



Funded under the Rights, Equality and Citizenship Programme 2014-2020
of the European Commission

AI Discrimination and Algorithmic Fairness – Technical Solutions and Legal Constraints

Prof. Dr. Philipp Hacker, LL.M. (Yale)
European New School of Digital Studies, European
University Viadrina

Outline

- I. Algorithmic Discrimination under
EU Law**
- II. Algorithmic Fairness**
- III. Normative and Legal Constraints
for Algorithmic Fairness**

Part I:

Algorithmic Discrimination under EU Law

Examples of Algorithmic Discrimination

Job selection algorithms

**The
Guardian**

Amazon ditched AI recruiting tool that favored men for technical jobs

Specialists had been building computer programs since 2014 to review résumés in an effort to automate the search process

Source: <https://www.theguardian.com/technology/2018/oct/10/amazon-hiring-ai-gender-bias-recruiting-engine>

Examples of Algorithmic Discrimination

Medical AI

RESEARCH ARTICLE

ECONOMICS

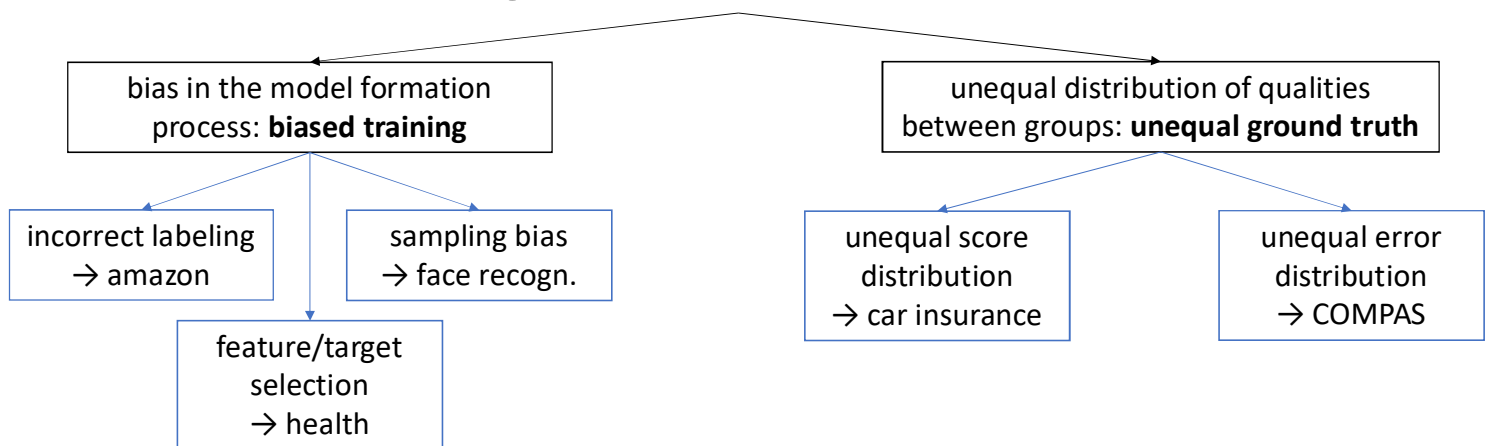
Dissecting racial bias in an algorithm used to manage the health of populations

Ziad Obermeyer^{1,2*}, Brian Powers³, Christine Vogeli⁴, Sendhil Mullainathan^{5*†}

Source: Obermeyer et al., 366 Science 447 (2019)

The Sources of Algorithmic Discrimination

Algorithmic Discrimination



The Law of Algorithmic Discrimination

Anti-Discrimination Law (Hacker, Teaching Fairness to Artificial Intelligence, Common Market Law Review 2018)

- 1) Coverage of algorithmic discrimination
- 2) AI discrimination: mostly indirect discrimination
- 3) Justification: legitimate aim and discriminatory practice proportionate
 - a) biased training → rather (-)
 - b) unequal ground truth → rather (+)
- Problem of differentiation between sources of discrimination
- 4) Enforcement problems: no access to data and model
- 5) The GDPR as a merely partial remedy

Part II:

Algorithmic Fairness

Algorithmic Fairness

Definitions of fairness: 2 main groups (Dwork, 2012; Friedler et al., 2016; Pessach/Shmueli, 2020)

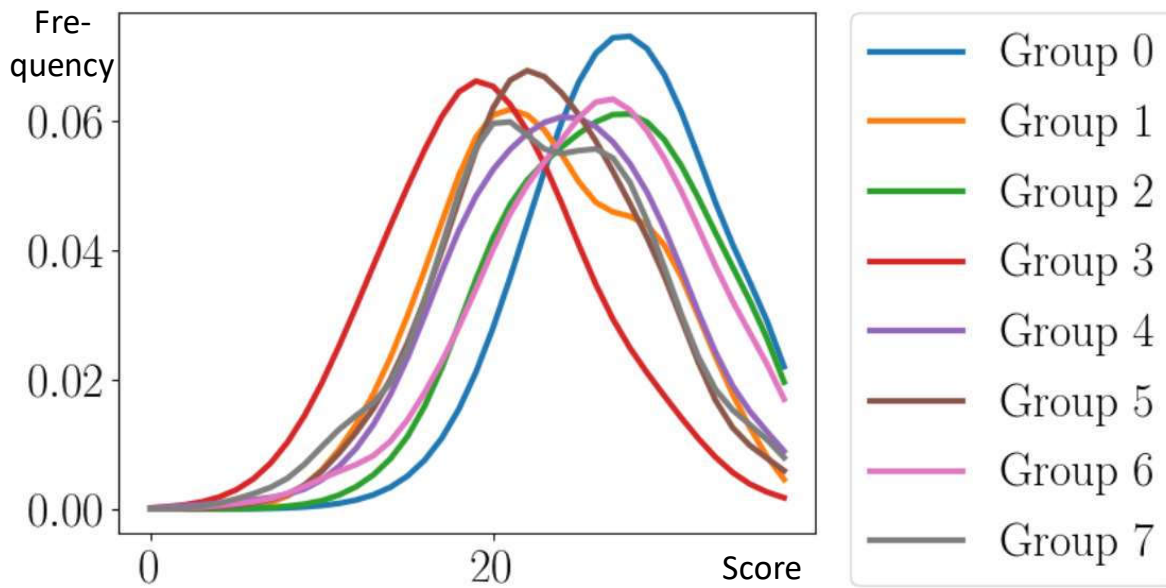
- 1) Individual Fairness: similar input \rightarrow similar output
 - Aristotle, Nichomachean Ethics, Book V, § 3, 1131a10
 - CJEU: equality before the law, Art. 20 ChFR
 - 2) Group Fairness: e.g., same positive selection rate for each group (statistical parity)
 - Outcome-egalitarian concept (Binns, 2018)
 - Near impossibility of indirect discrimination
- \rightarrow Trade-off necessary: more GF \leftrightarrow less IF

Bridging the Divide: Our Model

Zehlike/Hacker/Wiedemann, Matching Code and Law, 34 Data Mining and Knowledge Discovery 2020, 163:

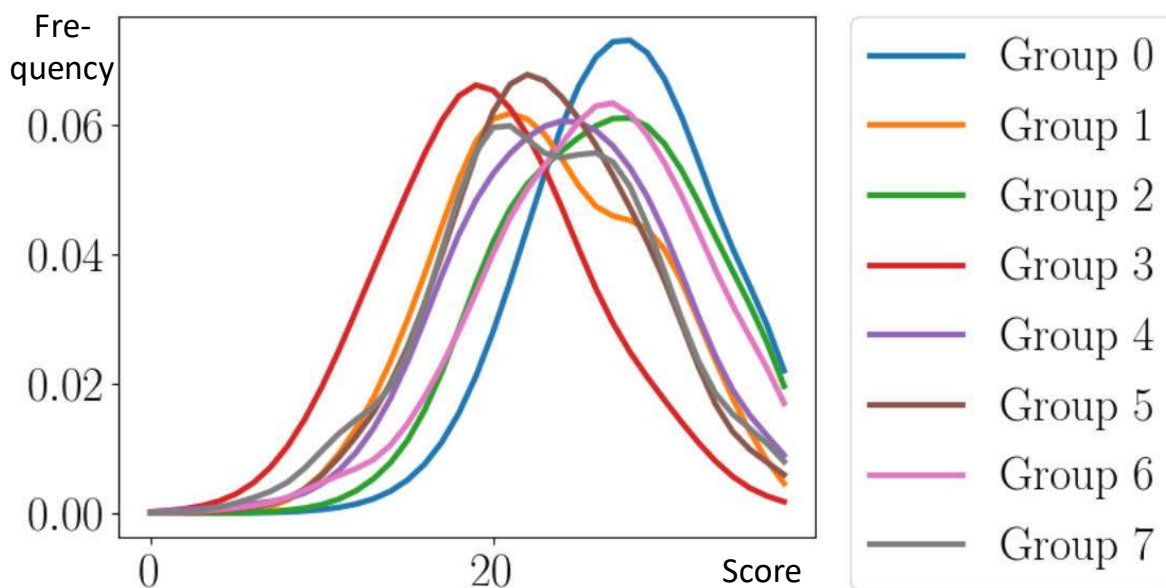
- Continuous interpolation between measures of individual and group fairness
- Parameter $\theta \in [0;1]$: degree of approximation of group distributions
 - $\theta = 0 \rightarrow$ individual differences are fully preserved (IF)
 - $\theta = 1 \rightarrow$ group distributions fully mapped onto barycenter (GF)
- Minimal information loss for decision maker through optimal transport

LSAT Scores (ethnicity): descriptive statistics



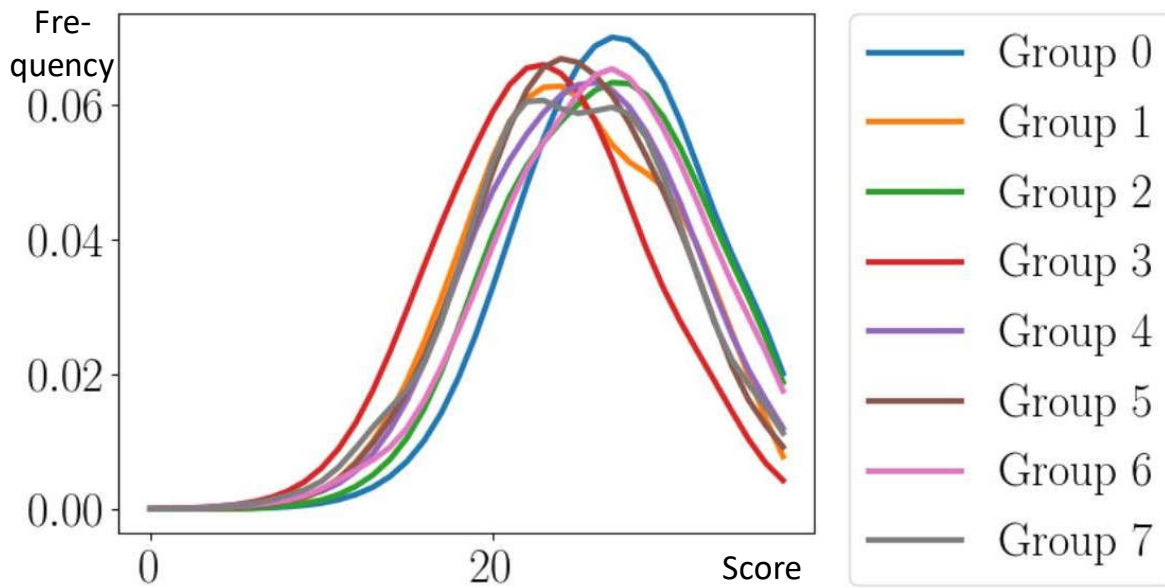
Source: Zehlike/Hacker/Wiedemann, 34 Data Mining and Knowledge Discovery 2020, 163

LSAT Scores (ethnicity) for $\theta = 0$



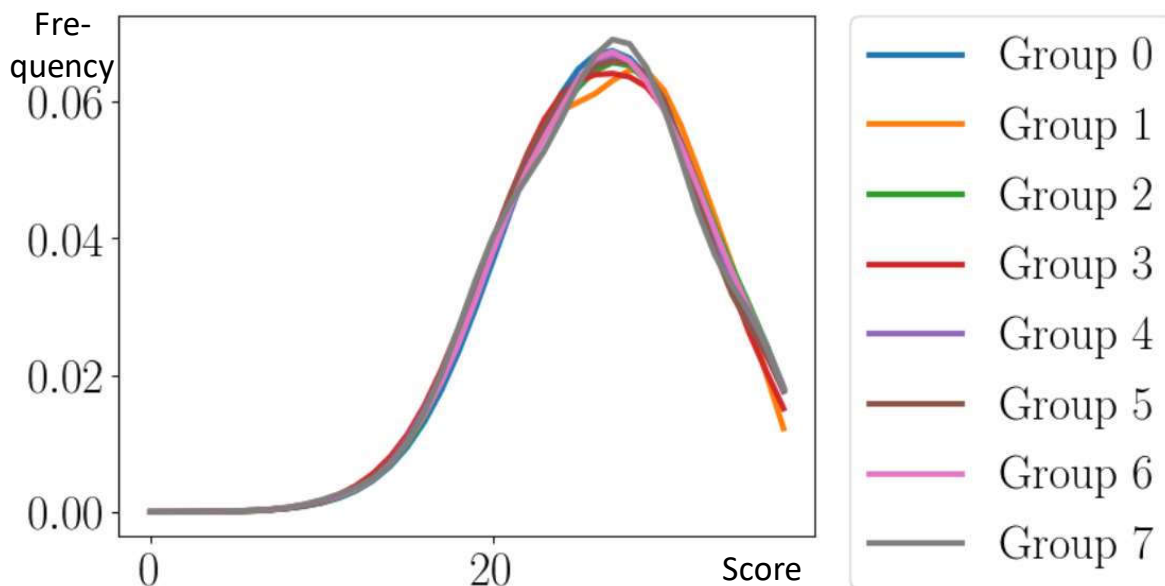
Source: Zehlike/Hacker/Wiedemann, 34 Data Mining and Knowledge Discovery 2020, 163

LSAT Scores (ethnicity) for $\theta = 0.5$



Source: Zehlike/Hacker/Wiedemann, 34 Data Mining and Knowledge Discovery 2020, 163

LSAT Scores (ethnicity) for $\theta = 1$

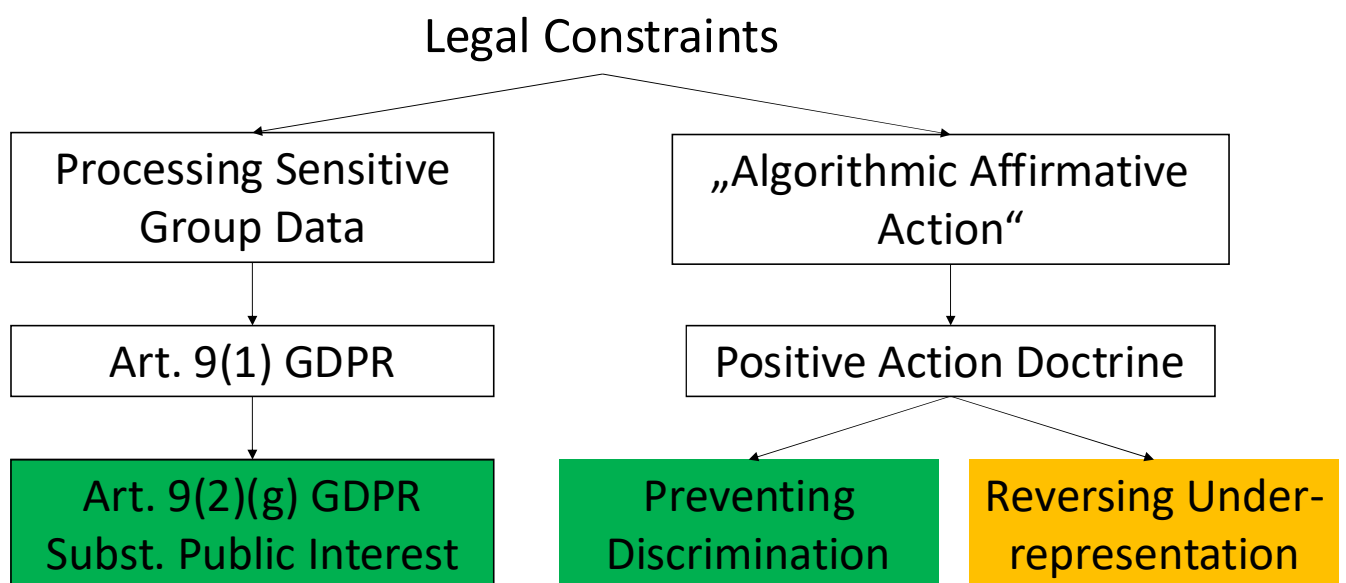


Source: Zehlike/Hacker/Wiedemann, 34 Data Mining and Knowledge Discovery 2020, 163

Part III:

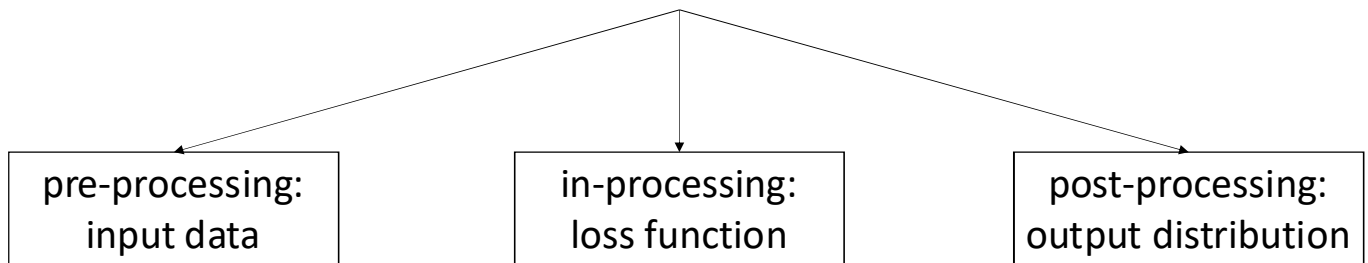
Normative and Legal Constraints for Algorithmic Fairness

The Legality of Algorithmic Fairness



Algorithmic Fairness Procedures

types of algorithmic fairness



Algorithmic Affirmative Action

CJEU guidelines:

1) During selection phase (results):

Marschall: restrictive criteria: only on the basis of all available information of the specific case
→ human in the loop, no automatic re-ranking

?

post-
processing
approaches

2) Before selection phase (opportunity):

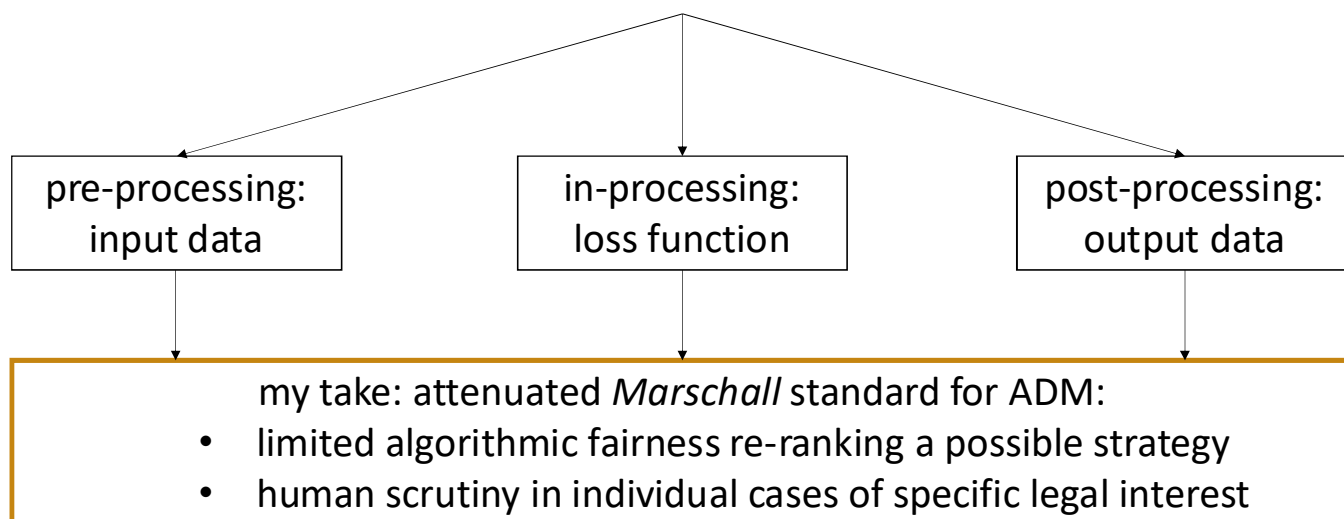
Badeck: lenient criteria: even quota possible

?

pre-/in-
processing
approaches

Algorithmic Fairness Procedures

types of algorithmic fairness



Conclusion

- 1) Enforcement bottleneck in AI discrimination
- 2) Abundance of fairness metrics in Computer Science
- 3) Our model:
 - a) Bridges individual and group fairness
 - b) While minimizing information loss
- 4) Legal constraints:
 - a) Data protection law: sensitive data
 - b) Anti-discrimination law: positive action doctrine
- 5) Allow automated fairness interventions to reduce discrimination, incentivize HMT to mitigate underrepresentation

Thank you!

hacker@europa-uni.de

