

**Fairness in AI: fair-by-design methodology and mitigation strategies,  
development and integration**

The goal of the project is the development, implementation and deployment methods to create fair-by-design methodologies for new and fair Artificial Intelligence (AI) approaches and bias mitigation strategies for already existing Machine Learning (ML) models. This goal is in line with the Aequitas EU project, that aims at the design of an environment for developers and users to create controlled experiments for providing effective methods and engineering guidelines to repair, remove, and mitigate bias, and provide fairness-by-design guidelines and software engineering techniques to design new bias-free systems. More precisely, the activity of this project is situated at the intersection of WP4 and WP5 (albeit tight connections with the other WPs are present as well). The methodologies developed will also have to be made part of the AI-on-Demand platform fostered by the EU, in particular a strong link with the AI4Europe project (see WP3) is present as well.

The activities performed in the project will have a twofold nature: on one hand novel techniques and research methodologies will be developed and evaluated, and on the other hand these approaches will be deployed and integrated in existing platforms (e.g., the AIoD platform) and others currently under development (the Aequitas platform). Part of the activity will involve as well the exploration of already existing research items and solutions which could be exploited. All of this will require the candidate to work both on research aspects and on more implementation/application-related issues. The research activity will also involve collaboration with partners from the academia and the industry, especially for the integration of already existing tools and approaches, and to provide support to partners non-expert in the AI field.

**Fairness in AI: metodologia fair-by-design e strategie di mitigazione,  
sviluppo e integrazione**

L'obiettivo del progetto è lo sviluppo e l'implementazione di metodi per creare metodologie fair-by-design per approcci di intelligenza artificiale (AI) nuovi ed equi e strategie di mitigazione dei pregiudizi nei modelli di Machine Learning (ML). Questo obiettivo è in linea con il progetto Aequitas EU, che mira alla progettazione di un ambiente per sviluppatori e utenti per creare esperimenti controllati per fornire metodi efficaci e linee guida ingegneristiche per riparare, rimuovere e mitigare i pregiudizi e fornire linee guida di equità fin dalla progettazione e tecniche di ingegneria del software per progettare nuovi sistemi privi di bias. Più precisamente, l'attività di questo progetto si situa all'intersezione tra WP4 e WP5 (sebbene siano presenti anche stretti collegamenti con gli altri WP). Le metodologie sviluppate dovranno inoltre far parte della piattaforma AI-on-Demand promossa dall'UE, in particolare è presente anche un forte legame con il progetto AI4Europe (vedi WP3).

Le attività svolte nel progetto avranno una duplice natura: da un lato verranno sviluppate e valutate nuove tecniche e metodologie di ricerca, e dall'altro questi approcci saranno implementati e integrati in piattaforme esistenti (ad esempio, la piattaforma AIoD) e altre attualmente in fase di sviluppo (la piattaforma Aequitas). Parte dell'attività riguarderà anche l'esplorazione di elementi di ricerca già esistenti e soluzioni che potrebbero essere sfruttati. Tutto ciò richiederà al candidato di lavorare sia sugli aspetti della ricerca che su questioni più relative all'implementazione / applicazione. L'attività di ricerca coinvolgerà anche la collaborazione con partner del mondo accademico e dell'industria, in particolare per l'integrazione di strumenti e approcci già esistenti, e per fornire supporto a partner non esperti nel campo dell'IA.