

Affidabilità nell'AIoD: Garantire Fiducia, Responsabilità e Governance Etica nella Piattaforma

Progetto di Ricerca

L'AI on Demand platform (AIoD) è una piattaforma Europea emergente che sta diventando una parte essenziale per l'adozione di soluzioni di intelligenza artificiale in Europa in diversi settori. Con l'aumento del suo utilizzo, diventa fondamentale affrontare questioni legate alla fiducia, alla responsabilità e alla governance etica. Garantire l'affidabilità e la trasparenza nelle piattaforme AIoD è cruciale per evitare usi inappropriati, bias, violazioni della privacy e per promuovere un'adozione diffusa e responsabile delle tecnologie IA.

Obiettivi del Progetto

L'obiettivo principale di questo progetto di ricerca è aumentare il trust nella piattaforma AIoD fornendo sul rispetto di standard etici, promuovendone la fiducia degli utenti e offrendo una governance etica robusta. Nello specifico, il progetto si focalizza su:

1. **Sviluppo di standard etici per l'AIoD:** Definire un framework etico solido che guidi lo sviluppo, la distribuzione e l'utilizzo dei modelli di intelligenza artificiale, riducendo al minimo i rischi legati all'uso improprio della tecnologia e garantendo la conformità con le normative vigenti in materia di privacy e sicurezza.
2. **Integrazione del concetto di fiducia nella piattaforma:** Implementare meccanismi che promuovano la trasparenza delle decisioni prese dall'IA, permettendo agli utenti di comprendere e fidarsi delle risposte fornite dal sistema. Questo include il tracciamento e la spiegabilità dei modelli, l'identificazione di bias e la verifica dell'equità nei processi decisionali.
3. **Creazione di un ambiente di test per la fiducia:** Integrare un ambiente dedicato per il test e la validazione delle componenti di fiducia nella piattaforma AIoD (piattaforma sperimentale in ambito fairness).

Piano di Attività

Mese 1-2: Analisi Preliminare

- Analisi delle normative etiche e regolamenti per l'IA.
- Roadmap operativa e rapporto preliminare sulle linee guida etiche.

Mese 3-4: Definizione del Framework Etico

- Sviluppo di un framework etico basato su trasparenza, responsabilità e non-discriminazione.
- Progettazione dei meccanismi di fiducia (auditabilità, spiegabilità, gestione dei bias).

Mese 5-8: Integrazione di un ambiente di sperimentazione fairness nella piattaforma

- Prototipazione dell'ambiente di test per valutare il trust nei modelli IA.
- Test interni dei meccanismi di fiducia e auditabilità.
- Ottimizzazione dell'ambiente di test per scenari complessi.

Mese 9-11: Validazione e Collaborazioni Esterne

- Coinvolgimento di esperti esterni per la revisione e feedback.
- Miglioramenti finali al framework etico e all'ambiente di test.
- Test di robustezza

Mese 12: Conclusione e Disseminazione

- Redazione della documentazione finale e pubblicazione dei risultati.
- Organizzazione di un evento di disseminazione e pubblicazioni scientifiche.
- **Deliverable:** Rapporto finale, pubblicazioni e presentazione pubblica dei risultati.