Algorithmic accountability and fairness

A computer scientist’s perspective
What is algorithmic accountability?

• Addresses problems with algorithms that interact with society and affect it, e.g. ADM-systems (algorithmic decision making) - especially the learning ones

• Who is accountable?
What is algorithmic accountability?

• Addresses problems with algorithms that interact with society and affect it, e.g. ADM-systems (algorithmic decision making) - especially the learning ones

• Who should feel accountable?

• What is fair?

• How can we implement algorithmic accountability?
Operationalization

Data Scientist

Researcher

Development of analytic method

Implementation

Development of analytic method

Implementation

Development of analytic method

Implementation

Method selection

Data collection

Person or Institution

Data collection

Data collection

Data collection

Data selection

Trained Decision System

Data

Interpretation of result

Decision of action

Feedback

Quality and fairness measures

Chain of responsibilities

Concept and visualization by Katharina Zweig
Who should feel accountable?

Demand-Supply Models
Definitions of Fairness

01. Oxford Dictionary
   The quality of treating people equally or in a way that is right or reasonable.

02. Lexico Dictionary
   Impartial and just treatment or behavior without favoritism or discrimination.

03. Webster’s Dictionary
   Fair or impartial treatment: lack of favoritism toward one side or another

04. Collins Dictionary
   Fairness is the quality of being reasonable, right and just.
Definitions of Fairness

Equality

Equity

https://edtrust.org/the-equality-line/equity-and-equality-are-not-equal/
Group fairness vs. Individual fairness

**Group fairness:**
- Protected groups should be treated similarly to the advantaged group or the populations as a whole.
- Does not consider the individual merits.
- May result in choosing the less qualified members of a group.

**Individual fairness:**
- Individuals should be treated consistently.
- Assumes a similarity metric of the individuals that may be hard to find.
- This kind of fairness is rarely used
Diversity
(Group) Fairness measures

!! Most of these measures are in conflict with each other !!
Fairness measures

- Independence
  - Relaxed Independence
  - Conditional Independence
- Separation
- Equalized Correlation
- Overall Accuracy Equality
- Sufficiency
- Conditional Use Accuracy
- Well Calibration
- Treatment Equality
- False Positive Error Rate Balance
- Balance for Positive Class

Maybe I just flip a coin...
Fairness measures

- Independence
  - Relaxed Independence
  - Conditional Independence
- Separation
- Equalized Correlation
- Overall Accuracy Equality
- Sufficiency
- Conditional Use Accuracy
- Well Calibration
- Treatment Equality
- False Positive Error Rate Balance
- Balance for Positive Class

...maybe they can help

Society

Scientists

Domain experts (Philosophy, IT, Law, ...)

Politics
When do we need regulation?

ADM-system need to be regulated, normed, and/or controlled if they
a) contain a learning or learned component
b) that makes decisions about humans or their belongings or that gives access to limited resources
c) independent of whether a human is in the loop or not
d) with respect to the logic and mechanism of their decision making.

Don't worry, we figured out who is responsible!
Why only then?

Both need to be checked for product safety and security.

Additional need to check for:

• Bias in data
• Data quality and representativeness
• Correct operationalization of human values
• Result fairness and quality and
• Justification and explainable decision making (for the possibility of appealing)

ADM systems deciding about things

ADM systems deciding about people and resources
How much control is necessary?

Needs to be differentiated by the

a) total potential individual and societal damage of using the ADM system in a given social context and

b) the dependency of the scored/classified subject on the decision.
Five classes of transparency and accountability requirements

Class 0
(Post-hoc analysis in suspicious cases)

Class 1
Permanent control by a black-box-analysis, and method used, ...

Class 2
Transparency about the quality index values, input data, algorithm used, ...

Class 3
Only explainable systems, no ADM-

Class 4
High potential damage

Low potential damage

Low dependency

Strong dependency

Concept and visualization by Katharina Zweig and Tobias D. Krafft
Sources

Maryam Haeri: Paper unpublished yet

Christopher Koska: https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Bibliothek/Doi_Publikationen/Ethik_fuer_Algorkithmiker._Was_wir_von_erfolgreichen_Professionsethiken_lernen_koennen._Final..pdf

https://www.vzbv.de/sites/default/files/downloads/2019/05/02/19-01-22_zweig_krafft_transparenz_adm-neu.pdf (p.13 and p.29)

Cliparts: https://webstockreview.net/