



## **I position paper della Fondazione ENIA**

Data di ricezione: 11.10.25 / Data accettazione: 12.3.25 / Data di pubblicazione: 31.12.25  
doi: 10.82015/NNR.2025.100118

### **Autovalutazione etica dei sistemi di Intelligenza Artificiale.**

#### **Un framework di governance, rischio e responsabilità organizzativa**

*Ethical Self-Assessment of Artificial Intelligence Systems.*

*A Framework for Governance, Risk, and Organizational Responsibility*

Arianna Ciracò, Stefano Gorla, Valeria Lazzaroli, Rebecca Paternostro\*

#### **1. Etica dell'intelligenza artificiale e governo della decisione**

L'introduzione dell'intelligenza artificiale (IA) nei processi organizzativi sta incidendo in modo profondo sul modo in cui le decisioni vengono prese, giustificate e governate. Chi oggi siede nei consigli di amministrazione o ha responsabilità di indirizzo strategico non può più considerare l'adozione di sistemi di IA come una scelta puramente tecnica o

---

\* Arianna Ciracò, co-direttrice del Dipartimento AI Ethics di Fondazione Ente Nazionale per l'Intelligenza Artificiale (ENIA) e Responsabile nazionale del Dipartimento "AI e Cybersecurity" di Fondazione AIGA – Tommaso Bucciarelli / E-mail: [studiolegale@studiociraco.it](mailto:studiolegale@studiociraco.it); Stefano Gorla, co-direttore del Dipartimento AI ETHICS di Fondazione Ente Nazionale per l'Intelligenza Artificiale (ENIA) / E-mail: [stefano.gorla@acmcert.net](mailto: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](mailto:presidenza@enia.ai); Rebecca Paternostro, comitato etico di Fondazione Ente Nazionale per l'Intelligenza Artificiale (ENIA) / E-mail: [rpaternostro@outlook.it](mailto:rpaternostro@outlook.it).



delegabile. Ogni sistema che apprende, classifica, prevede o raccomanda introduce una nuova forma di mediazione tra l'intenzione umana e l'esito della decisione, ridefinendo in modo sostanziale il perimetro della responsabilità.

I sistemi di IA operano attraverso modelli probabilistici addestrati su dati che riflettono storie, contesti e scelte pregresse. Le decisioni che ne derivano non sono mai neutrali, né completamente trasparenti. Esse emergono dall'interazione tra progettazione algoritmica, qualità dei dati, contesto d'uso e obiettivi organizzativi. Questo rende più complesso, ma anche più necessario, esercitare una governance consapevole.

La distribuzione dei processi decisionali comporta una corrispondente distribuzione delle responsabilità associate attraverso strumenti che ne permettano il riconoscimento, l'assunzione e il governo. Dal punto di vista dell'*Enterprise Risk Management* (ERM), l'intelligenza artificiale porta in primo piano una tipologia di rischio che molte organizzazioni non sono ancora abituate a trattare in modo strutturato. I rischi etici e sociali associati all'uso dell'IA non si manifestano sempre come eventi improvvisi o perdite immediatamente misurabili. Essi si accumulano nel tempo, attraversano funzioni diverse, incidono sulla fiducia, sulla legittimità e sulla relazione tra organizzazioni e persone. Sono rischi che, se ignorati, tendono a emergere quando è ormai difficile intervenire in modo correttivo.

In questo senso, l'etica dell'IA non può essere intesa come "integrazione esterna" al sistema di gestione del rischio, ma una sua componente interna e necessaria, che consente di portare alla luce aree di vulnerabilità che sfuggono alle metriche tradizionali, offrendo ai vertici aziendali e ai board una visione più completa del profilo di rischio associato all'adozione delle tecnologie intelligenti. L'etica diventa così una forma di presidio anticipatorio, capace di segnalare criticità prima che queste si traducano in danni giuridici, reputazionali o sociali difficilmente reversibili.

Per i consigli di amministrazione, questa evoluzione comporta un cambio di prospettiva. Governare la tecnologia, adesso, significa interrogarsi non solo su ciò che i sistemi sono in grado di fare, ma su ciò che è opportuno fare, in coerenza con il mandato fiduciario verso gli stakeholder e con il ruolo dell'organizzazione nella società. La delega



tecnologica non solleva il board dalle proprie responsabilità di indirizzo e controllo; al contrario, le rende più complesse.

Il governo dell'IA richiede, dunque, la capacità di porre domande nuove, di leggere segnali deboli e di integrare dimensioni etiche, operative e strategiche in un unico quadro decisionale. Uno dei limiti più frequenti che si osservano nelle organizzazioni è la distanza tra i principi dichiarati e le pratiche effettive. Codici etici, policy e linee guida sull'uso responsabile dell'IA sono ormai diffusi, ma spesso restano scollegati dai processi di controllo interno e dai flussi informativi che alimentano le decisioni di vertice. In particolare, quei principi che non vengono adeguatamente tradotti in criteri osservabili e valutabili rimangono, di fatto, inaccessibili al governo dell'organizzazione. Non entrano nei sistemi di reporting, non orientano le priorità, non supportano l'assunzione di responsabilità.

È in questo spazio che l'autovalutazione etica dei sistemi di IA assume un valore concreto. L'autovalutazione non è un esercizio autoreferenziale né un tentativo di certificare la propria "bontà" etica. È, piuttosto, uno strumento di consapevolezza e di disciplina organizzativa. Attraverso processi strutturati di autovalutazione, l'organizzazione è chiamata a rendere esplicite le proprie scelte, a riconoscere i compromessi impliciti e a confrontarsi con le conseguenze delle proprie decisioni tecnologiche.

Dal punto di vista dei sistemi di controllo interno, l'autovalutazione etica rappresenta un punto di connessione tra le diverse linee di difesa. Per chi opera nella prima linea, essa favorisce una maggiore responsabilizzazione rispetto all'uso concreto dei sistemi di IA. Per le funzioni di risk management e compliance, costituisce una base strutturata per l'identificazione e il monitoraggio dei rischi etici lungo il ciclo di vita dei sistemi. Per le funzioni di *assurance*, infine, crea le condizioni per una valutazione più sostanziale della qualità della governance dell'IA, che vada oltre la mera conformità formale. Per i board, strumenti di autovalutazione etica permettono di ridurre una delle principali asimmetrie informative che caratterizzano la governance dell'intelligenza artificiale. Rendono discutibili e comparabili nel tempo elementi che altrimenti resterebbero



impliciti, offrendo una base più solida per le decisioni strategiche e per l'esercizio dell'*oversight*.

In questo modo, l'etica entra a pieno titolo nei processi di ERM e nei sistemi di *assurance*, diventando una dimensione governabile e integrata. L'autovalutazione etica va infine letta come un processo dinamico, non come una fotografia statica. I sistemi di IA evolvono, apprendono, cambiano comportamento nel tempo. Allo stesso modo, devono evolvere i meccanismi di controllo e di responsabilità. L'etica, in questo senso, non è un vincolo all'innovazione, ma una delle condizioni che rendono l'innovazione sostenibile, affidabile e socialmente legittima. Assumere seriamente la governance etica dell'IA significa riconoscere che le decisioni tecnologiche sono sempre anche decisioni umane, con effetti che vanno oltre l'organizzazione e incidono sulla società nel suo complesso. In questa prospettiva, l'autovalutazione etica diventa una leva essenziale per i board e per il management, uno strumento che consente di esercitare una responsabilità consapevole e di orientare lo sviluppo dell'intelligenza artificiale in modo coerente con i valori, il profilo di rischio e la missione dell'organizzazione. La sezione che segue si colloca in questo quadro, proponendo un modello operativo di autovalutazione che permette di tradurre tali principi in pratiche di controllo strutturate e integrabili nei sistemi di ERM e di *assurance*, rendendo l'etica una componente viva e governabile dell'intelligenza artificiale.

## **2. Autovalutazione etica dei sistemi di intelligenza artificiale**

Quando si parla di sistemi di IA sempre più viene in rilievo il riferimento all'etica, quale disciplina. Essa, in effetti, è una branca dell'etica applicata e si occupa proprio di indagare quali impatti si producono sulla società per l'effetto dell'uso di sistemi tanto pervasivi e dalle potenzialità *disruptive*. L'etica è una bussola che deve orientare l'agire umano verso la direzione "*giusta*", in considerazione del contesto di riferimento. Per comprendere a pieno la sua funzione è quantomai efficace prendere le mosse dal concetto di *Infosfera*, elaborato dal filosofo Luciano Floridi per indicare quello spazio nel



quale l'analogico e il digitale non sono più distinguibili, in cui c'è confusione e commistione tra *l'online* e *l'onlife*.

Prendendo le mosse da tale prospettiva, ecco che con più facilità è possibile comprendere come l'indagine contemporanea sull'etica applicata vada condotta: ossia avendo come obiettivo quello di realizzare mature "società dell'informazione", che abbiano il progetto umano come perno centrale. L'indagine, infatti, in relazione all'etica dell'IA deve avere ad oggetto i fini, i valori ed i principi ai quali ancorare lo sviluppo di questa tecnologia. In tal senso va letto anche l'obiettivo a base del Regolamento UE 2024/1689, ovvero promuovere uno sviluppo della tecnologia più potente ad oggi mai creata, che abbia al centro l'uomo, nella sua accezione più piena, comprensiva del fascio di diritti e libertà fondamentali che la sfera personale racchiude.

L'etica, dunque, insieme al diritto, costuisce il guardrail che orienta ed informa lo sviluppo tecnologico. Non a caso il lavoro prodotto del Gruppo Indipendente di Esperti di Alto Livello, sull'intelligenza artificiale (HLEG), istituito dalla Commissione europea nel giugno 2018, che ha elaborato le Linee Guida Etiche dell'IA, rappresenta uno dei documenti cardine sulla cui base è stato strutturato *l'Artificial Intelligence Act*. Il Documento promuove l'idea di un'IA affidabile, in considerazione dell'assoluto convincimento che solo la *Trustworthiness* permetta uno sviluppo ed una diffusione efficace e soddisfacente di questa tecnologia. I tre pilastri su cui si deve basare l'intero progetto sono: la legalità, l'eticità e la robustezza; ecco, dunque, che diritto ed etica assumono un rilievo centrale, al pari della tecnologia, in un rapporto ad equilibri variabili, che muta, con il mutare di molti fattori tra cui il contesto, la platea di soggetti coinvolti e le diverse categorie di rischio, di volta in volta ritenute esistenti. Il Documento è strutturato in tre capitoli: il primo, partendo da un approccio basato sui diritti fondamentali, individua i principi etici e i valori correlati che devono essere rispettati, nella progettazione, nella distribuzione e nell'utilizzo dei sistemi di IA. I principi individuati sono: il rispetto della vita umana, la prevenzione dei danni, l'equità e l'esplicabilità; il secondo, che elenca sette requisiti che i sistemi devono soddisfare, in attuazione dei principi e fornisce metodi tecnici e non, che possono essere utilizzati a tal



fine. Importante soffermarsi nel declinare i sette criteri: intervento e sorveglianza umani, robustezza tecnica e sicurezza, riservatezza e *governance* dei dati, trasparenza, diversità non discriminazione ed equità, benessere sociale ed ambientale e, infine, *accountability*. Il terzo capitolo, più operativo, fornisce una lista di controlli, concreta e non esaustiva, per la valutazione dell'IA.

A valle di tale lavoro, l'HLEG sull'IA ha presentato l'ALTAI, ovvero una lista di valutazione definitiva per un'IA affidabile, che costituisce uno strumento pratico per tradurre le linee guida etiche in una *checklist* operativa. Ma, a monte di tutte le riflessioni sin qui fatte, sta un presupposto imprescindibile: la necessità e l'urgenza di diffondere una cultura tecnologica e, in particolare, dell'intelligenza artificiale, a tutti i livelli, con il fine ultimo di permettere che l'individuo si autodetermini in relazione a tale tecnologia. Non sarà, infatti, possibile realizzare alcuna società pienamente consapevole nell'autodeterminarsi, sia come gruppo, sia in ogni suo singolo componente, in assenza di un'effettiva conoscenza delle caratteristiche, dei limiti, delle potenzialità dell'intelligenza artificiale. Non potendo qui affrontare diffusamente questo ultimo aspetto, è però utile richiamare una delle tante (anche se non sufficienti) iniziative intraprese dalle più alte Istituzioni, nel senso di stimolare l'attenzione verso tali tematiche. Ecco, dunque, in tal senso, che merita soffermarsi sulla Raccomandazione sull'etica dell'intelligenza artificiale (AI) di UNESCO (2023) nella quale si evidenzia la necessità che le interazioni con i sistemi di IA «*non siano dannose ma contribuiscano alla pace, alla dignità umana, alla sostenibilità ed alla sicurezza*». Tra i principi cardine di detta raccomandazione risulta quello di garantire che «*I sistemi di IA devono essere sottoposti a Valutazione etica dell'impatto per l'intero ciclo di vita al fine di prevenire eventuali danni agli essere umani e assicurare la promozione dei diritti umani attraverso l'intelligenza artificiale*» e che «*L'intelligenza artificiale deve essere sviluppata e usata in modo tale da massimizzare i benefici per la diversità e l'inclusività, compresa la diversità culturale, salvaguardando la non discriminazione, promuovendo la libertà di espressione e raggiungendo la parità di genere*».

Tutto ciò si collega ad un concetto di *accountability* ossia di responsabilizzazione sia del



soggetto che sviluppa il sistema di IA (provider), o modello di IA, sia di quello che la implementa all'interno della propria organizzazione (deployer). Il primo, diverrà responsabile dello sviluppo di un sistema/modello nel rispetto dei principi etici mentre il secondo, della relativa implementazione di un sistema/modello idoneo a garantire il rispetto dei principi etici e, quindi, dei diritti e delle libertà degli esseri umani. Questi sono: diritto alla vita; diritto alla libertà ed alla sicurezza; diritto al rispetto della vita privata e familiare attraverso il diritto alla riservatezza dei propri dati personali; diritto ad essere informato, educato e formato al fine che sviluppi un'autodeterminazione della propria identità; divieto di discriminazione. Per tale motivo lo sviluppo del sistema e modello di IA dovrà essere antropocentrico, al fine di garantire i sopracitati diritti in ogni fase di sviluppo della tecnologia di IA.

In questo panorama, in continua evoluzione, l'essere umano deve essere supportato dalla tecnologia, non sostituito da essa, al fine di potersi realizzare quale individuo, immerso in un contesto sociale di cui fa parte e contribuisce come cittadino, contribuendo allo sviluppo del Paese attraverso la sua forza lavoro e alla circolazione dell'economia attraverso le proprie scelte quale consumatore. Per questi motivi non può essere abbandonata la formazione quale forma di sapere e di conoscenza continua che rappresenta la prima forma attraverso il quale il soggetto apprende l'uso della nuova tecnologia di questo nuovo sistema. Su questo punto la Raccomandazione dell'UNESCO (2023) invita ad agire nell'educazione e nella ricerca: *Promuovendo programmi di consapevolezza generale sugli sviluppi dell'intelligenza artificiale e sul loro impatto sui diritti umani dei bambini; Incoraggiando la collaborazione trasversale tra l'educazione alle competenze tecniche e l'educazione all'etica dell'IA e promuovendo ricerche interdisciplinari includendo discipline diverse dalle STEM (scienza, tecnologia, ingegneria e matematica); Sostenendo la consulenza scientifico-politica per lo sviluppo di politiche sull'intelligenza artificiale e contribuendo a coltivare la consapevolezza dell'etica dell'IA.*

Il provider dovrà anche considerare quali impatti i sistemi/modelli dell'IA potranno comportare sull'ambiente circostante, valorizzando gli aspetti richiamati dalla Raccomandazione dell'UNESCO (2023) relativi a: *Valutazione dell'impatto ambientale*



*diretto e indiretto durante il ciclo di vita del sistema di IA e assicurando il rispetto delle regolamentazioni ambientali; Sviluppo e adozione di applicazioni per la resilienza al rischio di catastrofe basate sul diritto e assistite dall'etica della IA; monitoraggio, protezione e rigenerazione dell'ambiente e degli ecosistemi, e la preservazione del pianeta; Divieto dei sistemi di IA per i quali vi sia una probabilità alta di impatti negativi sproporzionati sull'ambiente. Detto questo risulta importante riassumere, di seguito, i principi cardine etici del sistema/modello di IA:*

- I. **Supervisione umana:** sia nella fase di progettazione, sviluppo che in quella relativa alla implementazione, sarebbe necessario che le singole responsabilità fossero attribuibili a persone fisiche (provider e deployer);
- II. **Trasparenza:** Chi è influenzato dalla tecnologia di IA deve essere consapevole dei potenziali impatti che la stessa potrà avere nell'ambiente circostante e, potenzialmente, nei propri diritti;
- III. **Alfabetizzazione:** La formazione è vista come elemento centrale che può influenzare tutto il sistema di IA dalla sua progettazione al suo uso andando a comportare anche eventuali compromissioni di diritti e libertà altrui;
- IV. **Responsabilità:** Chi agisce nel ciclo di vita dell'IA deve essere consapevole e responsabile delle scelte e delle soluzioni che andrà ad implementare; a tal fine specifici obblighi sono imposti, soprattutto nel caso di specifici sistemi e modelli di IA, in capo ai provider ed ai deployer.
- V. **Sicurezza e protezione:** La sicurezza deve esistere in ogni fase del ciclo di vita dell'IA dalla sua progettazione al suo mantenimento e, a tal fine, devono essere implementate apposite misure tecniche ed organizzative;
- VI. **Divieto di discriminazione:** Occorre evitare i c.d. bias (pregiudizi, deviazioni dallo standard). Dipendono da fattori sociali, tecnici, dipendenti dagli errori sistemici introdotti dalle scelte relative alla progettazione algoritmica, dai dati sporchi, dalle procedure di campionamento, dai protocolli di segnalazione o dalle supposizioni errate



che causano una discrepanza tra gli input e gli output generati dal sistema. I bias si sostanziano in: violazioni dei diritti; stereotipi; decisioni errate o inefficienti; discriminazioni di individui e gruppi.

**VII. Diritto alla protezione dei dati personali:** affinché sia garantito il diritto del soggetto ad essere lasciato solo, ossia il diritto alla tutela della propria sfera privata e familiare, occorrerà che sia garantita la protezione del dato personale nelle tre dimensioni ossia disponibilità, integrità e riservatezza ponendo particolare attenzione anche ai principi di privacy by design e privacy by default che impongono il principio di responsabilizzazione che si esplica nella giustificazione delle scelte relative al trattamento dei dati personali da parte del Titolare del trattamento ai sensi anche di quanto previsto dall'art. 5 GDPR.

Quanto sin qui detto assumerà particolare rilievo, ad esempio, nell'ambito sanitario dove i rischi per una violazione dei diritti e delle libertà degli esseri umani divengono maggiormente tangibili. In sanità, infatti, l'impiego dei sistemi e modelli di IA assumerà particolare rilievo in quanto potrà comportare particolari benefici: diagnosi precoci delle malattie, nuove sperimentazioni di farmaci, nuove strumentazioni tecnologiche. Al contempo l'ambito sanitario rappresenta il campo dove maggiormente sono presenti dati personali particolari (quali i dati sanitari) e che, pertanto, una loro violazione potrebbe comportare una violazione dei diritti e delle libertà degli esseri umani. Sarà necessario, perciò, introdurre apposite linee guida che definiranno i ruoli e le responsabilità nonché le conformità normative previste affinché vi sia il rispetto dei principi etici sopra descritti. Ciò anche secondo quanto enunciato dall'UNESCO (2023), nella sua Raccomandazione, dove invita i governi e i portatori di interessi del settore sanitario a lavorare sulle seguenti questioni:

*«Sviluppare linee guida per le interazioni umano-robot e per il loro impatto sulle relazioni umano-umano, con particolare attenzione alla salute mentale e fisica degli esseri umani, in modo tale che queste interazioni umano-robot rispettino i diritti umani, le libertà fondamentali, la promozione della diversità e la protezione delle*



*persone vulnerabili e questo soprattutto quando si usano sistemi assistiti da IA per le neurotecnologie e le interfacce cervello-computer; Impiegare i sistemi di IA in modo tale che questi abbiano un impatto positivo sulla salute umana, monitorando la comparsa della malattia, e per condividere i dati su scala globale; Promuovere ricerche d'equipe sugli effetti delle interazioni a lungo termine delle persone con i sistemi di IA, prestando particolare attenzione all'impatto psicologico e cognitivo che questi sistemi possono avere sui bambini e sui giovani».*

### **3. Simulazione di implementazione dell'etica**

Sin qui è stata condotta una panoramica generale, in ordine ai principi, valori e modelli di riferimento, che entrano in rilievo quando si affronta la complessa questione dell'etica applicata all'intelligenza artificiale. Adesso è utile soffermarsi su una simulazione d'implementazione dell'etica all'interno dell'organizzazione, utilizzando quale schema di riferimento la norma ISO/IEC TR 24368, che rappresenta un autorevole strumento operativo. Gestire l'etica dell'IA è diventato ormai un punto imprescindibile della *governance* dei sistemi di IA. Ma una corretta gestione non è completa se non possiede il pilastro dei controlli. Attraverso metodologie strutturate è quindi possibile che le organizzazioni verifichino l'implementazione dei principi etici, la loro condivisione e applicazione.

Un supporto all'implementazione e al controllo può essere fornito dalla consultazione dello standard ISO/IEC TR 24368 che descrive proprio come affrontare i principi e i temi etici dell'IA. Esso offre una panoramica delle preoccupazioni etiche e sociali, fornendo principi, processi, metodi di riferimento senza prescrivere un sistema di valori specifico. È destinato a tecnologi, regolari, gruppi d'interesse e società nel loro complesso a cui propone una mappatura di standard internazionali rilevanti per affrontare tali preoccupazioni. Attraverso capitoli molto dettagliati guida il lettore ad affrontare in maniera consapevole e puntuale molti degli aspetti relativi all'etica dell'IA e ai suoi principi.



Introduciamo, adesso, la leva dei controlli a partire da un questionario di autovalutazione composto da 19 classi di principi suddiviso in 176 domande alle quali l'azienda deve rispondere con un valore di maturità/implementazione da 1 a 4. Inoltre, ad ogni classe l'organizzazione attribuisce un peso che deriva dal contesto, dal mercato, dall'applicazione/uso dell'IA e dall'impatto nel non rispetto della classe di riferimento. Per ogni classe viene calcolato il valore medio delle risposte e verificato l'impatto che questa produce. Definiamo  $\epsilon_n$  la domanda relativa n-esima ad una classe  $E_i$  i-esima e  $w_i$  il peso indicato dall'organizzazione per la classe i-esima. Riportiamo un esempio di autovalutazione con i relativi calcoli. I valori indicati sono puramente indicativi. La Figura 1 illustra un estratto dell'autovalutazione, mentre l'immagine seguente (Fig. 2) mostra le classi con il risultato ed il relativo peso attribuito. I risultati sono stati analizzati e il grafico radar illustra la copertura relativa alle varie classi confrontata con le valutazioni peggiori e quelle migliori, i pesi sono stati mantenuti uguali (Fig. 3). Emergono dal grafico aree (classi) relativamente coperte, cioè con un alto livello di implementazione e altre classi scoperte (esempio privacy, safety e security, etc.).

Questo permette una visione immediata di dove intervenire e sanare i gap. È stata anche considerata la distribuzione, in questo caso normale, dei risultati (Fig. 4) dimostrando una curva e delle risposte coerenti leggermente allargata verso le code, con un valore medio di 8.63 e una deviazione standard di 4.12.

Per una valutazione più accurata si sono analizzati rispetto alla presentazione del questionario e rispetto ad ordinando i risultati in ordine crescente. I grafici seguenti riportano l'andamento dei risultati (Fig. 5 e 6). La valutazione tra dati non ordinati rispetto a quelli ordinati ha portato ad ulteriori considerazioni sulle classi (Fig. 7); infatti, alcune classi si sono spostate dimostrando una maggior o minor attenzione (maturità di implementazione) e sensibilità sui principi. Come si può notare (Fig. 8) dalle frecce verdi e rosse, come alcune classi sono migrate verso il basso o l'alto; ad esempio, le classi relative ai rischi, ai quadri etici, ai diritti umani, etc. sono migrate verso il basso, cioè, posseggono un punteggio di maturità maggiore, mentre privacy,



pratiche di lavoro etc, sono migrate verso l'alto ottenendo un punteggio minore. Questo implica una maggior/minor copertura e, quindi, una maggior/minor attenzione al tema.

**Figura 1 - Estratto dell'autovalutazione**

<b>SCOPO</b>	<b>3,33</b>	<b>Basso</b>
Lo scopo e l'ambito del sistema AI sono definiti e documentati.	4	
Sono identificati destinatari e stakeholder interni/esterni.	3	
Sono esplicitati benefici attesi e potenziali impatti/limiti.	3	
<b>RIFERIMENTI NORMATIVI</b>	<b>2,93</b>	<b>Medio</b>
È stato definito un meccanismo per aggiornare i riferimenti nel tempo.	4	
Sono elencati i riferimenti applicabili e le loro versioni (es. ISO/IEC 22989, 23894, 42005, 26000, 27001/27002, 27701, 38505-1, 38507, 9241-210).	3	
Il glossario è allineato a ISO/IEC 22989 e condiviso nel team.	2	
Definizioni chiave chiarite: bias, fairness, privacy, sicurezza, sostenibilità, tracciabilità, valore nella catena (provider/producer/customer/partner/subject).	3	
Bias/fairness: ISO/IEC TR 24027; Trasparenza/spiegabilità: ISO/IEC TR 24028; Casi d'uso: ISO/IEC TR 24030.	4	
Gestione del rischio AI secondo ISO/IEC 23894 (in linea con ISO 31000).	3	
Governance dell'AI per le organizzazioni: ISO/IEC 38507.	2	
Governance dei dati (ISO/IEC 38505-1), sicurezza (ISO/IEC 27001/27002) e privacy (ISO/IEC 27701).	3	
ISO Guide 82 (sostenibilità) e ISO 26000 (responsabilità sociale) collegati alla fiducia nell'AI.	3	
Regolamento Europeo AIAct	3	
Regolamento Europeo GDPR	3	
Rispetto norme internazionali nell'uso/promozione AI.	3	
Tutela interessi legittimi (IP, sicurezza) e valori degli stakeholder.	2	
Adozione di norme internazionali quando la legge locale è insufficiente	3	
Evitare complicità con pratiche contrarie a tali norme	3	
<b>RISCHI E IMPATTI</b>	<b>3,67</b>	<b>Basso</b>
Sono stati valutati i rischi di: Danno finanziario.	3	
Sono stati valutati i rischi di: Danno psicologico.	4	
Sono stati valutati i rischi di: Danno alla salute/sicurezza fisica.	4	
Sono stati valutati i rischi di: Danno alla proprietà intangibile (es. reputazione, IP).	3	
Sono stati valutati i rischi di: Impatto su sistemi sociali/politici (es. interferenze elettorali).	4	
Sono stati valutati i rischi di: Impatto su libertà civili (es. privacy, censura).	4	

*Fonte: Elaborazione degli autori*

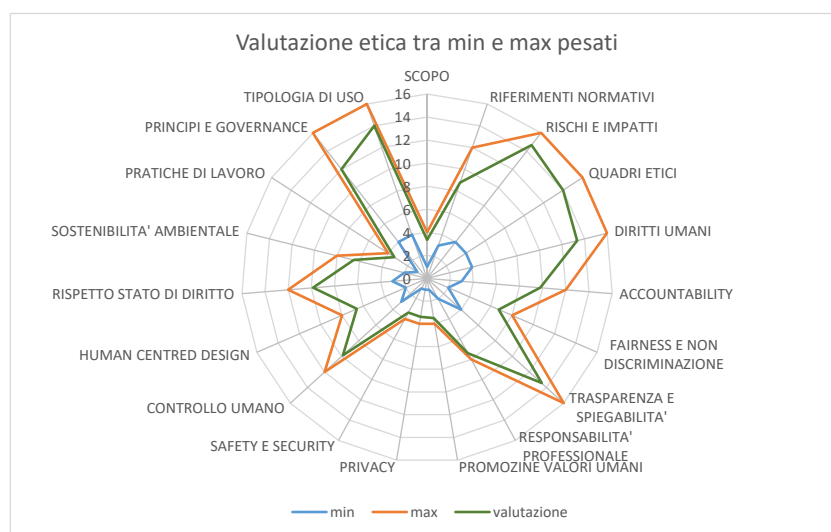
**Figura 2 - Classi con il loro risultato ed il relativo peso attribuito**



ITEM	e	w	e*w
SCOPO	3,33	1	3,333
RIFERIMENTI NORMATIVI	2,93	3	8,800
RISCHI E IMPATTI	3,67	4	14,667
QUADRI ETICI	3,50	4	14,000
DIRITTI UMANI	3,33	4	13,333
ACCOUNTABILITY	3,27	3	9,800
FAIRNESS E NON DISCRIMINAZIONE	3,38	2	6,750
TRASPARENZA E SPIEGABILITÀ	3,36	4	13,429
RESPONSABILITÀ PROFESSIONALE	3,70	2	7,400
PROMOZIONE VALORI UMANI	3,50	1	3,500
PRIVACY	3,38	1	3,385
SAFETY E SECURITY	3,38	1	3,375
CONTROLLO UMANO	3,29	3	9,857
HUMAN CENTRED DESIGN	3,30	2	6,600
RISPETTO STATO DI DIRITTO	3,29	3	9,857
SOSTENIBILITÀ AMBIENTALE	3,25	2	6,500
PRATICHE DI LAVORO	3,36	1	3,364
PRINCIPI E GOVERNANCE	3,00	4	12,000
TIPOLOGIA DI USO	3,50	4	14,000

Fonte: Elaborazione degli autori

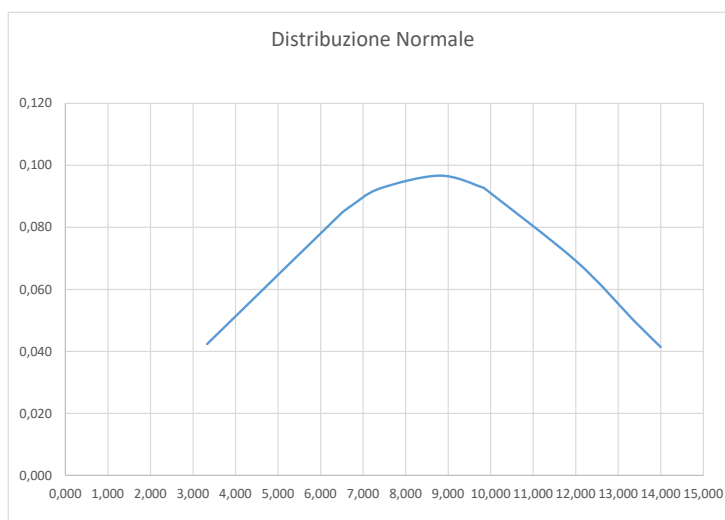
Figura 3 – Radar di valutazione etica



Fonte: Elaborazione degli autori

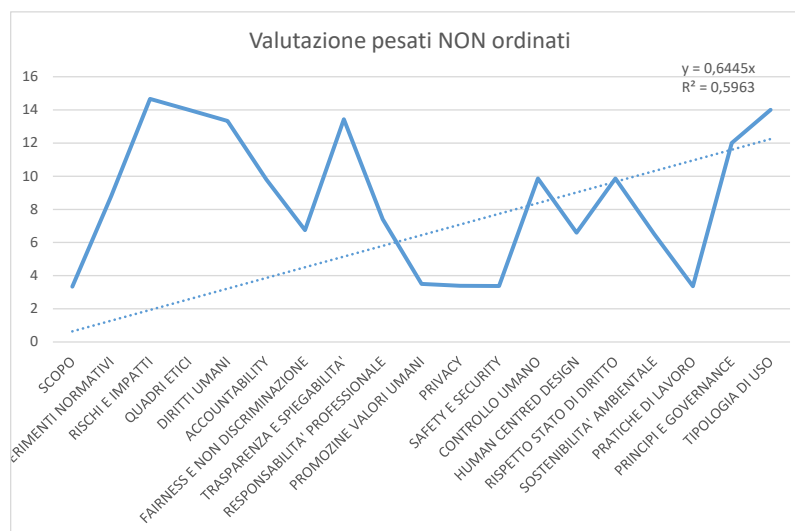


**Figura 4 - Distribuzione normale dei risultati**



Fonte: Elaborazione degli autori

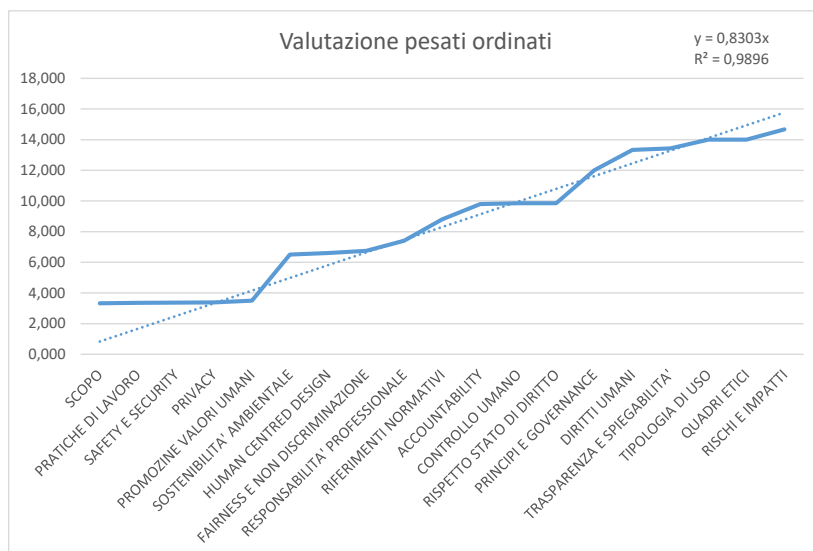
**Figura 5 - Valutazione pesati non ordinati**



Fonte: Elaborazione degli autori

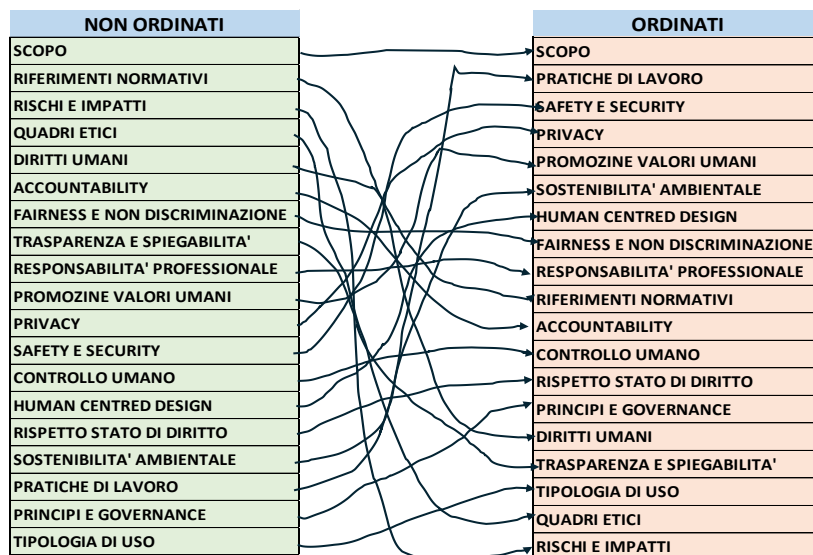


**Figura 6 - Valutazione pesati ordinati**



Fonte: Elaborazione degli autori

**Figura 7 - Valutazione dati non ordinati rispetto a quelli ordinati**



Fonte: Elaborazione degli autori



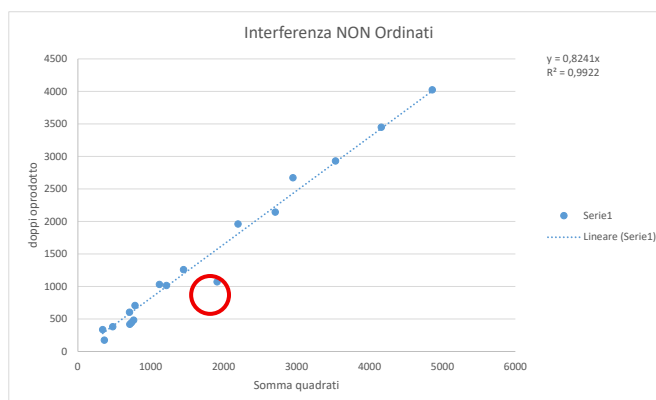
**Figura 8 - Valutazione dati non ordinati rispetto a quelli ordinati**

NON ORDINATI		ORDINATI
SCOPO	↔	SCOPO
RIFERIMENTI NORMATIVI	↓	PRATICHE DI LAVORO
RISCHI E IMPATTI	↓	SAFETY E SECURITY
QUADRI ETICI	↓	PRIVACY
DIRITTI UMANI	↓	PROMOZINE VALORI UMANI
ACCOUNTABILITY	↓	SOSTENIBILITA' AMBIENTALE
FAIRNESS E NON DISCRIMINAZIONE	↓	HUMAN CENTRED DESIGN
TRASPARENZA E SPIEGABILITA'	↓	FAIRNESS E NON DISCRIMINAZIONE
RESPONSABILITA' PROFESSIONALE	↔	RESPONSABILITA' PROFESSIONALE
PROMOZINE VALORI UMANI	↑	RIFERIMENTI NORMATIVI
PRIVACY	↑	ACCOUNTABILITY
SAFETY E SECURITY	↑	CONTROLLO UMANO
CONTROLLO UMANO	↑	RISPETTO STATO DI DIRITTO
HUMAN CENTRED DESIGN	↑	PRINCIPI E GOVERNANCE
RISPETTO STATO DI DIRITTO	↑	DIRITTI UMANI
SOSTENIBILITA' AMBIENTALE	↑	TRASPARENZA E SPIEGABILITA'
PRATICHE DI LAVORO	↑	TIPOLOGIA DI USO
PRINCIPI E GOVERNANCE	↑	QUADRI ETICI
TIPOLOGIA DI USO	↑	RISCHI E IMPATTI

Fonte: Elaborazione degli autori

Un'ulteriore analisi è stata quella di considerare le interferenze (calcolate come quadrato di un binomio, valori assoluti e doppio prodotto) reciproche tra le varie domande, questo ha evidenziato come alcuni valori (classi) risultino "spuri" ovvero da attenzionare. Si può notare dai grafici (Fig. 9, 10, 11 e 12) che questi valori emergono in maniera più evidente quando i dati non sono stati ordinati rispetto ai loro valori. I grafici sono "duplicati" in quanto sono stati invertiti gli assi per una maggior chiarezza.

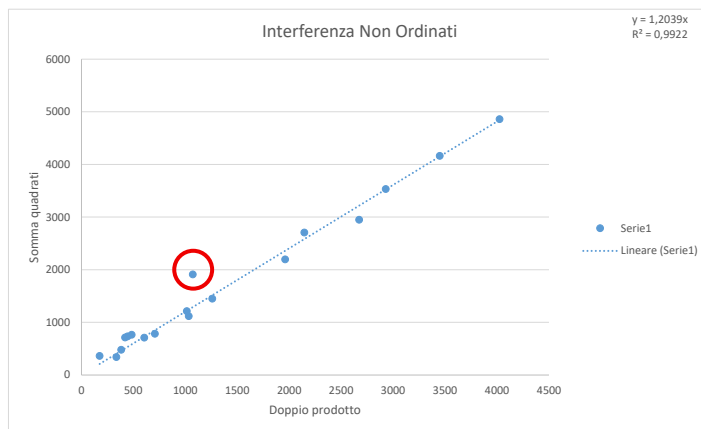
**Figure 9 - Interferenza non ordinati**



Fonte: Elaborazione degli autori

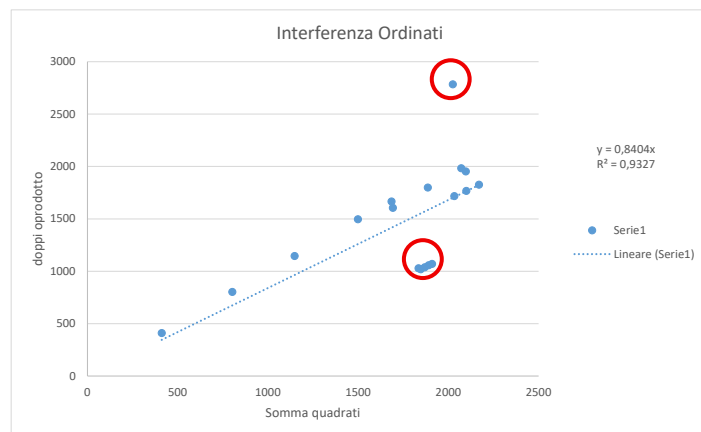


**Figure 10 - Interferenza non ordinati**



*Fonte: Elaborazione degli autori*

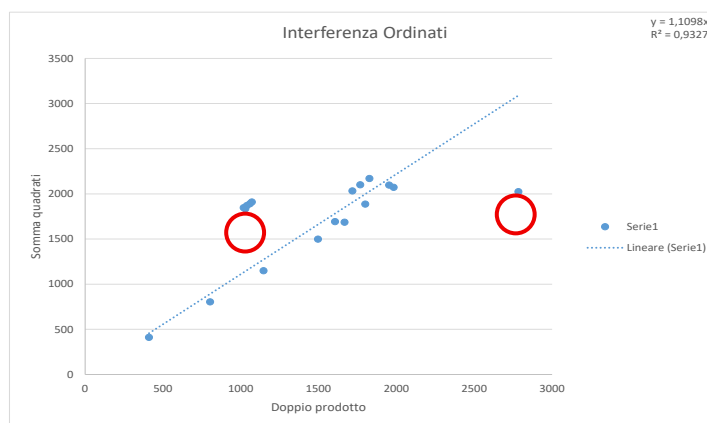
**Figure 11 - Interferenza ordinati**



*Fonte: Elaborazione degli autori*



**Figure 12 - Interferenza ordinati**

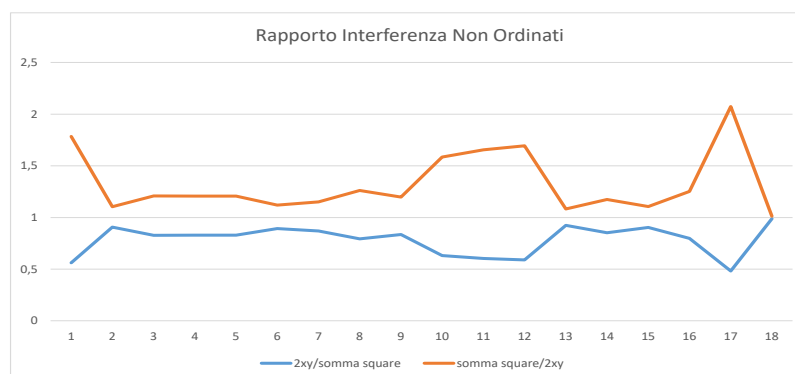


*Fonte: Elaborazione degli autori*

Questi punti riguardano, nel caso dei Non Ordinati la classe relativa allo scopo, che interagisce con le altre classi, mentre nel caso dei dati Ordinati riguardano le classi Scopo, Riferimenti Normativi, Rischi e Impatti e Responsabilità Professionale. Ciò permette di attivare dei piani d'azione considerando che ogni azione intrapresa impatterà sulle altre classi individuate.

Sono stati realizzati due grafici (Fig. 13 e 14) che riportano il rapporto tra l'interferenza (doppio prodotto) e la somma dei valori assoluti, anche in questo caso i grafici sono due invertendo gli assi.

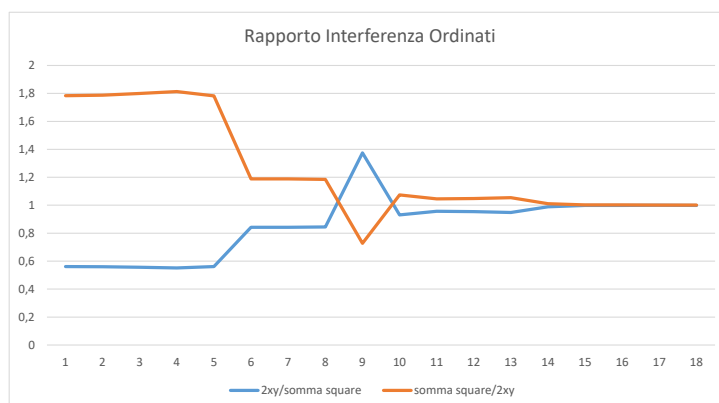
**Figura 13 - Rapporto interferenza non ordinati**



*Fonte: Elaborazione degli autori*



**Figura 14 - Rapporto interferenza ordinati**



Fonte: Elaborazione degli autori

Come si può notare i grafici sono praticamente speculari. I punti di picco sono quelli dove l'interferenza risulta maggiore, come anche illustrato nelle tabelle riportate sopra. L'implementazione di un sistema di gestione dell'etica, con i relativi controlli, permette alle organizzazioni di verificare la propria *governance* e l'implementazione dei sistemi IA garantendo il rispetto dei principi e dei diritti; attraverso un sistema di autovalutazione, è quindi possibile effettuare la fotografia dello stato di implementazione dell'etica nell'azienda dando informazioni sulle aree di miglioramento e fornendo strumenti al management per indirizzare le risorse e completare i gap esistenti.

#### **4. Dall'autovalutazione etica alla capacità di governo dell'intelligenza artificiale**

L'esercizio di autovalutazione etica presentato in questo lavoro non va letto come una mera applicazione tecnica di un questionario, né come un adempimento formale orientato alla conformità. Al contrario, esso rappresenta una dimostrazione concreta di come i principi etici possano essere tradotti in strumenti operativi capaci di generare conoscenza organizzativa e supportare decisioni informate.



Attraverso la strutturazione delle classi di principi, la ponderazione contestuale dei pesi e l'analisi dei risultati, l'autovalutazione consente di restituire una fotografia articolata dello stato di maturità etica di un'organizzazione rispetto all'uso dell'intelligenza artificiale. Ciò che emerge non è un giudizio binario, ma una rappresentazione differenziata delle aree di maggiore e minore presidio, che riflette scelte, priorità e sensibilità spesso implicite nei processi decisionali quotidiani.

Uno degli elementi di maggiore rilevanza del modello proposto risiede nella sua capacità di rendere visibili le asimmetrie. Le differenze tra classi maggiormente presidiate e classi più fragili – come quelle relative a privacy, sicurezza, responsabilità o pratiche di lavoro – non indicano semplicemente lacune, ma raccontano il modo in cui l'organizzazione ha storicamente distribuito attenzione, risorse e responsabilità. In questo senso, l'autovalutazione etica agisce come uno strumento di riflessione critica, capace di far emergere disallineamenti tra intenzioni dichiarate e pratiche effettive. Particolarmente significativo è anche il contributo dell'analisi delle interferenze tra le classi di principi. L'evidenza che alcune dimensioni come lo scopo del sistema, i riferimenti normativi, la gestione dei rischi e la responsabilità professionale, interagiscono in modo sistemico con le altre conferma una intuizione spesso sottovalutata: nell'intelligenza artificiale, le scelte non sono mai isolate. Ogni intervento su una dimensione etica produce effetti su più livelli della governance, rendendo necessario un approccio integrato e non settoriale. Da questa prospettiva, l'autovalutazione non fornisce soltanto un'indicazione su *dove* intervenire, ma aiuta a comprendere *come* intervenire, evidenziando le interdipendenze che devono essere considerate nella definizione dei piani di azione. È qui che il valore del modello supera la dimensione diagnostica e assume una funzione strategica, offrendo al management e ai board una base strutturata per orientare le decisioni, allocare le risorse e rafforzare i presidi di governance.

Il lavoro svolto dimostra, inoltre, come l'etica dell'intelligenza artificiale possa essere integrata nei sistemi di controllo interno e di gestione del rischio senza essere semplificata o ridotta a checklist superficiali. Al contrario, la complessità del modello



riflette la complessità dei sistemi che intende governare. La possibilità di analizzare i risultati in forma aggregata, di confrontare scenari e di monitorare l'evoluzione nel tempo rende l'autovalutazione uno strumento coerente con una concezione dinamica del rischio e della responsabilità. In definitiva, l'autovalutazione etica proposta in questo paper mostra come sia possibile passare dall'etica dichiarata all'etica governata. Non si tratta di stabilire una volta per tutte se un sistema di intelligenza artificiale sia "etico", ma di dotarsi di strumenti che consentano di interrogare continuamente le scelte tecnologiche alla luce dei loro impatti, dei contesti di utilizzo e delle aspettative della società.

In un tempo in cui l'intelligenza artificiale è sempre più pervasiva e decisiva, la capacità di misurare, discutere e governare l'etica diventa parte integrante della responsabilità di chi guida le organizzazioni. L'autovalutazione etica, così concepita, non è un punto di arrivo, ma un dispositivo abilitante come spazio strutturato di consapevolezza che permette alle organizzazioni di apprendere, correggere e migliorare il proprio rapporto con l'intelligenza artificiale, rendendola una tecnologia non solo potente, ma anche affidabile, sostenibile e socialmente legittima.