto, previsto e valutato attraverso dati, correlazioni e modelli

probabilistici; il giudizio, che implica esperienza, prudenza e capacità di riconoscere l'irripetibilità dei casi concreti, rischia di essere sostituito da procedure automatizzate orientate alla performance.

La specificità di tale rischio consiste nel fatto che esso non opera soltanto sul piano esterno dell'organizzazione sociale, ma incide sulla rappresentazione della soggettività. Quando l'essere umano è descritto attraverso dati, indicatori e previsioni, tende a essere percepito come insieme di variabili misurabili, più che come soggetto capace di libertà, relazione e assunzione morale. L'automazione decisionale può così produrre una riduzione silenziosa dell'umano, mediante la sua traduzione in parametri funzionali.

Questa trasformazione produce un effetto più sottile della semplice delega tecnica: induce a considerare il criterio algoritmico come più neutrale, oggettivo ed efficiente del giudizio umano. Ma l'algoritmo non è neutro. Esso incorpora dati storici, scelte progettuali, criteri di ottimizzazione, interessi organizzativi e assunzioni implicite. La sua apparente oggettività può rendere più difficile individuare la riconducibilità della decisione, contestarne l'esito o comprenderne le ragioni effettive.

In questa prospettiva, la spiegabilità non riguarda soltanto la possibilità tecnica di ricostruire il funzionamento di un modello, ma anche la possibilità giuridica e morale di riferire una decisione a un soggetto responsabile. Se l'esito algoritmico è assunto come dato oggettivo e inevitabile, il circuito della responsabilità decisionale si indebolisce: chi decide può invocare il sistema, mentre chi subisce la decisione può incontrare difficoltà nel comprenderla, contestarla o chiederne la revisione.

È qui che emerge il rischio antropologico: l'uomo non viene soltanto assistito dalla tecnica, ma reinterpretato attraverso di essa. Fragilità, incertezza e libertà vengono percepite come difetti da superare, anziché come dimensioni costitutive della condizione umana. L'ideale tecnocratico tende così a sostituire la domanda sul bene con la domanda sull'efficienza, trasformando la tecnica da strumento in criterio normativo.

Il paradigma tecnocratico si afferma precisamente quando l'efficienza del mezzo tende a sostituire la riflessione sul fine. L'algoritmo, in tal caso, non è più soltanto uno strumento al servizio di decisioni umane, ma diviene criterio ordinatore delle pratiche sociali: seleziona, classifica, anticipa, valuta e orienta. Il rischio non è l'esistenza della tecnica, ma la sua elevazione a misura esclusiva del valore, della razionalità e persino della giustizia.

Da qui deriva l'importanza di un criterio umano di contenimento. Esso non coincide con un atteggiamento antitecnologico, né con la rinuncia all'innovazione; rappresenta piuttosto la condizione perché il progresso non si rovesci in dominio. Riconoscerlo significa ammettere che non tutto ciò che è tecnicamente possibile è anche giuridicamente legittimo o antropologicamente desiderabile, e che la velocità dell'automazione deve essere valutata alla luce della dignità umana.

La riflessione agostiniana evocata da Magnifica Humanitas consente di leggere tale dinamica in profondità. La torre di Babele diviene immagine di una civiltà che cerca sicurezza nel dominio tecnico e nell'autosufficienza, mentre Gerusalemme ricostruita da Neemia rappresenta una comunità fondata su ascolto, corresponsabilità e riconoscimento della misura. La scelta non è dunque tra progresso e rifiuto del progresso, ma tra una tecnica orientata alla comunione e una tecnica convertita in potere.

La metafora di Babele, letta in tale prospettiva, consente di comprendere il rischio di una razionalità tecnica che aspira all'autosufficienza e all'uniformità. Essa non descrive il fallimento del progresso in quanto tale, ma quello di un progresso privo di orientamento, incapace di riconoscere la pluralità dei soggetti e la dimensione relazionale della convivenza. L'alternativa non è tra tecnologia e umanesimo, ma tra un'innovazione guidata dalla responsabilità e una lasciata alla sola logica della potenza.

4. Immaginario, potere e responsabilità: Tolkien come chiave ermeneutica del potere tecnologico

Anche il riferimento a Tolkien contenuto nell'enciclica papale permette di rendere visibile, attraverso l'immaginario, la dinamica morale del potere tecnologico. L'Unico Anello rappresenta uno strumento che promette forza, semplificazione e controllo, ma che progressivamente consuma chi lo possiede e ne altera la capacità di giudizio. In modo analogo, l'intelligenza artificiale non regolata o assunta come mezzo di dominio può trasformarsi da ausilio della decisione in dispositivo di orientamento delle scelte.

Il parallelismo non intende demonizzare la tecnologia, ma mostrare il rischio della fascinazione del potere. Come l'Anello seduce promettendo dominio senza fatica, così l'automazione può sedurre promettendo esiti rapidi, impersonali e apparentemente infallibili. Il pericolo nasce quando tale promessa induce a rinunciare al giudizio, all'imputazione personale e alla fatica del discernimento.

Perfino la figura di Saruman e la trasformazione di Isengard in apparato produttivo e distruttivo assumono valore paradigmatico. La tecnica separata dalla sapienza non resta neutrale: tende a organizzarsi come dominio, sfruttamento e devastazione del creato. La critica tolkieniana della meccanizzazione cieca si salda, così, con la critica contemporanea alla concentrazione dei dati, delle infrastrutture digitali e della capacità di calcolo in pochi centri di potere.

In questa prospettiva, la parola di Gandalf — il richiamo a compiere il bene possibile nel tempo dato, senza pretendere di dominare tutte le maree del mondo — assume il valore di un principio di misura. Governare la tecnologia non significa rifiutarla, ma collocarla entro un orizzonte di custodia, discernimento e responsabilità. Il futuro non è già scritto dagli algoritmi: resta affidato alla libertà morale dell'uomo e alla sua capacità di non delegare integralmente il discernimento all'efficienza automatizzata.

Conclusioni

La questione dell'intelligenza artificiale non si esaurisce nella conformità dei sistemi né nella mitigazione dei rischi tecnici: investe il modo in cui diritto, istituzioni e organizzazioni intendono preservare la dignità umana nell'ambiente tecnologico. L'AI Act, il NIST AI Risk Management Framework e Magnifica Humanitas, pur operando su piani diversi, convergono su un punto decisivo: l'automazione non può diventare uno spazio sottratto al giudizio umano e alla riconducibilità della decisione a un soggetto responsabile.

Il pericolo più profondo non è l'errore algoritmico in sé, ma l'abitudine a considerare l'umano come dato, funzione o prestazione calcolabile. È qui che la tecnica cessa di essere semplice strumento e diviene criterio di ordinamento sociale. Per questo la misura non è un freno al progresso, ma la condizione che consente al progresso di restare umano e di non convertirsi in dominio.

La regolazione dell'IA è dunque una scelta culturale prima ancora che tecnica: impedire che l'efficienza computazionale attenui il dovere personale e collettivo di decidere, rispondere e discernere. Una governance autenticamente antropocentrica non si limita a controllare gli algoritmi; riafferma che nessun sistema automatizzato può sostituire la responsabilità morale del soggetto umano.

Governare l'intelligenza artificiale significa, in definitiva, scegliere quale forma di convivenza si intende costruire: una società governata dall'efficienza algoritmica o una comunità capace di subordinare la potenza tecnica alla libertà, alla dignità della persona e al bene comune. È su questa alternativa che si misura la responsabilità giuridica e morale del nostro tempo.