

## **Oggetto attività di ricerca**

Ci sono molti strumenti di intelligenza artificiale, fonti di dati e pratiche di interpretazione utilizzati in Europa. Non è chiaro se danneggino una comunità o un individuo a causa di pregiudizi ingiusti "insinuati" durante lo sviluppo di questi strumenti, possibilmente dovuti ai dati utilizzati per addestrarli, testarli e convalidarli o agli algoritmi implementati o alle pratiche di interpretazione sviluppate intorno a questi applicativi.

In questo contesto, l'obiettivo del progetto è affrontare le molteplici manifestazioni di pregiudizi e iniquità nell'IA da tutte queste dimensioni. In particolare, l'attenzione si concentrerà sul

- (1) fornire linee guida e metodologie per lo sviluppo di applicazioni di AI fair-by-design, sfruttando tecniche di ingegneria del software per la progettazione di sistemi privi di bias;
- (2) tecniche di machine learning per assessment e mitigazione di comportamenti non equi in applicazioni di AI.

## **Dettaglio attività da svolgere**

La ricerca ha come oggetto il supporto alle attività di (1) definizione di tecniche e metodi fair-by-design ed elaborazione delle linee-guida sull'utilizzo di tecnologie di AI prive di bias; (2) definizione di metodi e tecniche per l'assessment e la mitigazione di bias in AI . Le due attività si svolgeranno in parallelo nel corso della ricerca.

In particolare, il supporto alle attività di preparazione dei data set includerà:

- revisione dello stato dell'arte
- selezione di alcune tecniche di machine learning per assessment/reparation di bias
- selezione di alcuni metodi per una progettazione fair-by-design
- test e sperimentazione delle tecniche su diversi casi di studio