Sesto Piano d'Azione nazionale per il governo aperto 2024 - 2026

Obiettivo B Accompagnare la diffusione e l'innovazione delle politiche di apertura a tutti i livelli di governo

Impegno 5 - Promozione dell'inclusività e dei diritti nell'accesso alle tecnologie e nell'utilizzo dell'Intelligenza Artificiale

Rischi e opportunità dell'IA per migliorare equità e accessibilità

Report sessione di discussione gruppo 1

07 ottobre 2025

Partecipanti presenti alla sessione

1. Laura Ferrari, Rete per i diritti umani e digitali
2. Nunzia Imperato, Commissione opportunità Lega Coop Emilia-Romagna
3. Maria Letizia Mancinelli, Ministero della Cultura - Istituto Centrale per il Catalogo e la Documentazione
4. Claudia Mazzanti, ActionAid Italia
5. Antonella Negri, Ministero della Cultura - Istituto Centrale per la Digitalizzazione del Patrimonio Culturale
6. Elena Ruga, Comune di Tremezzina (CO) - Associazione Comunicatori Pubblici
7. Valentina De Riz (silenziosa)

Staff Formez/Facilitazione

1. Fabio Riva, facilitatore IAF
2. Flavia Marzano, staff Formez

Sviluppo della discussione

Il facilitatore 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 ha esplorato il tema dell'addestramento dei sistemi di IA da prospettive complementari: chi lavora direttamente con dati culturali, chi si occupa di diritti e advocacy, chi opera nella comunicazione pubblica, chi coordina progetti di ricerca sull'inclusività. Questa diversità ha permesso di identificare cinque aree critiche interconnesse.

Trasparenza e certificazione dei dati di addestramento - La qualità e la provenienza dei dati utilizzati per l'addestramento sono un prerequisito fondamentale per garantire equità. È necessario rendere trasparenti: l'origine dei dataset, le modalità di raccolta, i settori della popolazione coinvolti, i criteri di selezione. Questa tracciabilità aiuta a comprendere quali scelte stanno alla base dell'addestramento e quali bias potrebbero essere stati introdotti sin dall'inizio.
Nel settore dei beni culturali sono emerse esperienze concrete di addestramento su dataset specializzati e certificati. Per classificare automaticamente immagini del patrimonio culturale non si possono utilizzare banche dati generiche di internet: servono dati pertinenti al dominio specifico, con competenze esperte che guidino il riconoscimento di elementi particolari (aureole, iconografie sacre, stili artistici). Questo approccio garantisce risultati più accurati e culturalmente appropriati. I dati devono essere documentati, tracciati nelle loro versioni successive, e accompagnati da metadati completi che ne descrivano caratteristiche e limiti.
La nuova normativa italiana (legge 132/2025) introduce obblighi di trasparenza per i fornitori della PA sulle modalità di addestramento, richiedendo che possano spiegare procedure, fonti dei dati, controlli di qualità. Questo rappresenta un passo avanti importante verso standard più rigorosi.

Addestramento responsabile e correzione delle distorsioni - L'addestramento di un algoritmo non è un processo statico ma iterativo, che richiede continui aggiustamenti. L'algoritmo "impara" dagli esempi forniti e la qualità, completezza e rappresentatività dei dati influenzano direttamente i risultati. Dataset distorti, parziali o sbilanciati generano modelli con bias e scarsa capacità di generalizzazione.
Ci sono esperienze concrete di collaborazione tra tecnici ed esperti di dominio: un processo di rodaggio iniziale permette ai tecnici di comprendere il contesto applicativo e le aspettative, mentre gli esperti di dominio vengono gradualmente introdotti agli aspetti tecnici. Questo lavoro interdisciplinare, fatto di prove, test e correzioni successive, porta a risultati soddisfacenti. I modelli vengono via via aggiornati in base alle correzioni applicate e ai dati forniti.
Importante è distinguere tra informazioni prodotte da esperti e informazioni generate dall'IA: l'arricchimento automatico dei dati tramite IA deve essere sempre marcato e distinto dalle descrizioni fatte da catalogatori umani. Questo permette trasparenza e consente valutazioni critiche sui risultati.
Sul piano della formazione, si evidenzia la necessità di percorsi educativi specifici per un addestramento responsabile ed etico. La correzione delle distorsioni richiede inoltre **valutazioni eterogenee:** i risultati non dovrebbero essere testati solo dal gruppo di sviluppo, ma da utenti con formazioni e prospettive diverse, per cogliere problematiche non immediatamente visibili.

Governance condivisa e strutture di controllo – Ci vorrebbe una governance multilivello: a livello centrale, un'autorità con funzioni di vigilanza simili a quelle del Garante della

Privacy; a livello capillare, figure responsabili in ciascuna organizzazione che adotta sistemi di IA, con meccanismi aperti per raccogliere segnalazioni da parte degli utenti.
La governance dovrebbe prevedere comitati etici e tecnici con competenze multidisciplinari, in grado di integrare aspetti etici, tecnici e di dominio specifico, supervisionare l'addestramento, la provenienza dei dati, e le logiche decisionali degli algoritmi.
A livello operativo, servirebbe un responsabile per la trasparenza dei dati - presso ciascuna amministrazione o organizzazione - facilmente raggiungibile per segnalazioni immediate, con un meccanismo di segnalazione sempre aperto, quindi anche in ottica di accountability, con la possibilità di accesso da parte di ogni singolo utente.
Fondamentale è anche il meccanismo di feedback: non basta ricevere segnalazioni, occorre rispondere all'utente spiegando come è stata gestita la problematica. Questo circolo virtuoso aumenta la consapevolezza collettiva e la fiducia nei sistemi.
La governance deve inoltre essere dinamica e aggiornata costantemente, data la velocità di evoluzione della tecnologia.
Molto importante è il coinvolgimento della società civile nei processi di valutazione. Test con utenti eterogenei, consultazioni pubbliche, e l'integrazione di organizzazioni della società civile nelle fasi di verifica sono visti come strumenti efficaci per garantire che i sistemi rispondano a valori di equità e diritti fondamentali.

Partecipazione attiva e consapevolezza diffusa - Un tema trasversale riguarda il ruolo attivo che ciascuno può e deve avere nell'addestrare i sistemi di IA. Gli algoritmi imparano da ciò che viene loro insegnato: correggere bias linguistici o segnalare risposte inadeguate contribuisce concretamente al miglioramento collettivo. Esempi concreti mostrano come interventi puntuali (segnalare traduzioni stereotipate, chiedere risposte più equilibrate) modifichino effettivamente il comportamento dei modelli.
Tuttavia, l'IA non è neutrale: riflette i dati che le vengono forniti, inclusi pregiudizi e stereotipi presenti nella società. La differenza cruciale rispetto agli esseri umani è che l'IA non ha pregiudizi emotivi: cataloga e generalizza senza discriminare intenzionalmente. Questo rappresenta sia un limite (difficoltà a cogliere sfumature contestuali) sia un vantaggio (assenza di manipolazioni emotive deliberate). Resta però fondamentale che gli utenti consapevoli sappiano filtrare criticamente le risposte, mentre gli utenti inconsapevoli devono essere tutelati attraverso formazione, alfabetizzazione e strumenti di governance efficaci.
Sul piano della comunicazione, emergono sfide importanti: il divario linguistico tra tecnici specialisti e utenti finali va colmato attraverso figure di mediazione **e** vocabolari condivisi. La comunicazione diventa ancora più complessa quando l'IA viene utilizzata per generare contenuti informativi: serve vigilanza per evitare disinformazione, bolle informative e fake news. L'utilizzo dell'IA nella comunicazione pubblica richiede quindi standard elevati di trasparenza e qualità.
La formazione diffusa è vista come condizione necessaria: inserire moduli su IA e algoritmi nei percorsi educativi (non solo informatici ma anche in ingegneria e altre discipline) contribuisce a creare cittadini capaci di pensiero critico e decisioni informate. Anche i

comunicatori pubblici e gli operatori della PA necessitano di competenze aggiornate per gestire consapevolmente questi strumenti.

L'implementazione dei meccanismi di ricorso senza giudice (redress mechanism) - Si tratta diun tema importante previsto dagli articoli 85 e 86 dell'AI Act europeo. Questi strumenti di ricorso permetterebbero ai cittadini di contestare violazioni dei propri diritti causate da sistemi di IA senza dover affrontare procedure giudiziarie complesse e costose. La normativa italiana recente (legge delega) non ha ancora dettagliato questi aspetti, lasciando aperta la questione di quale autorità dovrebbe gestire tali ricorsi: se l'Agenzia per la Cybersicurezza Nazionale (ACN), l'AGID o altra. Un modello potrebbe essere il funzionamento del Garante della Privacy.
Questi meccanismi rappresentano una garanzia soprattutto per le categorie più fragili, che difficilmente si rivolgerebbero a un giudice. Serve una spinta da parte della società civile affinché l'Italia implementi tempestivamente procedure chiare, accessibili e con un'autorità competente ben definita per poter identificare chiaramente chi risponde delle decisioni prese dai sistemi di IA (accountability) e attraverso quali canali i cittadini possono far valere i propri diritti.

Conclusioni ed esiti della sessione

In conclusione, le convergenze principali riguardano le seguenti proposte:

- **garantire piena trasparenza e tracciabilità dei dati di addestramento**: documentare origine, modalità di raccolta, criteri di selezione, l'aggiornamento dei modelli. I fornitori di sistemi IA per la PA devono essere in grado di spiegare le proprie procedure di addestramento.
- **promuovere collaborazione interdisciplinare nell'addestramento**: tecnici ed esperti di dominio devono lavorare insieme in processi iterativi di test, valutazione e correzione. Distinguere chiaramente output generati dall'IA da contenuti prodotti da esperti umani.
- **istituire una governance multilivello** con chiari principi di accountability: autorità centrale di vigilanza (sul modello del Garante Privacy), comitati etici e tecnici multidisciplinari, responsabili per la trasparenza in ciascuna organizzazione, meccanismi aperti di segnalazione e feedback strutturati.
- **diffondere formazione e consapevolezza**: educazione su IA e algoritmi nei percorsi scolastici e professionali; alfabetizzazione digitale per cittadini e operatori pubblici; coinvolgimento attivo della società civile nei processi di valutazione e controllo.
- **implementare rapidamente i meccanismi di ricorso senza giudice** previsti dall'AI Act europeo, individuando chiaramente l'autorità competente e garantendo procedure accessibili soprattutto per le categorie fragili.