NEURAL NEXUS Review
Rivista scientifica ufficiale di **ENIA**
Ente Nazionale per l'Intelligenza Artificiale
ISSN 3103-604X
https://neuralnexus-review.org/

I position paper della Fondazione ENIA

Data di ricezione: 11.8.25 / Data accettazione: 11.30.25 / Data di pubblicazione: 31.12.25
doi: 10.82015/NNR.2025.100117

Un confronto tra standard ISO dedicati all'Etica dell'intelligenza artificiale

A comparison of ISO standards dedicated to the ethics of artificial intelligence

Stefano Gorla, Valeria Lazzaroli*

1. Introduzione

L'Intelligenza Artificiale (IA) rappresenta uno dei vettori di trasformazione più profondi e pervasivi del XXI secolo, non soltanto per la sua capacità di automatizzare processi o incrementare l'efficienza operativa, ma per il modo in cui riconfigura la natura stessa delle decisioni all'interno dei sistemi sociali, economici e istituzionali.

A differenza delle precedenti ondate tecnologiche, l'IA non si limita a fornire strumenti di supporto all'azione umana, potendo interviene direttamente nei meccanismi di selezione, classificazione, previsione e allocazione delle risorse, incidendo in modo strutturale su diritti, responsabilità e rapporti di potere.

* Stefano Gorla, co-direttore del Dipartimento AI ETHICS di Fondazione Ente Nazionale per l'Intelligenza Artificiale (ENIA) / E-mail: stefano.gorla@acmcert.net; Valeria Lazzaroli, Presidente di Fondazione Ente Nazionale per l'Intelligenza Artificiale (ENIA) e Lead Auditor ISO IEC 42001 / E-mail: presidenza@enia.ai.

In ambiti ad alta criticità come la sanità, la finanza, la pubblica amministrazione, le risorse umane, la sicurezza e l'industria, i sistemi di IA operano ormai come infrastrutture decisionali che orientano scelte con effetti reali e talvolta irreversibili su individui e collettività. La concessione di un credito, la priorità di una prestazione sanitaria, la selezione di un candidato, l'assegnazione di un rischio assicurativo o l'attivazione di un controllo di sicurezza non sono più esclusivamente il risultato di un giudizio umano, ma l'esito di processi ibridi nei quali modelli algoritmici, dati storici e regole organizzative concorrono a produrre decisioni che assumono una forma sempre più opaca, distribuita e difficilmente attribuibile a un singolo attore.

Questa trasformazione segna un passaggio epistemico rilevante: la tecnologia smette di essere un semplice artefatto tecnico, assumendo invece il ruolo di attore socio-tecnico, inserito in un ecosistema complesso di norme, valori, aspettative e responsabilità. Entro tale prospettiva, l'adozione dell'IA non è mai neutrale, poiché ogni sistema algoritmico incorpora implicitamente scelte normative, assunzioni valoriali e priorità organizzative che si riflettono sugli esiti prodotti. I dati, i codici e l'architettura stessa di queste tecnologie diventano così luoghi di sedimentazione di decisioni che, pur apparendo tecniche, hanno una natura eminentemente politica ed etica.

È proprio questa dimensione, spesso sottovalutata nelle prime fasi di diffusione dell'IA, ad aver reso evidente l'insufficienza di approcci puramente tecnologici o di compliance frammentata, fondati su controlli ex post o su adattamenti incrementali di norme preesistenti. L'integrazione dell'IA nei processi decisionali ha infatti messo in crisi categorie tradizionali del diritto, del risk management e della governance aziendale. Concetti come responsabilità, imputabilità, diligenza professionale e controllo assumono nuove configurazioni quando le decisioni

sono mediate da modelli probabilistici, addestrati su grandi volumi di dati e capaci di apprendere comportamenti emergenti non sempre pienamente prevedibili.

La delega decisionale all'IA, anche quando parziale, produce una redistribuzione delle responsabilità che non può essere affrontata attraverso strumenti concettuali pensati per sistemi deterministici o per catene decisionali lineari. Ne consegue la necessità di ripensare la governance dell'IA come un dominio autonomo, dotato di propri principi, strumenti e meccanismi di controllo, in grado di tenere insieme innovazione tecnologica, tutela dei diritti fondamentali, fiducia degli stakeholder e sostenibilità organizzativa.

In questo contesto si colloca l'emergere di due grandi famiglie di strumenti di regolazione dell'IA, profondamente diverse per natura giuridica ma complementari nella funzione. Da un lato, gli standard volontari internazionali sviluppati in ambito ISO/IEC, che forniscono cornici concettuali, linee guida operative e, in alcuni casi, veri e propri sistemi di gestione orientati alla governance dell'IA. Dall'altro, le regolamentazioni cogenti, tra cui spicca il Regolamento europeo UE 2024/1689 (c.d. AI Act) che introduce obblighi legali, classificazioni di rischio e meccanismi sanzionatori direttamente applicabili negli Stati membri.

La coesistenza di questi due livelli di regolazione non rappresenta una duplicazione ridondante, ma riflette la complessità del fenomeno IA e la necessità di affrontarlo attraverso strumenti differenziati ma coordinati.

Gli standard ISO/IEC si collocano in una logica di soft law tecnica e organizzativa. Essi non impongono obblighi giuridici diretti, ma definiscono modelli di riferimento condivisi a livello internazionale, capaci di orientare le pratiche delle organizzazioni verso livelli più elevati di maturità, controllo e responsabilità. In

particolare, gli standard di sistema introducono una logica di governance basata su processi, ruoli, responsabilità, valutazione dei rischi e miglioramento continuo, consentendo di trasformare principi astratti in pratiche verificabili. Questa dimensione è cruciale nel contesto dell'IA, dove l'assenza di strutture organizzative adeguate rappresenta uno dei principali fattori di rischio, spesso più rilevante delle caratteristiche tecniche dei singoli modelli.

Le regolamentazioni cogenti, al contrario, rispondono all'esigenza di tutelare interessi pubblici fondamentali, come i diritti umani, la sicurezza e il corretto funzionamento del mercato. A tal fine, il Regolamento europeo adotta un impostazione risk-based, introducendo una classificazione dei sistemi di IA in base al livello di rischio e impone requisiti stringenti per quelli considerati ad alto rischio, incidendo direttamente sulle modalità di progettazione, sviluppo, implementazione e utilizzo. Tuttavia, la norma giuridica, per sua natura, definisce il "cosa" deve essere rispettato, ma difficilmente entra nel dettaglio del "come" le organizzazioni debbano strutturarsi per garantire un controllo effettivo e continuo dei sistemi di IA nel tempo. È proprio in questo spazio operativo che gli standard ISO/IEC trovano la loro funzione strategica. L'integrazione tra standard volontari e regolamentazione cogente diventa dunque un elemento chiave per un'adozione consapevole dell'IA. Non si tratta di scegliere tra compliance legale ed eccellenza organizzativa, ma di riconoscere che la conformità normativa, soprattutto in un ambito dinamico e tecnologicamente complesso come quello dell'IA, non può essere raggiunta in modo sostenibile senza un adeguato sistema di governance. Gli standard ISO/IEC offrono alle organizzazioni un linguaggio comune, strumenti strutturati e meccanismi di audit che consentono di rendere dimostrabile ciò che il legislatore richiede in termini di controllo, responsabilità e gestione del rischio.

La governance dell'IA, dunque, si configura come un dominio ibrido, nel quale diritto, tecnica ed etica non operano come sfere separate, ma come dimensioni interdipendenti di un unico sistema socio-tecnico. Comprendere il ruolo e le differenze tra questi strumenti è quindi una condizione necessaria per evitare approcci superficiali o meramente reattivi all'adozione dell'IA. Le organizzazioni che affrontano l'IA esclusivamente come un problema tecnologico tendono a sottovalutare gli impatti sistemici e a esporsi a rischi reputazionali, legali e operativi difficilmente governabili a posteriori. Al contrario, un approccio strutturato, fondato sull'integrazione tra standard di gestione e regolamentazione, consente di collocare l'IA all'interno di un quadro di decisione consapevole, nel quale l'innovazione non è in opposizione alla responsabilità, ma ne diventa una componente essenziale. È in questa prospettiva che la governance dell'IA deve essere letta non come un vincolo all'adozione tecnologica, ma come l'infrastruttura abilitante che rende possibile un uso dell'IA coerente con i valori democratici, la sostenibilità economica e la fiducia degli stakeholder, ponendo le basi per una trattazione sistematica dei modelli di gestione che verranno analizzati nei paragrafi successivi.

2. ISO 26000: la responsabilità sociale come fondamento etico

La norma ISO 26000, pubblicata dalla *International Organization for Standardization* (2010), rappresenta uno dei riferimenti più solidi e al tempo stesso più fraintesi nel dibattito contemporaneo sulla governance dell'IA[1].
Nata come linea guida internazionale sulla responsabilità sociale delle

[1] International Organization for Standardization (2010). *ISO 26000: Guidance on social responsibility*. Geneva: International Organization for Standardization.

organizzazioni, la norma ISO 26000 svolge una funzione più profonda e strutturale: fornire una grammatica etica condivisa per orientare le decisioni organizzative in contesti ad alta complessità e impatto sociale. La natura non prescrittiva della norma, che non si colloca nel perimetro delle norme certificabili, è stata spesso letta come un limite in ambienti fortemente orientati alla compliance. Eppure, l'assenza di requisiti tecnici puntuali e di controlli formali non ne riduce la rilevanza, facendone invece una cornice epistemica capace di attraversare domini tecnologici diversi, anticipando questioni che la normazione tecnica e la regolazione giuridica tendono a intercettare solo in una fase successiva. In tal senso, la ISO 26000 opera come fondamento etico-culturale su cui possono innestarsi sistemi di gestione, controlli organizzativi e obblighi legali, fornendo un orientamento stabile in un contesto tecnologico intrinsecamente instabile.

La distinzione tra norme certificabili e linee guida, centrale nell'architettura ISO, assume qui un significato particolare. Le norme di sistema – come ISO 9001[2] e ISO/IEC 27001[3] – sono progettate per essere auditabili e certificabili, e dunque per rendere verificabile l'adozione di determinati processi e controlli. Al contrario, la ISO 26000 non mira a dimostrare la conformità a requisiti specifici, orientando piuttosto il modo in cui un'organizzazione definisce le proprie responsabilità nei confronti della società.

Lo scarto tra la conformità a requisiti specifici e la definizione delle proprie responsabilità nei confronti della società è particolarmente rilevante nel campo dell'IA, dove i rischi associati ai sistemi algoritmici non sono sempre e

[2] International Organization for Standardization. (2015). *ISO 9001:2015 – Quality management systems — Requirements*. Geneva: ISO.

[3] International Organization for Standardization. (2022). *ISO/IEC 27001:2022 – Information security, cybersecurity and privacy protection — Information security management systems — Requirements*. Geneva: ISO.

immediatamente riconducibili a violazioni formali, ma esito di dinamiche sistemiche che coinvolgono scelte organizzative, culturali e valoriali. L'adozione di un sistema di IA può risultare tecnicamente conforme e legalmente accettabile, pur producendo effetti discriminatori, opachi o socialmente dannosi.

È in questo spazio grigio, tra ciò che è formalmente lecito e ciò che è eticamente sostenibile, che la ISO 26000 esercita la propria funzione più rilevante. I principi su cui la norma si fonda – responsabilità, trasparenza, comportamento etico, rispetto degli stakeholder, legalità, rispetto delle norme internazionali di comportamento e tutela dei diritti umani – assumono una valenza specifica quando applicati alle tecnologie intelligenti. La responsabilità, intesa come accountability, richiede una riflessione sulla distribuzione delle decisioni tra esseri umani e sistemi automatizzati. In presenza di modelli che apprendono da dati storici, che operano su base probabilistica e che possono produrre risultati non pienamente prevedibili, l'accountability diventa una proprietà emergente del sistema organizzativo nel suo complesso e non invece una qualità attribuibile a un singolo attore. La trasparenza, a sua volta, non coincide semplicemente con la disponibilità di informazioni tecniche, ma riguarda la capacità dell'organizzazione di spiegare, giustificare e rendere comprensibili le decisioni automatizzate a coloro che ne subiscono gli effetti. Il comportamento etico non si esaurisce nel rispetto delle leggi vigenti e implica una valutazione preventiva degli impatti sociali e una disposizione attiva alla mitigazione dei rischi, anche quando questi non sono ancora codificati in obblighi normativi.

Il rispetto degli interessi degli stakeholder rappresenta una questione particolarmente articolata. I sistemi algoritmici incidono su una pluralità di soggetti, spesso non immediatamente visibili o rappresentati nei processi decisionali interni all'organizzazione. Utenti finali, cittadini, lavoratori, comunità

locali e persino generazioni future possono essere influenzati da scelte tecnologiche apparentemente circoscritte. La ISO 26000 invita le organizzazioni a riconoscere questa pluralità di interessi e a integrare il dialogo con gli stakeholder come parte integrante della governance, superando una visione ristretta centrata esclusivamente su clienti e azionisti. Tale approccio risulta essenziale per l'IA, dove la distanza tra chi progetta, chi utilizza e chi subisce gli effetti dei sistemi può essere particolarmente ampia. La tutela dei diritti umani, esplicitamente richiamata dalla norma, rappresenta forse il punto di contatto più diretto tra ISO 26000 e il dibattito contemporaneo sull'IA.

Sebbene la norma non menzioni specificamente i sistemi algoritmici, essa fornisce un quadro concettuale immediatamente applicabile a tecnologie che influenzano l'accesso a risorse, opportunità e servizi essenziali. L'IA può incidere sul diritto alla non discriminazione, alla privacy, alla libertà di espressione, al lavoro e alla dignità personale, spesso in modo indiretto e mediato. La norma ISO 26000 consente di affrontare questi impatti come dimensioni strutturali dell'adozione tecnologica, richiedendo alle organizzazioni di valutare sistematicamente le conseguenze delle proprie scelte. In questo senso, la rilevanza della ISO 26000 risiede nel suo ruolo di cornice etica di responsabilità capace di mediare gli effetti sociali, economici e ambientali della tecnologia.

La norma consente di tematizzare questioni quali la supervisione umana delle decisioni automatizzate, la tracciabilità delle scelte algoritmiche, la prevenzione di discriminazioni basate su genere, età, origine etnica o condizioni di vulnerabilità, e la protezione della dignità umana in applicazioni particolarmente sensibili come la sorveglianza, il credit scoring o la selezione del personale. Aspetti, questi, spesso affrontati in modo frammentario nei dibattiti tecnici, trovano nella ISO 26000 un principio di unificazione concettuale.

Un ulteriore elemento di rilievo per la norma ISO 26000 riguarda la dimensione organizzativa e lavorativa. L'adozione dell'IA infatti modifica ruoli, competenze e condizioni di lavoro, introducendo nuove forme di controllo, valutazione e sorveglianza. La norma ISO 26000 comporta la considerazione di questi impatti come parte integrante della responsabilità sociale dell'organizzazione. In ambito Human Resource Management (HRM), l'uso dell'IA pone questioni delicate in termini di equità, trasparenza e autonomia professionale, che non possono essere risolte esclusivamente attraverso requisiti tecnici o legali. La norma ISO 26000 accompagna l'integrazione di queste tecnologie avanzate in modo coerente con il rispetto delle persone, evitando derive strumentali o invasive.

La ISO 26000 consente di estendere l'attenzione etica oltre i confini dell'organizzazione, introducendo criteri di valutazione dei rischi sociali lungo la supply chain e sollecitando l'inclusione di clausole contrattuali relative all'uso lecito ed etico dell'IA. L'estensione della responsabilità sociale verso le complessive catene di fornitura digitale risulta decisiva nel campo dell'IA, articolato in ecosistemi complessi che coinvolgono fornitori di dati, sviluppatori di modelli, piattaforme cloud e integratori di sistema. La norma ISO 26000, quindi, ridimensiona il rischio di una responsabilità sociale frammentata lungo le catene globali di produzione tecnologica.

Attraverso il suo focus sulla sostenibilità, la norma ISO 26 000 interviene anche sulla dimensione ambientale offrendo un quadro per integrare considerazioni ambientali nelle scelte tecnologiche, promuovendo l'efficienza, l'uso responsabile delle infrastrutture *cloud* e l'allineamento con le strategie di *Environmental Social Governance* (ESG). Anche in questo caso, la norma non fornisce metriche tecniche, ma orienta il processo decisionale verso una valutazione sistemica degli impatti.

Nel complesso, dunque, la ISO 26000 definisce il modo in cui un'organizzazione dovrebbe decidere di adottarli, utilizzarli e controllarli. Essa prepara il terreno culturale e organizzativo su cui possono innestarsi norme tecniche più specifiche e sistemi di gestione certificabili, come quelli previsti dalla ISO/IEC 42001. La ISO 26000 svolge dunque una funzione essenziale nella governance dell'IA: non come alternativa alla normazione tecnica o alla regolazione giuridica, ma come fondamento etico che consente di dare coerenza, profondità e legittimità alle scelte organizzative in un'epoca in cui la tecnologia diventa un autonomo attore socio-tecnico.

3. ISO/ IEC 24368: etica *by design*

La norma ISO/IEC 24368[4] si colloca in una posizione cruciale nel panorama della governance dell'IA e affronta uno dei nodi più complessi e spesso elusi del dibattito contemporaneo: la traduzione operativa dell'etica all'interno dei sistemi tecnici. Se la ISO 26000 fornisce una cornice valoriale e culturale che orienta le decisioni organizzative, la ISO/IEC 24368 rende quei valori tecnicamente significativi, intervenendo direttamente nei processi di progettazione, sviluppo, addestramento, implementazione e monitoraggio dei sistemi di IA.

In questa prospettiva, l'etica non è più concepita come un insieme di principi esterni o successivi alla tecnologia, ma come una proprietà emergente del sistema socio-tecnico nel suo complesso, che deve essere intenzionalmente incorporata nelle architetture, nei dati e nei processi algoritmici. La nozione di

[4] International Organization for Standardization & International Electrotechnical Commission (2024) *ISO/IEC 24368: Information technology — Artificial intelligence — Overview of ethical and societal concerns*. Geneva: International Organization for Standardization.

etica by design, centrale nella ISO/IEC 24368, non va interpretata come un'estensione metaforica di pratiche di design responsabile, ma come un preciso orientamento normativo che riconosce la natura intrinsecamente normativa dei sistemi di IA: ogni scelta tecnica, dalla selezione delle variabili alla definizione delle funzioni obiettivo, dall'architettura del modello alle modalità di interazione con l'utente, implica una presa di posizione rispetto a ciò che viene considerato rilevante, desiderabile o accettabile.

La ISO/IEC 24368 ha origine dal presupposto che tali scelte non possano essere lasciate all'implicito e che debbano invece essere discusse e governate all'interno di un quadro strutturato. La norma, dunque, si concentra sul problema, eminentemente tecnico, di come valori già riconosciuti possano essere incorporati senza perdere coerenza, tracciabilità e controllabilità lungo l'intero ciclo di vita dei sistemi di IA.

Nella progettazione, ciò implica una riflessione preventiva sugli scopi del sistema, sui contesti d'uso previsti e su quelli ragionevolmente prevedibili, nonché sugli impatti potenziali sugli individui e sui gruppi sociali coinvolti. L'*etica by design* si manifesta come capacità di anticipazione, ovvero come attitudine a interrogarsi sulle conseguenze sistemiche delle scelte tecnologiche prima che esse si materializzino. Tale capacità si si traduce nella necessità di documentare le assunzioni progettuali, le alternative considerate e le motivazioni che hanno condotto a determinate decisioni, creando una base informativa che potrà essere successivamente oggetto di audit e revisione.

Nella fase di sviluppo e addestramento, la ISO/IEC 24368 pone particolare attenzione al ruolo dei dati come veicolo primario di valori e bias, quindi come rappresentazioni che riflettono strutture sociali, storie di esclusione e asimmetrie di potere che possono essere amplificate dai modelli di IA. La norma

sollecita una valutazione sistematica dei rischi etici associati ai dati, includendo la provenienza, la qualità, la rappresentatività e le modalità di raccolta. In questa prospettiva, la prevenzione non può essere ridotta a un problema statistico e richiede una comprensione più ampia delle dinamiche sociali che i dati incorporano. La norma non prescrive tecniche specifiche di debiasing, ma richiede che le scelte adottate siano consapevoli, motivate e coerenti con i valori dichiarati dall'organizzazione. Ancora una volta, l'accento è posto sulla tracciabilità delle decisioni, più che sull'illusione di una neutralità algoritmica irraggiungibile.

Durante l'implementazione, l'*etica by design* assume una dimensione relazionale attenta alle interazioni con utenti, operatori e altri sistemi. Entro questa prospettiva, la ISO/IEC 24368 invita a considerare l'esperienza dell'utente come parte integrante dell'architettura etica del sistema, ponendo attenzione alla comprensibilità, alla prevedibilità e alla possibilità di intervento umano. La spiegabilità viene intesa come capacità del sistema di fornire informazioni significative e contestualizzate a coloro che ne subiscono gli effetti. Pur riconoscendo che non tutte le decisioni automatizzate possono essere completamente spiegate in termini semplici, la norma richiede che l'organizzazione si interroghi su quali livelli di spiegazione siano appropriati in relazione ai rischi e agli impatti del sistema.

La fase di monitoraggio rappresenta un ulteriore tassello nell'approccio della ISO/IEC 24368. La dimensione etica, infatti, rappresenta una qualità dinamica che può degradarsi nel tempo a causa di cambiamenti nei dati, nei contesti d'uso o nelle interazioni con altri sistemi. La norma sollecita quindi l'adozione di meccanismi di osservazione continua degli output e degli effetti del sistema, al fine di individuare deviazioni, effetti inattesi o impatti negativi emergenti. Ciò

implica la definizione di indicatori, soglie di attenzione e procedure di intervento che consentano di mantenere l'allineamento tra i valori dichiarati e il comportamento effettivo del sistema. L'*etica by design* si estende, così, naturalmente in un'*etica by operation*, che riconosce la necessità di un governo continuo dell'IA.

La norma ISO/IEC 24368 è stata concepita come elemento raccordo tra linguaggi disciplinari diversi così da essere utilizzata da team multidisciplinari che includono sviluppatori, data scientist, decision maker e figure di governance. Essa fornisce un lessico comune che consente di discutere questioni etiche in termini operativi, evitando l'astrazione eccessiva da una parte e il riduzionismo tecnico dall'altra. In questo senso, la norma svolge un ruolo di mediazione rendendo possibile un dialogo strutturato tra chi definisce i valori e chi costruisce i sistemi.

La complementarità tra ISO/IEC 24368 e ISO 26000 emerge con chiarezza proprio in questa capacità di raccordo tra differenti estrazioni disciplinari e professionali, portando la riflessione etica lungo l'intero ciclo di vita dell'IA e fornendo criteri operativi che possono essere integrati in procedure, controlli e audit. La norma riconosce così la necessità di trattare la dimensione etica come un profilo progettuale soggetto a scelte, trade-off e vincoli. In questo senso, la ISO/IEC 24368 rappresenta un passo significativo verso una governance dell'IA che non separa valori e tecnologia, ma li considera come elementi co-costitutivi di un unico sistema socio-tecnico. Nel contesto attuale, caratterizzato da una crescente attenzione regolatoria e da aspettative sociali sempre più elevate, la norma assume un valore strategico per le organizzazioni che intendono adottare l'IA in modo responsabile. Pur essendo una norma volontaria e non certificabile, essa offre un riferimento essenziale per dimostrare la diligenza organizzativa

nella gestione dei rischi etici, anticipando requisiti che stanno progressivamente entrando nel perimetro delle regolamentazioni cogenti. Nel quadro complessivo della governance dell'IA, la ISO/IEC 24368 svolge dunque una funzione di cerniera collegando la dimensione valoriale della responsabilità sociale alla dimensione gestionale dei sistemi di IA, rendendo possibile il passaggio dall'*etica come principio* all'*etica come pratica*. È in questa funzione di traduzione e integrazione che la norma esprime il proprio contributo più rilevante, ponendo le basi per una successiva formalizzazione della governance dell'IA all'interno di sistemi di gestione certificabili.

4. ISO/ IEC 42001: sistema di gestione dell'Intelligenza Artificiale

La ISO/IEC 42001[5] introduce il primo sistema di gestione certificabile specificamente dedicato alle tecnologie intelligenti, riconoscendo formalmente l'impossibilità di governarla come le precedenti, e segna una svolta operativa nel panorama della governance dell'IA.

Con questo standard, l'IA viene definitivamente collocata nell'alveo delle infrastrutture organizzative critiche, al pari della qualità, della sicurezza delle informazioni o della continuità operativa. La scelta di adottare la *High Level Structure* comune agli standard ISO di sistema riflette una precisa presa di posizione epistemica: l'IA non è una tecnologia isolata, bensì un insieme di processi, decisioni, responsabilità e controlli che devono essere integrati nella gestione complessiva dell'organizzazione. In questa prospettiva, la norma mette al centro la capacità dell'organizzazione di comprendere le tecnologie intelligenti, integrarle, governarle e renderne conto nel tempo.

[5] International Organization for Standardization & International Electrotechnical Commission. (2023). *ISO/IEC 42001 – Artificial intelligence management system*. Geneva: International Organization for Standardization.

La ISO/IEC 42001 trasforma la governance in pratica strutturata che affronta il problema complesso della sostenibilità organizzativa associato all'IA. Assumendo che l'etica e la responsabilità non siano affidabili a iniziative episodiche, a comitati isolati, né alle buone intenzioni individuali, l'ISO/IEC 42001 ne sostiene l'istituzionalizzazione all'interno di un sistema di gestione dotato di ruoli, processi, controlli e meccanismi di miglioramento continuo. La governance dell'IA viene trattata, dunque, come una funzione manageriale soggetta agli stessi criteri di pianificazione, attuazione, verifica e riesame che caratterizzano gli altri sistemi di gestione maturi.

La norma richiede un'analisi del contesto e delle parti interessate particolarmente approfondita come strumento di base per la promozione di un allineamento tra le aspettative degli stakeholder coinvolti in questo complesso ambiente socio-tecnico. In questo modo, la ISO/IEC 42001 invita le organizzazioni a interrogarsi non solo sugli obiettivi interni, ma anche sugli impatti esterni delle proprie soluzioni di IA, riconoscendo che la legittimità dell'uso di queste tecnologie dipende dalla capacità di contemperare interessi potenzialmente confliggenti. La governance dell'IA si configura così come un esercizio di responsabilità in chiave preventiva, nel quale l'organizzazione è chiamata a esplicitare le proprie assunzioni, a valutare i contesti di applicazione e a riconoscere i limiti entro cui l'automazione decisionale può essere considerata accettabile.

La gestione del rischio – che include rischi etici, legali, reputazionali, operativi e strategici – rappresenta uno degli assi portanti dello standard ISO/IEC 42001. Tale attività non si esaurisce nell'identificazione dei possibili scenari di malfunzionamento, includendo altresì l'analisi degli impatti possibili sugli individui e sulla società, nonché delle conseguenze derivanti da un uso improprio o non previsto dei sistemi. Il trattamento del rischio diventa così un processo continuo, che accompagna l'intero ciclo di vita dell'IA e che richiede decisioni informate, documentate e periodicamente riesaminate.

La definizione dei ruoli e delle responsabilità costituisce un altro elemento distintivo

della ISO/IEC 42001. Si prevede che l'organizzazione definisca chiaramente i soggetti, i ruoli e le responsabilità corrispondenti lungo la complessiva catena decisionale dell'IA. In particolare, si richiede di rendere esplicita la distribuzione delle responsabilità, attribuendo ruoli chiari per la supervisione, il controllo e l'intervento umano. Lo standard ISO/IEC 42001 interviene, quindi, a supporto delle prescrizioni giuridiche contribuendo a ridurre il divario tra la responsabilità formale e la responsabilità sostanziale, rafforzando pure i diritti fondamentali nel contesto tecnologico emergente.

I controlli su dati, modelli e processi decisionali rappresentano il cuore operativo del sistema di gestione. La norma riconosce che la qualità dell'IA dipende in larga misura dalla qualità dei dati e dalla coerenza dei processi di trattamento che li trasformano in decisioni. Invece di prescrivere soluzioni tecniche specifiche, la ISO/IEC 42001 adotta un approccio basato sui controlli, richiedendo che l'organizzazione definisca, implementi e mantenga misure adeguate in funzione del contesto e del rischio, permettendo di adattare il sistema di gestione a tecnologie diverse e in rapida evoluzione. Muovendo in questa direzione, la norma privilegia la verificabilità dei processi rispetto alla standardizzazione delle soluzioni, rendendo possibile l'audit senza irrigidire l'innovazione.

Il monitoraggio continuo e il miglioramento rappresentano l'elemento che conferisce alla ISO/IEC 42001 una dimensione temporale esplicita. La norma riconosce che una governance efficace non può essere episodica e che deve essere sostenuta nel tempo attraverso meccanismi di osservazione, valutazione e revisione. Il sistema di gestione dell'IA diventa un dispositivo di apprendimento organizzativo, capace di adattarsi ai cambiamenti tecnologici e normativi senza perdere coerenza, migliorando in modo continuo la capacità dell'organizzazione di governare l'IA in modo responsabile.

Altro aspetto rilevante dello standard ISO/IEC 42001 consiste nella possibilità di sottoporre il sistema di IA a verifiche indipendenti. L'audit consente di trasformare dichiarazioni di principio e policy interne in evidenze verificabili, rendendo la governance dell'IA osservabile dall'esterno. Per auditor, giuristi e autorità di controllo,

la ISO/IEC 42001 fornisce un linguaggio comune e una struttura di riferimento che facilita la valutazione della diligenza organizzativa.

La relazione tra la ISO/IEC 42001 e la conformità normativa è di natura strategica. Pur non essendo uno strumento giuridico, lo standard offre un supporto concreto alla dimostrazione di conformità ai requisiti regolatori traducendo obblighi legali in processi organizzativi strutturati che diventano un fattore di resilienza normativa.

In conclusione, la ISO/IEC 42001 può essere letta come il punto di convergenza tra etica, tecnica e diritto: non si sostituisce alla riflessione etica, né alla regolamentazione giuridica, ma fornisce l'infrastruttura organizzativa per renderle operative e sostenibili.

5. AI Act dell'Unione Europea – Regolamento cogente

Il Regolamento UE 2024/1689 sull'Intelligenza Artificiale (c.d. AI Act)[6] rappresenta il primo tentativo di traduzione delle preoccupazioni etiche, sociali e di sicurezza in obblighi giuridici vincolanti. La sua rilevanza risiede nel contenuto prescrittivo e pure nella scelta di fondo che lo ispira: il riconoscimento dell'IA come infrastruttura regolabile in quanto tale, e non invece come semplice estensione di tecnologie digitali preesistenti.

L'impostazione *risk-based* adottata dal legislatore europeo costituisce il cuore concettuale del Regolamento, che introduce una disciplina modulata in relazione al grado di rischio associato all'applicazione dell'IA in un particolare contesto. La classificazione del grado di rischio associato all'applicazione dell'IA in differenti domini rappresenta un atto normativo che incorpora valutazioni di ordine etico e politico. Stabilire che l'applicazione di un sistema di IA nel campo dell'occupazione, della gestione dei lavoratori e nell'accesso al lavoro sia ad alto rischio significa riconoscere che l'automazione decisionale nel mondo del lavoro può incidere in modo significativo

[6] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence.

sulla vita delle persone, rendendo necessaria l'introduzione di obblighi e garanzie rafforzati[7].

In questa prospettiva, il rischio diventa una categoria di mediazione tra etica e diritto, analoga a quella che, nei sistemi di gestione, consente di tradurre valori e principi in controlli organizzativi.

Per i sistemi classificati come ad alto rischio, il Regolamento introduce un insieme articolato di obblighi che riguardano l'intero ciclo di vita delle tecnologie intelligenti, dalla progettazione all'uso operativo. Tra questi figurano la gestione documentata del rischio, la qualità e la governance dei dati, la tracciabilità delle operazioni, la supervisione umana, la robustezza tecnica e la sicurezza informatica.

I requisiti del Regolamento, pur formulati in termini giuridici, si pongono in rapporto di complementarietà con i sistemi di gestione normati in ambito ISO. Essi presuppongono infatti l'esistenza di processi organizzativi stabili, ruoli chiaramente definiti e meccanismi di controllo e monitoraggio continui. In assenza di tali elementi, la conformità al Regolamento rischia di rimanere puramente formale. Il Regolamento stabilisce ciò che deve essere garantito in termini di tutela dei diritti e di gestione del rischio, ma non entra nel dettaglio delle architetture organizzative necessarie per assicurare tali garanzie in modo sistematico. Dal canto loro, gli standard ISO forniscono un modello di sistema di gestione che consente di rendere operativi e verificabili requisiti che il legislatore formula in termini generali. In questa relazione, la norma tecnica non si pone come alternativa al diritto, ma come infrastruttura abilitante della conformità giuridica.

Dal punto di vista della filosofia delle tecnologie, il Regolamento può essere interpretato come un tentativo di ricondurre le tecnologie computazionali avanzate entro un perimetro di governabilità democratica, riconoscendo che l'automazione si inserisce in relazioni sociali asimmetriche richiedendo garanzie e obblighi differenziati in relazione

[7] Regulation (EU) 2024/1689. Allegato III – Sistemi di IA ad alto rischio di cui all'articolo 6, paragrafo 2.

ai particolari ambiti di applicazione. Le norme ISO sostengono la capacità delle organizzazioni di governare i propri sistemi di IA in termini strettamente operativi.

La governance dell'IA emerge, dunque, come spazio intermedio tra norma giuridica e pratiche tecnico-organizzative. Il Regolamento richiede che i sistemi di IA ad alto rischio siano progettati e utilizzati in modo tale da consentire un intervento umano significativo. Questa prescrizione, apparentemente semplice, solleva interrogativi complessi sul ruolo dell'essere umano nei sistemi automatizzati e implica una riflessione organizzativa sulle competenze, sulle responsabilità e sulle condizioni necessarie affinché l'intervento umano sia realmente possibile, efficace e conforme a quanto previsto dal Regolamento. Il Regolamento indica il fine, lasciando aperta la questione dei mezzi, che viene invece affrontata attraverso gli standard internazionali.

Anche gli obblighi di tracciabilità, documentazione e responsabilità previsti dal Regolamento richiedono interventi attuativi di carattere organizzativo che passano dalle norme ISO.

Nell'ambito di questo rapporto di complementarietà tra normativa cogente e standard internazionali, la dimensione sanzionatoria del Regolamento emerge come elemento di rafforzamento dei processi, delle strutture e delle responsabilità organizzative (Fig. 1). La conformità normativa si configura quindi come processo continuo che richiede adattamenti organizzativi, attività di monitoraggio, riesame e interventi per il miglioramento continuo. La ISO/IEC 42001 offre una risposta concreta a questa esigenza, fornendo un modello di governance che consente di integrare la compliance nel funzionamento ordinario dell'organizzazione.

Il Regolamento ha attirato critiche consistenti in quanto ritenuto eccessivamente vincolante e di conseguenza capace di frenare le attività di ricerca, sviluppo e le applicazioni sperimentali. Tale lettura, che presuppone una contrapposizione tra innovazione e regolazione, non tiene in adeguata considerazione la complessità dei sistemi socio-tecnici emergenti, i costi notevoli da sostenersi nel medio-lungo periodo e la consistente incertezza rispetto al loro legittimo impiego. Al contrario, l'integrazione

del Regolamento con strumenti normotecnici come gli standard ISO favorisce l'apprendimento organizzativo e sostiene percorsi di evoluzione tecnologica in linea con i principi e i valori democratici.

Nel quadro complessivo della governance dell'IA, il Regolamento può dunque essere interpretato come il livello giuridico di un'architettura più ampia, che include dimensioni etiche, tecniche e organizzative. Esso stabilisce i confini entro cui l'IA può essere utilizzata legittimamente, ma lascia alle organizzazioni la responsabilità di costruire i meccanismi interni necessari per rispettare tali confini in modo sostanziale. È in questa integrazione che si gioca la possibilità di una transizione intelligente legittima, affidabile e socialmente sostenibile.

Figura 1 - Confronto strutturato tra i quattro strumenti

	ISO 26000	**ISO/IEC 24368**	**ISO/IEC 42001**	**Regolamento UE 2024/1689**
Natura	Linea guida	Norma tecnica	Norma di sistema	Regolamento
Certificabile	No	No	Sì	No
Focus	Etica e responsabilità	Etica by design	Governance IA	Conformità legale
Approccio	Valoriale	Tecnico-etico	Gestionale	Giuridico
Obbligatorietà	Volontaria	Volontaria	Volontaria	Cogente
Ruolo	Fondamento culturale	Implementazione etica	Controllo e audit	Enforcement

Fonte: Elaborazione degli autori