

Non-discrimination by design

What is the problem and how will AI help solve it?
What percentage of false negatives and false positives is acceptable, and why?

1

Problem definition

2

Data collection

Which data are required for this project and to what extent are these data already available within the organization? How are the data biased?

3

Data preparation

Which criteria are used for data selection and how do they reflect distinctions made between groups?

Modelling

4

How are criteria concerning explainability and fairness translated into a model selection strategy? How does the model perform on the predetermined success criteria in terms of false positives and false negatives?

6

Evaluation

Pick an implementation strategy and formulate an evaluation and exit strategy. Assess how the system would function in case a different model, fairness definition and/or algorithm had been chosen.

5

Implementation

Choose a specific application to test the system; ensure that this clearly defined application is representative of the entire domain in which the AI system will eventually be used. Adjust the model based on the results of the test case and consult stakeholders.

