Rischi e opportunità dell'IA per migliorare equità e accessibilità

Report sessione di discussione gruppo 3

7 ottobre 2025

Partecipanti presenti alla sessione

1. Antonella Fancello, AICA
2. Giuditta Bellosi, Period Think Tank
3. Tatiana Quercia, Lean In Network
4. Mirko Duradoni, UniMercatorum
5. Elisabetta Todeschini, Federmanager

Staff Formez/Facilitazione

1. Annalaura Ciampi, IAF

Sviluppo della discussione

Il/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.

Dopo la lettura condivisa della domanda e una breve presentazione dei partecipanti, la conversazione si è spostata su argomenti già affrontati negli incontri precedenti ma declinati sulla fase di addestramento del sistema: dall'importanza della scelta dei dati, alla formazione del team, ai sistemi di protezione per il cittadino.

Scelta dei dati e risoluzione dei bias - I dati usati dall'IA hanno le caratteristiche della società in cui viviamo, compresi i *bias*, che risultano inevitabili, ma il problema può diventare la soluzione stessa; il sistema può essere addestrato a dubitare di se stesso attraverso la *counter narrative*: se si inseriscono nel sistema due tipi di indicatori diversi (*master narrative* e *counter narrative*), questi potrebbero nel tempo imparare a riequilibrarsi a vicenda. Da valutare inoltre che quello che oggi è un bias, potrebbe non esserlo in futuro, o non esserlo stato in passato: oggi una persona con gli occhi viola è

considerata particolare, ma non sappiamo se lo sarà anche in futuro. Inoltre, non siamo in grado di riconoscere e individuare tutti i bias presenti: se ci affidiamo alla nostra selezione, rischiamo di non affrontarli tutti; se insegniamo al sistema un metodo per riconoscerli, sarà più efficace nel lungo periodo, anche se nell'immediato potrebbe non esserlo. Nel dubbio, il sistema può essere allenato "a dire meno". Per ovviare a questo problema si può prevedere di usare delle fonti certificate di dati, con una dinamica complicata per usi generalisti tipo chat GPT. Tuttavia, si può sempre chiedere che venga esplicitata la fonte o si possono creare delle liste di fonti da cui attingere dati selezionati.
I dati non sono neutri, a seconda delle modalità con cui vengono raccolti cambiano significato: servono sensibilità e competenze sociologiche per riconoscere, ad es., elementi storicizzati ed eliminarli; e servono metodi di mitigazione dei risultati nelle diverse fasi (raccolta, uso, manutenzione).

Team e contributo umano - L'IA act trova una possibile risposta al problema della selezione dei dati, ponendo attenzione alla formazione del team di addestramento: il team infatti dovrebbe avere una maggiore consapevolezza dell'importanza dei dati che vengono scelti e delle azioni umane successive ("il sistema va allineato e riallineato continuamente in base ai prompt che mangia"); questo implica che ci sia un'azione umana che elabora i dati creati dal sistema e decide se/quali correttivi applicare.
Il team di designer dovrebbe includere dei componenti con competenze etiche per intervenire sui temi che riguardano le discriminazioni e far sì che l'algoritmo le riconosca e le eviti, anche se questo meccanismo potrebbe generare a sua volta "dei mostri", ovvero delle forti incoerenze all'interno del sistema. A questo proposito infatti occorre sottolineare che spesso i team che portano avanti tecnologicamente lo sviluppo dell'IA sono "mono genere", con un tipo preciso di *background*, per cui oggi proprio all'interno del team c'è molta segregazione e discriminazione, oltre che un problema di linguistica ("che siano mele, pere o discriminare una persona per un lavoro, per il tecnico informatico non c'è differenza perché sono solo parametri"). Per fare fronte a questo problema si possono usare dei presidi etici (commissioni) per fare dei test aperti in *crowdsourcing* (con persone con competenze specifiche sulle tematiche o meno): se diverse IA danno risultati diversi alla stessa domanda si dovrebbe entrare nel merito e capirne il perché, ad es. se si affrontasse una questione medica, i risultati potrebbero essere diversi a causa di valutazioni del sistema basate su dati che prendono in considerazione etnie diverse del malato. A questo proposito è stata segnalata una buona pratica: a Barcellona si è sperimentata Decode Project, una OpenAI promossa da Francesca Bria, un modello aperto in cui i cittadini scelgono insieme quali dati utilizzare per prendere decisioni democratiche. Un'esperienza simile si è realizzata a Brindisi, con giurie di comunità di IA, per fare entrare i cittadini nelle scelte: secondo il principio che "gli esperti" non devono per forza avere una competenza specifica, come ad es. gli ingegneri, ma devono potersi basare sulla propria esperienza.
Protezione del cittadino - Dentro l'IA act si individua la necessità di un meccanismo di protezione, uno "*stop button* che dia la possibilità di bloccare il sistema ("senza incasinarlo") da usare quando ci si accorge che il sistema ha commesso un errore o lo sta perpetuando in diversi casi. Inoltre, ogni volta che la pubblica amministrazione usa un

meccanismo di IA per prendere una decisione, l'*IA act* cataloga quella scelta come ad alto rischio. Quando il sistema verrà più diffuso e i cittadini saranno più consapevoli, probabilmente ci sarà più attenzione su queste tematiche. Anche per questo l'*IA act* ha chiesto ad ogni Stato di individuare due enti indipendenti dal governo per la tutela del cittadino: benché la Commissione Europea si sia espressa negativamente quando è stata condivisa la bozza dell'Italia con i nominativi scelti (Agenzia per l'Italia digitale e Agenzia per la cybersicurezza nazionale) perché non ritenuti enti indipendenti, i nominativi non sono stati sostituiti. Da considerare che per modificare questi nominativi, individuati tramite legge, servirà una ulteriore legge.
Pensando all'operatività più quotidiana, si potrebbe creare un manuale utente del software, venduto alla P.A. o al cittadino, utilizzabile quando qualcosa non torna, o dare la possibilità all'utente di comunicare tramite *help desk* per verificare la scelta fatta dalla IA.
La creazione di immagini da parte del software infine, merita particolare attenzione: esistono già team di persone che controllano quello che viene restituito dal software e ne fanno una cernita, proteggendo l'utente da *fake*; essendo un lavoro ripetitivo ed estraniante, questo tipo di lavoro attualmente è localizzato principalmente in luoghi del terzo mondo.

Appendice - Risorse citate dai partecipanti

Materiale Lean In
https://drive.google.com/file/d/1voQO2KMvDkQzfSpiwEo8B0Db4NsvYjYP/view

https://drive.google.com/file/d/1dH_neFaRxjzYLD1JMjR30nGNAcg3dXvY/view
Decode project
https://decodeproject.eu/index.html

Conclusioni ed esiti della sessione

In conclusione, le convergenze principali riguardano le seguenti proposte:

- **dare importanza all'azione umana,** attraverso la selezione dei dati, gli allineamenti successivi all'interno del sistema e la formazione del team tecnico;
- **lavorare sulle fonti**, chiedendo sempre che siano esplicitate o usare liste specifiche di fonti da cui attingere dati;
- **allenare il sistema** con criteri di verifica interna;
- **creare team competenti e consapevoli**, con competenze etiche, di linguaggio ed atteggiamenti non discriminanti;
- **considerare buone pratiche** quelle che mettono al centro la componente umana, come è successo a Barcellona e Brindisi;
- **seguire le indicazioni dell'*IA act*** nel definire enti di controllo nazionali indipendenti rispetto al governo italiano e delle dinamiche di blocco del sistema nel momento in cui ci si accorge di errori (stop button);

- **perseguire l'equità**, ovvero tenere a mente che un sistema funzionante deve essere equo e questo richiede un processo più lento di quello che permette la discriminazione e l'imparzialità.