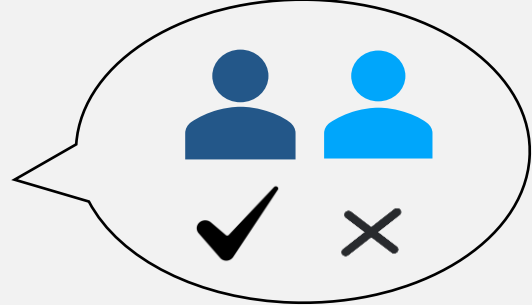
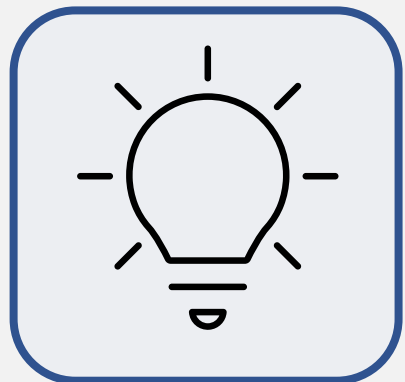


Functional Requirements in Interactive Auditing Tools

BIAS IN AUTOMATED DECISION MAKING



Potential of Interactive Toolkits



Understand & contextualize bias



Adhere to legal requirements



Find strategies for bias mitigation

UNDERSTANDING REQUIREMENTS



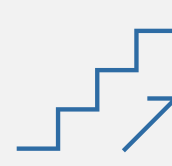
Analysing Previous Interview Studies



Design Considerations



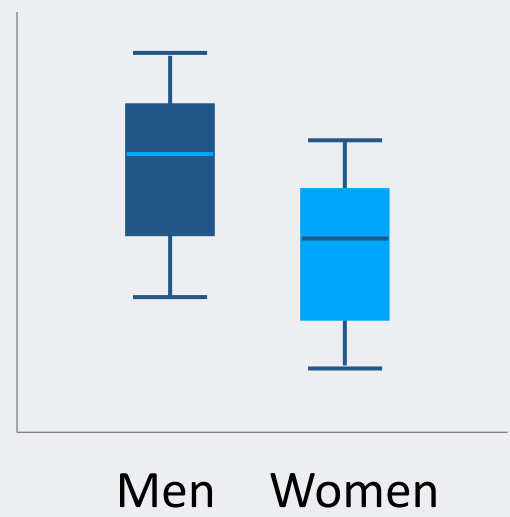
Inspection of Existing Tools



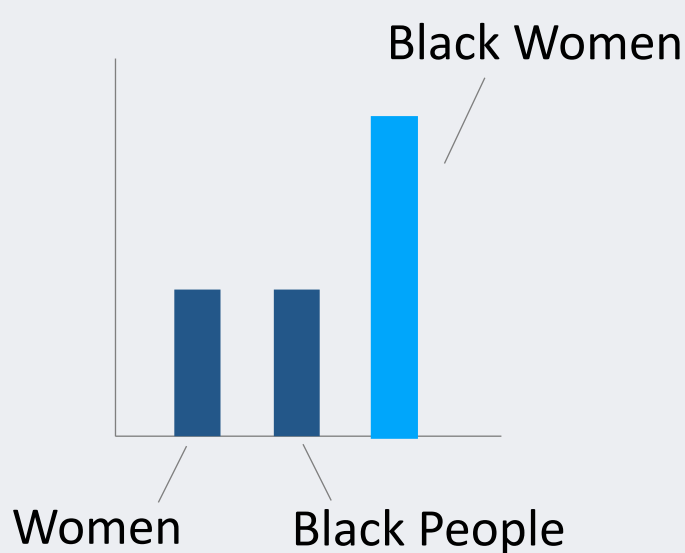
Actionable Insights for Tool Improvement

FUNCTIONAL REQUIREMENT CHECKLIST

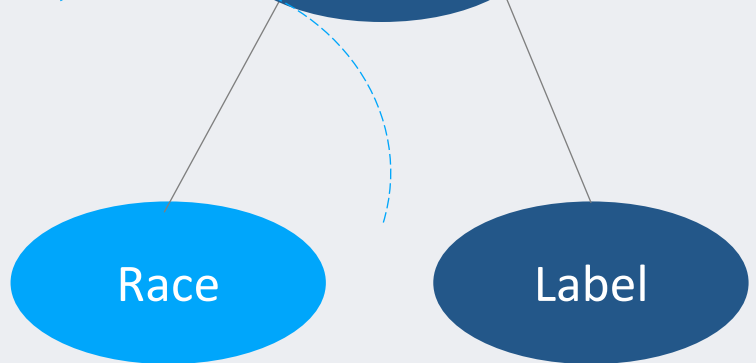
Prediction Probabilities



False Negative Rate

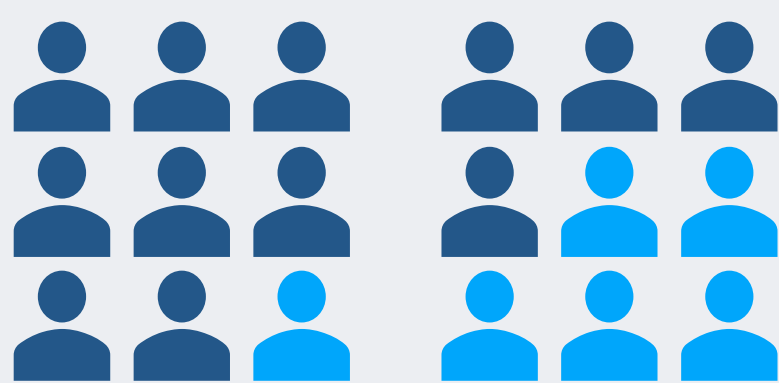


Zip Code



Train

Test
Representation
Bias?



BIAS DETECTION IN PREDICTIONS

Different forms of bias can be detected

Tool supports following bias definitions:

- Outcome based
- Actual vs. Outcome based
- Probability based
- Similarity based
- Causal based

Tool provides customizable metrics

Intersectional Analysis is supported

- Bias based on non-binary sensitive attributes
- Bias based on combination of sensitive attributes

Bias beyond sensitive attribute(s)

- Contextualization of differences in outcomes
- Indirect bias analysis

BIAS DETECTION IN INPUT DATA

- Relation between attributes and ground truth
- Relation between attributes
- Inspect subgroup sizes
- Inspect differences between train- and test set

SCALABILITY OF AUDIT

Prioritizing systemic biases

- Report of confidence intervals
- Grouping similar subgroups together
- Tool groups similar individuals together

Accounting for blind spots

- Automatic suggestion of subgroup biases
- Automatic suggestion of individual biases

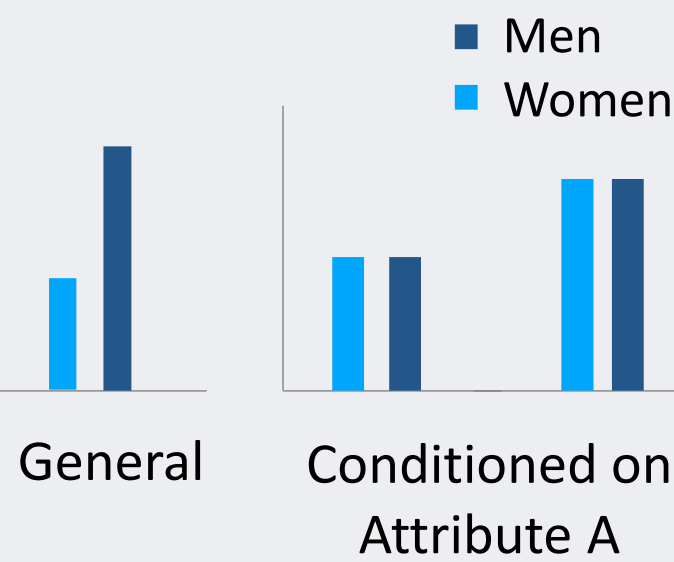
Define your equation here

Attribute A Σ \square \square

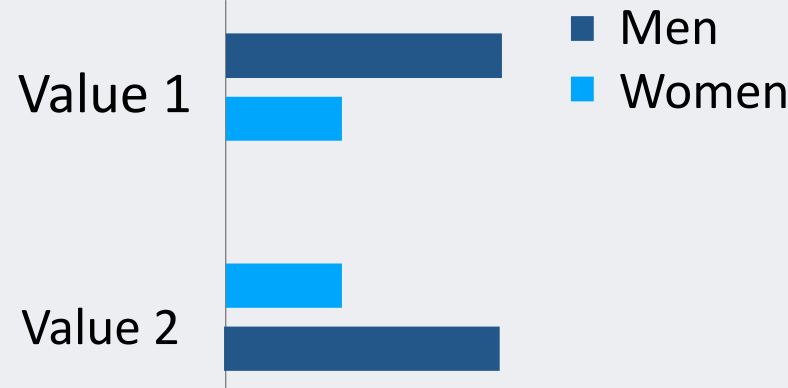
Attribute B \square \square \square

Attribute C $+$ $-$ \times

Approval ratio



Attribute A



★ Potential bias

