

THE AWARDS UNIVERSITY OF THE YEAR

'Fairness' in Machine learning for Healthcare Applications

Dr. Fani Deligianni,

fani.deligianni@glasgow.ac.uk

Lecturer (Assistant Professor)

nttps://www.gla.ac.uk/schools/computing/staff/fanideligianni Lead of the Computing Technologies for Healthcare Theme

WORLD CHANGING GLASGOW



Bias in Data Mining

- Limited samples of subgroups
- Limited features of subgroups
- Bias over bias



Historic Examples in Algorithmic Bias

Gender and Racial Discrimination has been captured by an algorithm and reproduced







Bias in Al Algorithms

Discrimination in Online Ads



































































































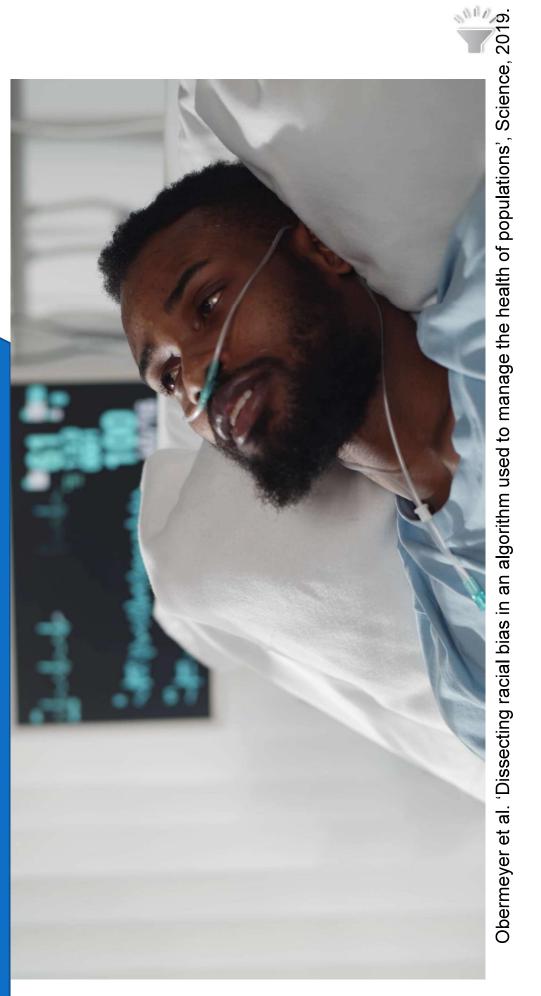




Proxies to sensitive attributes

- Anti-discrimination law prohibits unfair treatment based on sensitive attributes, such as gender or race
- Implicit features may correlate with sensitive attributes
- Inherently algorithms inherit the prejudices of prior decision makers

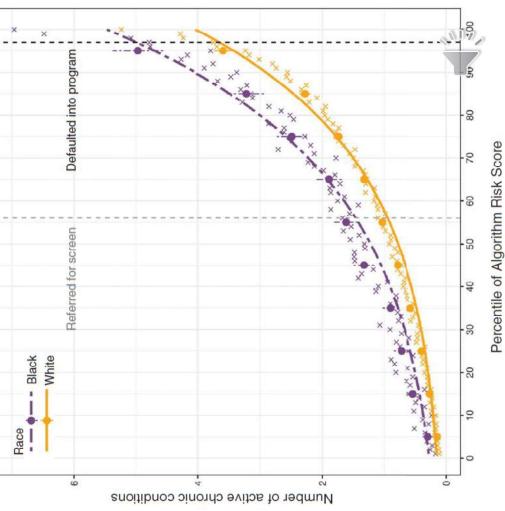
Racial Bias in Healthcare



Racial Bias in Healthcare



- At 97th percentile of risk score blacks have 26.3% more chronic illness than whites
- Significant evidence of disparities that favor white people

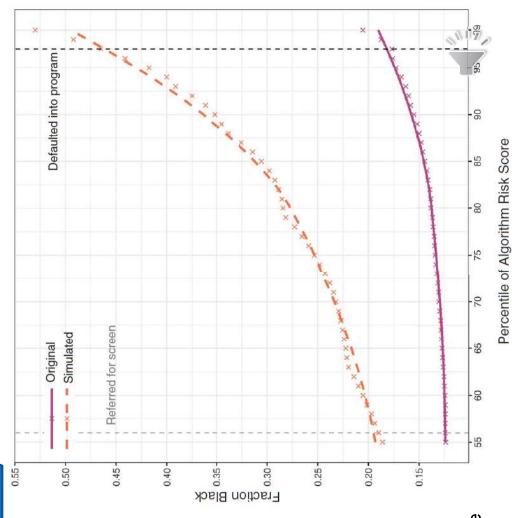


Obermeyer et al. 'Dissecting racial bias in an algorithm used to manage the health of populations', Science, 2019.

Racial Bias in Healthcare

- Simulate an algorithm with no prediction gaps between blacks and whites
- At 97th percentile among the fraction of black patients would rise from 17.7 to 46.5%
- Evaluate the performance of the model
- External Validation
- Generalisability to 'plausibly related populations'

Obermeyer et al. 'Dissecting racial bias in an algorithm used to manage the health of populations', Science, 2019.





Mechanisms of Bias

- The algorithm takes into consideration several features:
- Demographics, insurance type, diagnosis, procedure codes, medications and detailed costs
- Exclude race
- Prediction label is cost (The program is used to target high costs)
- The model is unbiased in terms of the cost prediction but biased in terms of the severity of the disease/multi-morbidity



Summary

- Historic data include discrimination bias in ways that are difficult to detect and eliminate
- Removing sensitive fields is important but is not adequate since the algorithm can identify proxies and learn based on them
- Quantifying calibration bias is used to detect bias that can disadvantage protective groups

References

- Barocas et al. 'Big Data's Disparate Impact', 2016.
- Obermeyer et al. 'Dissecting racial bias in an algorithm used to manage the health of populations', Sciences, 2019.