Rischi e opportunità dell'IA per migliorare equità e accessibilità

Report sessione di discussione gruppo 2

7 ottobre 2025

Partecipanti presenti alla sessione

1. Gianpaolo Sellitto, ANAC
2. Glenda Gentili, AgID
3. Sandro Libianchi, Medico Co.N.O.S.C.I
4.

Formez/Facilitazione

1. Fabiola De Toffol, IAF
2. Giovanni Allegretti, Formez
3. Francesca De Chiara, Formez

Sviluppo della discussione

La facilitatrice introduce la sessione e dà lettura della domanda:

Come possiamo assicurare equità nelle fasi di addestramento di un sistema di IA, al fine di ridurre discriminazioni e stereotipi?

Ciascun partecipante è invitato a contribuire alla discussione.

Segue una restituzione sintetica dei temi trasversali emersi dal confronto con attenzione alle diverse percezioni/priorità presentate, senza attribuzioni (ovvero senza riportare "chi ha detto cosa"). Laddove possibile sono segnalate le proposte che verificano un ampio livello di convergenza.

Il gruppo è composto da esperti che hanno condiviso considerazioni sul tema da esplorare. Sono state sottolineate le criticità derivanti dall'utilizzo di modelli nel "mondo reale" anche in assenza di condivisione e chiarezza sulla qualità dell'addestramento (es. diagnostica medica).

La discussione ha affrontato i seguenti temi chiave:

Trasparenza e comunicazione - È emersa la necessità di rendere più chiaro e comprensibile l'uso dell'intelligenza artificiale, in particolare nelle fasi di addestramento dei modelli. Oggi, soprattutto in ambito sanitario, capita che i referti indichino genericamente l'impiego dell'IA, senza precisare in che modo il sistema sia stato addestrato o in quale fase di utilizzo si trovi. Questo crea un problema di affidabilità: un modello appena avviato può produrre risultati meno solidi rispetto a uno consolidato, ma questa differenza non viene comunicata né ai professionisti né ai cittadini. Per affrontare questa opacità, si è discusso della necessità di introdurre obblighi chiari di dichiarazione:

chi utilizza sistemi di IA dovrebbe specificare in modo trasparente non solo che la tecnologia è impiegata, ma anche con quali dataset è stata addestrata, come viene aggiornata e qual è il livello di affidabilità raggiunto.
Per rendere comprensibile questa complessità anche ai non esperti, si è suggerito di sviluppare un sistema di classificazione della maturità dei modelli, una sorta di "scoring" con livelli (A, B, C, D). Ciò permetterebbe di distinguere, ad esempio, un sistema addestrato su milioni di dati e validato con criteri rigorosi da un modello in fase iniziale, con dataset limitati. Uno strumento del genere avrebbe un valore comunicativo importante, perché tradurrebbe parametri tecnici complessi in un linguaggio accessibile e utile per cittadini e professionisti.
Accanto a questi strumenti concreti, è stato evidenziato un problema culturale e comunicativo: oggi prevale un discorso fortemente promozionale sull'IA, che ne mette in luce solo i successi e raramente i rischi o i limiti. Questo contribuisce a creare un'illusione di perfezione, riducendo il senso critico e la consapevolezza del pubblico. In realtà, come ricordato, nessun sistema è privo di difetti: la tecnologia è in continua evoluzione e le sue criticità devono essere dichiarate con la stessa chiarezza dei progressi.

Partecipazione della società civile - Un secondo filone di discussione ha riguardato il ruolo della società civile nei processi di addestramento dell'intelligenza artificiale. La domanda centrale è stata: come garantire che cittadini, organizzazioni e attori non istituzionali possano avere voce in capitolo su un tema tradizionalmente dominato da sviluppatori, aziende e istituzioni? È emerso come, secondo alcuni, al momento la società civile venga coinvolta soprattutto come destinataria di informazioni, in una posizione passiva. Altri interventi hanno invece proposto di immaginare forme di partecipazione attiva. Ciò significherebbe, ad esempio: contribuire alla scelta dei temi e delle priorità su cui addestrare un sistema; partecipare alla definizione delle categorie logiche e delle fonti di apprendimento utilizzate dai modelli; avere accesso a strumenti di valutazione semplificati (come gli scoring di maturità dei modelli) per poter esercitare un controllo diffuso. Questa prospettiva nasce dall'idea che i sistemi di IA non debbano essere addestrati soltanto con dati tecnici, ma anche con documenti e pratiche di contesto che riflettano valori sociali. Tra le proposte operative: istituire punti di contatto pubblici, attraverso i quali i cittadini possano segnalare malfunzionamenti o presentare reclami rispetto ai risultati prodotti dall'IA; abilitare piattaforme di consultazione aperta sui criteri di addestramento, simili a quelle già sperimentate in progetti europei come Aequitas.

Bias e distorsioni - È stato inevitabile il richiamo ai bias che si annidano nei dataset di addestramento, ricordando che molti squilibri presenti nei dati – ad esempio la predominanza maschile in alcune professioni o la sottorappresentazione di minoranze – per lungo tempo sono stati considerati "normali" e quindi non percepiti come problematici. Solo l'evoluzione culturale e sociale ha portato a riconoscerli come distorsioni. Questo significa che il riconoscimento dei bias non è neutro, ma dipende dal contesto storico e culturale. Un ulteriore elemento critico riguarda i rischi di sovradiagnosi in ambito sanitario. Sistemi molto sensibili possono individuare anomalie irrilevanti, innescando catene di esami inutili, con conseguenti costi economici e psicologici. È stato

sottolineato che in questi casi resta centrale il principio dello "Human in the Loop": l'IA può evidenziare dati, ma la responsabilità finale deve rimanere al professionista umano, che interpreta e decide. Tuttavia, decidere di trascurare gli output di una macchina può comportare rischi professionali. Un'altra criticità è rappresentata dal testing nel contesto reale. Spesso ci si accorge dell'affidabilità di un modello solo una volta messo in uso, quando produce decisioni o diagnosi reali. Questo processo espone però a rischi significativi: i cittadini diventano inconsapevolmente "tester" di sistemi non ancora maturi. Un ulteriore problema riguarda i modelli che si auto-modificano nel tempo, adattandosi continuamente ai nuovi input. In questi casi si perde la tracciabilità del dataset iniziale e diventa difficile garantire stabilità e coerenza del comportamento. Questo è particolarmente problematico nei modelli general purpose. Infine, è stata richiamata l'attenzione sull'ambiguità d'uso di certi strumenti: sistemi progettati con finalità positive (ad esempio per mappare reti IoT - Internet of Things - e aumentarne la sicurezza) possono essere riutilizzati in modo malevolo, diventando strumenti per identificare vulnerabilità da sfruttare.

Governance e responsabilità - È importante investire nella governance perché, senza un quadro chiaro, si rischia di alimentare dinamiche di autocensura e deresponsabilizzazione. Ad esempio, in ambito sanitario, un radiologo che nutra dubbi su un referto prodotto con il supporto di un algoritmo potrebbe esitare a segnalare l'anomalia, temendo conseguenze legali o reputazionali. Se invece esistesse una struttura di governance trasparente, con regole chiare e spazi sicuri di confronto, sarebbe possibile riportare dubbi ed errori senza che ciò diventi un attacco alla persona. È emersa la proposta di istituire figure o organismi specifici, come un responsabile etico per l'IA, paragonabile al titolare del trattamento dati previsto dal GDPR. Tuttavia, diversi partecipanti hanno evidenziato che una sola persona difficilmente può concentrare competenze tecniche, giuridiche ed etiche. Da qui l'idea di ricorrere a comitati etici interdisciplinari, che riuniscono professionalità diverse e garantiscono una valutazione più equilibrata e completa. La governance riguarda anche i fornitori di soluzioni IA, che hanno un ruolo decisivo. Sono loro a progettare i sistemi, a definire le "chiavi" di funzionamento e spesso a promuoverli sul mercato come strumenti già pronti e affidabili. Questa dinamica commerciale, però, rischia di privilegiare il marketing rispetto alla trasparenza. Senza regole vincolanti, i fornitori possono enfatizzare solo gli aspetti positivi, omettendo i limiti e i rischi dei loro prodotti. Infine, è stato sottolineato come la governance non sia solo un tema tecnico o regolatorio, ma riguardi anche la fiducia sociale.

Formazione e cultura - In ambito formativo, oggi prevalgono corsi legati ai prodotti e ai sistemi commerciali: si insegna a utilizzare strumenti specifici, ma raramente si affrontano le implicazioni etiche, sociali e culturali dell'addestramento dei modelli. Questo approccio rischia di ridurre l'IA a una semplice competenza tecnica, senza sviluppare la capacità critica necessaria per governarla. È stata quindi sottolineata la necessità di una formazione interdisciplinare, che metta in dialogo competenze tecniche, etiche, giuridiche e sociali. Tuttavia, è emersa anche la difficoltà di concentrare tutte queste competenze in una sola figura professionale. Da qui il suggerimento di ispirarsi al modello dei comitati etici

pluridisciplinari, già adottati in altri campi (ad esempio nella ricerca biomedica), evitando di attribuire a un unico soggetto la responsabilità di valutare la correttezza dell'addestramento e dell'uso dei sistemi. Il tema formativo si intreccia con quello della consapevolezza dei decisori: chi prende decisioni deve avere una visione più ampia, per non rischiare di adottare tecnologie senza valutare le conseguenze. Tale aspetto si lega anche alla dimensione educativa e culturale diffusa. L'IA non riguarda solo gli esperti, ma tocca la vita quotidiana di tutti. Per questo è necessario promuovere una vera e propria alfabetizzazione critica della società, capace di formare cittadini consapevoli dei rischi e delle potenzialità, in grado di esercitare un controllo sulle modalità di funzionamento delle tecnologie che li riguardano.

Normativa e decreti attuativi - In merito alla recente legge delega (la Legge n. 132 del 23 settembre 2025), e alla necessità di preparare contenuti da includere nei decreti attuativi, è stata proposta la creazione di un glossario normativo condiviso. Le parole chiave dell'IA – come "trasparenza", "bias", "equità" – assumono significati diversi a seconda di chi le utilizza. Stabilire definizioni chiare e comuni, con valore normativo, aiuterebbe a ridurre ambiguità interpretative e garantirebbe maggiore coerenza nell'applicazione delle regole. Questo approccio è già stato adottato in altri provvedimenti legislativi e potrebbe diventare un valore aggiunto della normativa. È stato sottolineato come le raccomandazioni che saranno prodotte nel percorso partecipativo possano contribuire a orientare la scrittura dei decreti attuativi, fornendo spunti concreti che tengano conto non solo delle esigenze istituzionali, ma anche delle prospettive della società civile e dei professionisti dei settori più coinvolti. La discussione ha inoltre richiamato la necessità di collegare il dibattito normativo alla mobilitazione per i diritti digitali. La legge non può limitarsi a imporre regole dall'alto, ma deve favorire un percorso di crescita culturale e linguistica che consenta a cittadini e organizzazioni di comprendere e gestire le tecnologie in modo critico. In questo senso, la normativa dovrebbe essere accompagnata da strumenti di educazione, sensibilizzazione e comunicazione trasparente.

Conclusioni ed esiti della sessione

In conclusione, le convergenze principali riguardano le seguenti proposte:

- **trasparenza nell'uso dell'IA,** rendendo obbligatoria la dichiarazione dell'uso dell'IA nei documenti ufficiali (es. referti sanitari), specificando la fase di maturità del modello e i criteri di addestramento utilizzati, creando sistemi di classificazione (scoring A, B, C, D) per comunicare in modo semplice il livello di affidabilità dei modelli, contrastando la comunicazione solo promozionale, valorizzando anche rischi e limiti;
- **partecipazione della società civile,** attraverso l'istituzione di punti di contatto pubblici per segnalazioni e reclami sui sistemi IA, l'attivazione di piattaforme di consultazione aperta per coinvolgere cittadini e organizzazioni nelle scelte sull'addestramento, il coinvolgimento della società civile peri contribuire alla definizione dei temi, delle categorie e delle fonti di apprendimento dei modelli;

- **gestione dei bias e delle distorsioni,** con il monitoraggio e la correzione dei bias impliciti nei dataset, riconoscendo il loro legame con i contesti storici e culturali, applicando il principio dello *Human in the Loop* e creando banche dati pubbliche e condivise per raccogliere esperienze e favorire il confronto tra modelli;
- **governance e responsabilità,** istituendo figure di responsabile etico per l'IA o, meglio, comitati interdisciplinari che garantiscano valutazioni equilibrate, con spazi sicuri di confronto che permettano a professionisti e operatori di segnalare errori senza conseguenze punitive e regole vincolanti anche per i fornitori;
- **formazione e cultura,** con il superamento di una formazione esclusivamente tecnica o di prodotto, per sviluppare percorsi interdisciplinari che integrino aspetti etici, giuridici e sociali e con percorsi di alfabetizzazione critica sull'IA rivolti non solo a esperti, ma anche a cittadini e decisori;
- **normativa e decreti attuativi,** con l'aiuto di un glossario normativo condiviso, capace di ridurre le ambiguità terminologiche e integrando le raccomandazioni che saranno state elaborate all'interno dei testi normativi.