

DG.3

Social, Ethical and Legal AI-Fairness Methodologies

This content is based on research conducted for the AEQUITAS report "Preliminary Social, Ethical and Legal AI-Fairness Methodologies"

www.aequitas-project.eu

01.

Social, Ethical and Legal AI-Fairness Methodologies

Our Goal

To develop a methodology for the creation of the social, ethical and legal requirements for the 3 AEQUITAS Engines:



02.

Our Methodology



Development of AEQUITAS engines integrating Social, Ethical and Legal Perspectives

What do we need to address?

- What legal elements should be considered.
- What ethical elements should be considered, based on the Ethics Guidelines of Trustworthy AI (EGTAI).
- What social elements should be considered, based on an existing literature, methodologies, and processes.

The preliminary table shown below aims to highlight the requirements that must be considered for each of the AEQUITAS Engines at the data, model and interpretation levels.

	Awareness and Diagnostics Engine ADE	Repairation and Mitigation Engine RME	Fairness-by-Design Engine FDE
Data	Legal: Article 10 of the AI Act, specifically, those under art. 10.2, art. 10.3, and art. 10.4.	Technical repairation Stakeholder engagement + Re-diagnosis (e.g. re-evaluating art. 10.3)	Fair data collection methodology
Model	Ethical: ALTA/EGTAI Requirement 4 - Transparency	Technical repairation Stakeholder engagement + Re-diagnosis (e.g. re-evaluating art. 10.3)	AI Act Risk Classification
Interpretation	Legal: Article 14 of the AI Act	Technical repairation Stakeholder engagement + Re-diagnosis (e.g. re-evaluating art. 10.3)	Stakeholder Engagement

03.

User Stories

We aim to develop the requirements needed for the detection of bias in ethical, social, legal, economic and cultural contexts. More specifically, we aim to develop user stories that can diagnose and assess fairness in AI-systems at the data, model, and interpretation level.

Awareness and Diagnostics Engine (ADE)

TITLE	PRIORITY	ID	PROPOSER
Data compliance	xxxxxxxxxx	xxxxxxxxxx	xxxxxxxxxx

** As an AI developer (researcher, engineer, data scientists) I want to be able to check if the dataset is fair as regards the persons or groups on which the system is intended to be used (art. 10.3 & 10.4 AIA). So that I can evaluate the level of compliance with the AIA and act accordingly.

ACCEPTANCE CRITERIA → Elements to be identified/visualized for the dataset as regards the person or groups on which the system is intended to be used.

Relevancy	Representativeness	Free of errors (as far as possible)	Completeness
Appropriateness of Statistical properties	Characteristics of Geographical setting	Characteristics of Behavioural setting	Characteristics of Functional setting

Repairation and Mitigation Engine (RME)

TITLE	PRIORITY	ID	PROPOSER
Data repairation	xxxxxxxxxx	xxxxxxxxxx	xxxxxxxxxx

** As an AI developer (researcher, engineer, data scientists) I want to be able to repair unfair elements found in the data as expressed in user story ID: x.

ACCEPTANCE CRITERIA → The repairation method should include technical repairation in combination with the re-involvement of relevant stakeholder to ensure that social and legal aspects are still taken into consideration.

Fairness-by-Design Engine (FDE)

In contrast to the other engines, the Fairness-by-Design engine allows us to identify requirements that can be translated into tools and methodologies which we refer to as 'building blocks'. In turn, these building blocks can be aligned to the AI lifecycle so AI developers or other users of the AEQUITAS tool can follow fairness-by-design methodologies with ease.

TITLE	PRIORITY	ID	PROPOSER
Data repairation	xxxxxxxxxx	xxxxxxxxxx	xxxxxxxxxx

As an AI developer **I want** a visual representation of the AI lifecycle. **So that** I can have a smoother interaction with the FbD methodology

ACCEPTANCE CRITERIA → Lifecycle should include the following phases: Scoping, Risk Analysis, Development (Data, Algorithm, Interpretation), Evaluation (Testing, Deployment), Monitoring in Operation, Re-Evaluation, Decommissioning

Requirements needed to be implemented in the Fairness-by-Design engine:

The need to have a trustworthy AI elaboration process	The need to use EGTAI/ALTAI	The need to identify and involve multidisciplinary expertise	The need to identify and involve stakeholders
The need to align with the AI Act	The need to assess the impact on fundamental rights, democracy and the rule of law	The need to align with AI-related regulation (existing and upcoming)	

Building blocks of a holistic methodology for AI Fairness-by-Design

AI fairness Impact Assessment/AI fairness Readiness Assessment	Ethics Guidelines for Trustworthy AI and Assessment List for Trustworthy AI (HLEG AI)	Socio-technical Matrix
Stakeholder Identification Methodology (developed as part of Deliverable 6.1)	Stakeholder Engagement Methodology	Trustworthy AI Deliberation (based on the 7-step exercise for Trustworthy AI, developed for the Trustworthy AI Project [Erasmus+])
AI Fairness Regulatory Landscape Identification and Assessment	AI Act Risk Classification	AI Act High-Risk Requirements Guidance for Fairness
Fundamental Rights Impact Assessment	Fair software engineering methodologies architectures and methods:	<ul style="list-style-type: none"> Fair System Architecture Methodology Fairness Criteria Definition, Assessment and Monitoring Fair Data Collection Methodology
AI fairness Evaluation/Bias Audit	Monitoring in Operation through Critical Control Points	

What's next?

The requirements identified and observed will be used as building blocks for the 3 AEQUITAS ENGINES.



Want to know more about AEQUITAS?

Visit our website for more information: www.aequitas-project.eu

Follow us on social media: @aequitasEU

Contact us: info@aequitas-project.eu

