

Titolo: *Inequality and Artificial Intelligence: Trust and Acceptance, Use and Effectiveness of Generative AI and ADMs*

Dipartimento Scienze Politiche e Sociali SPS– Università di Bologna

Questo Progetto fa capo allo Spoke8 (coordinato dall'Università di Bologna) del partenariato esteso PE1 "FAIR: Future Artificial Intelligence Research", dedicato all'Intelligenza Artificiale e guidato dal CNR di Pisa.

Obiettivo del WP8.4 *"The Social Implications of Artificial Intelligence (AI): Data and Models, Acceptance and Use of Generative AI and ADMs"*

Il WP8.4 si pone l'obiettivo di ricostruire e valutare le istanze e gli interessi degli stakeholder (progettista, adottante, utente/ricevitore e pubblico in generale) che progettano, adottano, subiscono e accettano un sistema automatico di supporto alle decisioni (Automated Decision-Making system, ADM) basato su Machine Learning (ML). Lo stesso dicasi per l'utilizzo di sistemi di Generative AI (sullo stile di ChatGPT) che attraverso chatbot sono di supporto ai processi e alle pratiche quotidiane di diversi tipi di utenti. Per ciascuno degli stakeholder (sia esso un progettista, un adottante, un utente/ricevente o il pubblico in generale), saranno delineati gli immaginari socio-tecnici che influenzano le scelte e modellano le percezioni per valutare le asimmetrie di potere, l'impatto in termini di riproduzione di bias e di disuguaglianze sociali e la varietà dei contesti istituzionali di utilizzo.

Approccio metodologico: ci avvarremo di un mix di tecniche quantitative e qualitative: reverse-engineering, audit, interviste in profondità, indagine quantitativa rappresentativa a livello nazionale, modelli predittivi, studi controllati randomizzati (RCT) ed esperimenti controfattuali.

Questa combinazione di tecniche e metodi consentirà di avere un quadro chiaro di:

- qual è la relazione tra diversi sistemi di AI e la riproduzione delle disuguaglianze sociali (identificazione delle fonti di pregiudizio lungo tutto il percorso, a monte e a valle dell'algoritmo);
- quali sono le dimensioni sociali e istituzionali da considerare per ottenere sistemi di AI non discriminanti in diversi settori e casi di studio;
- come costruire una governance solida e rappresentativa di tutti gli stakeholder;
- come aumentare la consapevolezza delle implicazioni sociali di sistemi di AI nei diversi gruppi sociali e nella popolazione in generale (sulla base dei loro immaginari sociotecnici, motivazioni, fiducia e pratiche)
- quali fattori spiegano l'uso e l'adozione di sistemi di AI da parte dei diversi stakeholder in industrie e settori selezionati;
- l'impatto previsto dei sistemi di AI sul mercato del lavoro e sul benessere nei prossimi decenni.

Piano di attività dell'assegnista all'interno del gruppo di ricerca

- 1) Rassegna della letteratura sul tema delle implicazioni sociali dell'AI.
- 2) Selezione e analisi di casi studio e best practices.
- 3) Definizione questionario per una survey da somministrare a un campione rappresentativo della popolazione italiana.
- 4) Definizione di una traccia di intervista a testimoni privilegiati nell'uso di sistemi di AI.

Titolo: Inequality and *Artificial Intelligence: Trust and Acceptance, Use and Effectiveness of Generative AI and ADMs*

Dipartimento Scienze Politiche e Sociali SPS– Università di Bologna

The objective of WP8.4 is to reconstruct and evaluate the instances and interests of the stakeholders (designer, adopter, user/recipient and general public) who design, adopt, undergo and accept an Automated Decision-Making system (ADM) based on Machine Learning (ML). The same applies to the use of Generative AI systems (on the style of ChatGPT) that through chatbots support the daily processes and practices of different types of users. For each stakeholder (be it a designer, an adopter, a user/recipient or the general public), the socio-technical imaginaries that influence choices and shape perceptions will be outlined to assess power asymmetries, the impact in terms of reproducing bias and social inequalities, and the variety of institutional contexts of use.

Methods: we use a mix of quantitative and qualitative techniques: in-depth interviews, nationally representative quantitative survey, predictive models, randomized controlled trials (RCTs), counterfactual experiments, reverse-engineering and audits.

This mixed method approach will provide a clear picture of:

- what is the relationship between different AI systems and the reproduction of social inequalities (identification of sources of bias all along the way, upstream and downstream of the algorithm);
- what are the social and institutional dimensions to be considered in order to achieve non-discriminatory AI systems in different sectors and case studies;
- how to build a solid and representative governance of all stakeholders;
- how to raise awareness about the social implications of AI systems in different social groups and in the general population (based on their socio-technical imaginations, motivations, trust and practices)
- what factors explain the use and adoption of AI systems by different stakeholders in selected industries and sectors;
- the expected impact of AI systems on the labour market and on individuals' welfare in the coming decades.

Post-doc's activity plan within the research group

- 1) Literature review on social implications of AI.
- 2) Selection and analysis of case studies and best practices.
- 3) Work on a questionnaire for a survey to be administered to a nationally representative sample of the Italian population.
- 4) Definition of an interview outline for privileged witnesses in the use of AI systems.