Rischi e opportunità dell'IA per migliorare equità e accessibilità

Report sessione di discussione gruppo 4

7 ottobre 2025

Partecipanti presenti alla sessione

1. Leda Guidi, Compubblica
2. Martin Enrico Iglesias, Comune di Milano
3. Fabrizio Peresson, Confartigianato
4. Alessandro Selam, ANORC

Staff Formez/Facilitazione

1. Alessandra Pietropoli - Facilitatrice IAF

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.

Concetti di equità, bias e stereotipo - La discussione si apre con la considerazione che è molto importante chiarire il significato dei termini chiave — equità, bias, discriminazione, stereotipo — per evitare fraintendimenti terminologici e impostare un linguaggio comune. È emersa la necessità di una base formativa condivisa, anche di tipo filosofico ed etico, che consenta di riconoscere le forme di discriminazione esplicite e implicite già nella fase di progettazione e addestramento dei modelli. Questa consapevolezza è considerata la premessa per qualsiasi azione di prevenzione delle distorsioni. È importante mettere dei paletti di sicurezza e tornare proprio all'origine della domanda: "cosa sono le discriminazioni e gli stereotipi?". I bias devono essere considerati specifici al dominio applicativo: non esiste un'unica regola universale. A quel punto anche i bias sono circoscritti e si verificano più facilmente all'interno di quel servizio.

Formazione e competenze - Un tema trasversale e ricorrente riguarda la formazione, ritenuta fondamentale per ogni livello della filiera dell'intelligenza artificiale e condizione preliminare per ogni azione efficace sull'IA. Si sottolinea l'importanza di una formazione differenziata (verticalizzata) a seconda dei ruoli: sviluppatori, dipendenti pubblici, decisori politici, aziende e cittadini devono acquisire competenze specifiche coerenti con le proprie

responsabilità e anche con i propri obiettivi. L'azienda che sviluppa il sistema non ha gli stessi obiettivi che ha il dipendente della PA o che ha il cittadino. La formazione alla quale si fa riferimento non è solo tecnica, ma anche culturale, finalizzata alla comprensione dei principi di equità, trasparenza e responsabilità. Si tratta di uno strumento di alfabetizzazione collettiva alla cultura dell'IA, che deve includere anche i giovani e i non addetti ai lavori. Se non si fa questo come prima cosa non si riuscirà a fare il resto. Ci sono già linee guida che danno indicazioni in questo senso, ma tutto questo deve sedimentare e diventare una pratica comune. La formazione dei soggetti preposti al controllo è fondamentale: solo chi è adeguatamente formato può riconoscere e correggere un bias in modo efficace.

Differenze tra IA generativa e non generativa - Il gruppo distingue tra IA generativa e IA non generativa, evidenziando che i processi di addestramento e i rischi connessi ai bias sono diversi. L'IA generativa non punta a fornire risposte vere, ma "probabili" e questo la rende più vulnerabile a fenomeni di allucinazione o distorsione. Si osserva che, nella PA, la maggior parte dei sistemi attualmente in uso appartiene ancora alla categoria non generativa, che non deve essere dimenticata in favore delle tecnologie più recenti. Il livello di spiegabilità e attendibilità dei modelli deve essere calibrato in base al contesto applicativo: in ambiti sensibili (es. medico o amministrativo) è più importante che il sistema sia molto attendibile piuttosto che riuscire a spiegarne completamente il funzionamento. Anche per questa ragione è molto importante che tutta la IA non generativa non venga dimenticata.

Qualità, pulizia e rappresentatività dei dati - Il tema dei dati utilizzati per l'addestramento dell'IA è emerso come cruciale per assicurare equità. È stato ribadito che i dataset devono essere variegati, bilanciati e accuratamente puliti, rappresentando in modo equo i diversi gruppi sociali e culturali, validati con procedure *golden test* (ndr. un insieme di pratiche utilizzate per verificare che un sistema produca risultati coerenti con una versione di riferimento considerata "d'oro", cioè perfettamente corretta e validata). Sono stati citati esempi positivi, come quello della società informatica *in house* regionale del Friuli-Venezia Giulia, che gestisce tutti i dati sanitari della Regione e che, in collaborazione con il Garante della privacy, sta verificando le modalità per creare dei dati sintetici con attendibilità statistica. Questi dati possono essere forniti alle aziende interessate a sviluppare sistemi di IA in ambito sanitario, che possono essere addestrati conciliando tutela della privacy e qualità dei dati.

Trasparenza e documentazione dei processi di addestramento - Ampio spazio è stato dedicato alla trasparenza, considerata condizione essenziale per la fiducia pubblica nei sistemi di IA. Si è convenuto sull'importanza di documentare in modo chiaro le fasi di raccolta e trattamento dei dati, i criteri di selezione, gli algoritmi utilizzati e le modalità di verifica. È stata sottolineata la necessità di rendere tracciabili le scelte metodologiche e di mantenere un registro dei passaggi fondamentali dell'addestramento, accessibile ai soggetti coinvolti nella governance e al controllo civico. La trasparenza deve riguardare anche le gare d'appalto e i capitolati pubblici, che dovrebbero integrare requisiti tecnici e procedurali mirati alla riduzione dei bias e alla verifica periodica dei risultati. È utile chiedersi quali elementi di trasparenza nell'addestramento possiamo già garantire oggi:

infatti in un contesto di governo aperto, è importante immaginare che tipo di operazioni stanno dentro il quadro normativo che è stato approvato da poco.

Governance pubblica e meccanismi di controllo - Il gruppo riconosce il ruolo decisivo della PA nella definizione delle regole e delle procedure per l'uso dell'IA. La PA deve mantenere la responsabilità umana finale (Human in the Loop), garantendo che le decisioni restino sotto il controllo umano e che siano istituiti meccanismi di supervisione e correzione in caso di distorsioni. L'IA deve essere un sostegno all'attività umana, non deve sostituirla. Si fa riferimento alla figura dell'AI Officer e all'importanza di creare delle équipe multidisciplinari permanenti (composte da tecnici, giuristi, linguisti, esperti etici) incaricati di monitorare la correttezza dei sistemi e di valutare i rischi emergenti. È stato inoltre evidenziato che la governance deve includere processi partecipativi con il coinvolgimento della società civile e degli esperti di settore, per garantire trasparenza e controllo diffuso. La PA deve imporre regole chiare nei capitolati, garantendo trasparenza, auditabilità, e inclusione della società civile. Ci sono una serie di criteri che si possono richiedere come committenti, per esempio che gli algoritmi siano aperti, conoscibili, open source, riutilizzabili e rielaborabili; infatti non è detto che i provider condividano questi requisiti. Ci sono già delle linee guida condivise nella PA. L'IA si alimenta di dati e la PA ha ricchezza di dati. Le aziende che avranno accesso ai dati li sfrutteranno come da loro natura e sarà necessario mantenere attenzione alta sugli accordi che riguardano l'utilizzo dei dati. L'applicazione dell'IA nella Pubblica Amministrazione è infatti trasversale, ma la cosa più difficile da fare è mantenere allineate le funzioni con gruppi di lavoro trasversali. Questo anche se la PA funziona ancora per verticalità, che sono necessarie, ma sarà necessario sfidare anche questo aspetto.

Rischi legati alla localizzazione dei data center e alla sovranità dei dati - È emersa una preoccupazione rispetto alla localizzazione dei data center e alla dipendenza tecnologica da soggetti extraeuropei. Si teme la perdita di controllo sui dati e sull'uso che ne viene fatto al di fuori del contesto normativo europeo. Si propone pertanto di privilegiare fornitori e partner europei, promuovendo un modello di autonomia strategica e di sovranità digitale, in linea con i principi del futuro AI Act e delle normative europee sulla protezione dei dati.

Correzione delle distorsioni e tutela in caso di violazioni - Riguardo ai meccanismi di tutela, è stato sottolineato che le distorsioni devono essere rilevate e corrette tempestivamente, attraverso test aperti, audit periodici e canali di segnalazione accessibili. È ritenuto importante che esistano procedure chiare per la gestione delle violazioni nell'uso dei dati di training, con responsabilità definite sia per i fornitori sia per i committenti pubblici. È stato inoltre ricordato che la divisione dei dataset in porzioni distinte per training e con procedure di validazione (il già citato *golden test*) per verificare se il modello sta imparando e cosa sta imparando, può garantire una verifica oggettiva delle prestazioni del modello e contribuire a individuare le distorsioni. Occorre inoltre tenere conto del fatto che dei modelli molto grandi con tantissimi parametri funzionano, ma non sappiamo il perché: sono scatole nere, ma funzionano. Possiamo misurare gli output e stabilire qual è la percentuale accettabile. Un modello che raggiunge il 95% di

risposte esatte verso il golden test è giudicato affidabile e sufficientemente aderente alle nostre linee guida.

Conclusioni ed esiti della sessione

In conclusione, le convergenze principali riguardano le seguenti proposte:

- **necessità di una formazione diffusa e differenziata** per tutti i soggetti coinvolti nei processi di IA;
- **importanza di chiarire concetti e terminologia** per evitare ambiguità su bias e discriminazioni;
- **importanza dei dataset,** che devono essere rappresentativi, puliti e documentati, anche attraverso l'uso di dati sintetici per proteggere la privacy;
- **trasparenza come principio guida**: documentare i processi, rendere pubblici i criteri di addestramento e coinvolgere la società civile;
- **rafforzamento della governance pubblica**, con responsabilità chiare e figure dedicate (es. AI Officer);
- **definizione di regole chiare negli accordi di fornitura,** a tutela dei dati e per l'addestramento dei sistemi**;**
- **istituzione di tavoli interdisciplinari** permanenti di controllo e monitoraggio;
- **predilezione per sistemi verticali e specifici per dominio**, che riducono il rischio di bias generalizzati;
- **attenzione alla localizzazione dei data center** e alla tutela della sovranità dei dati europei;
- **adozione di meccanismi di verifica** e correzione continua dei bias e delle distorsioni.