Foundation of Data Science
Lecture 11, Module 1
Fall 2022

Rumi Chunara, PhD

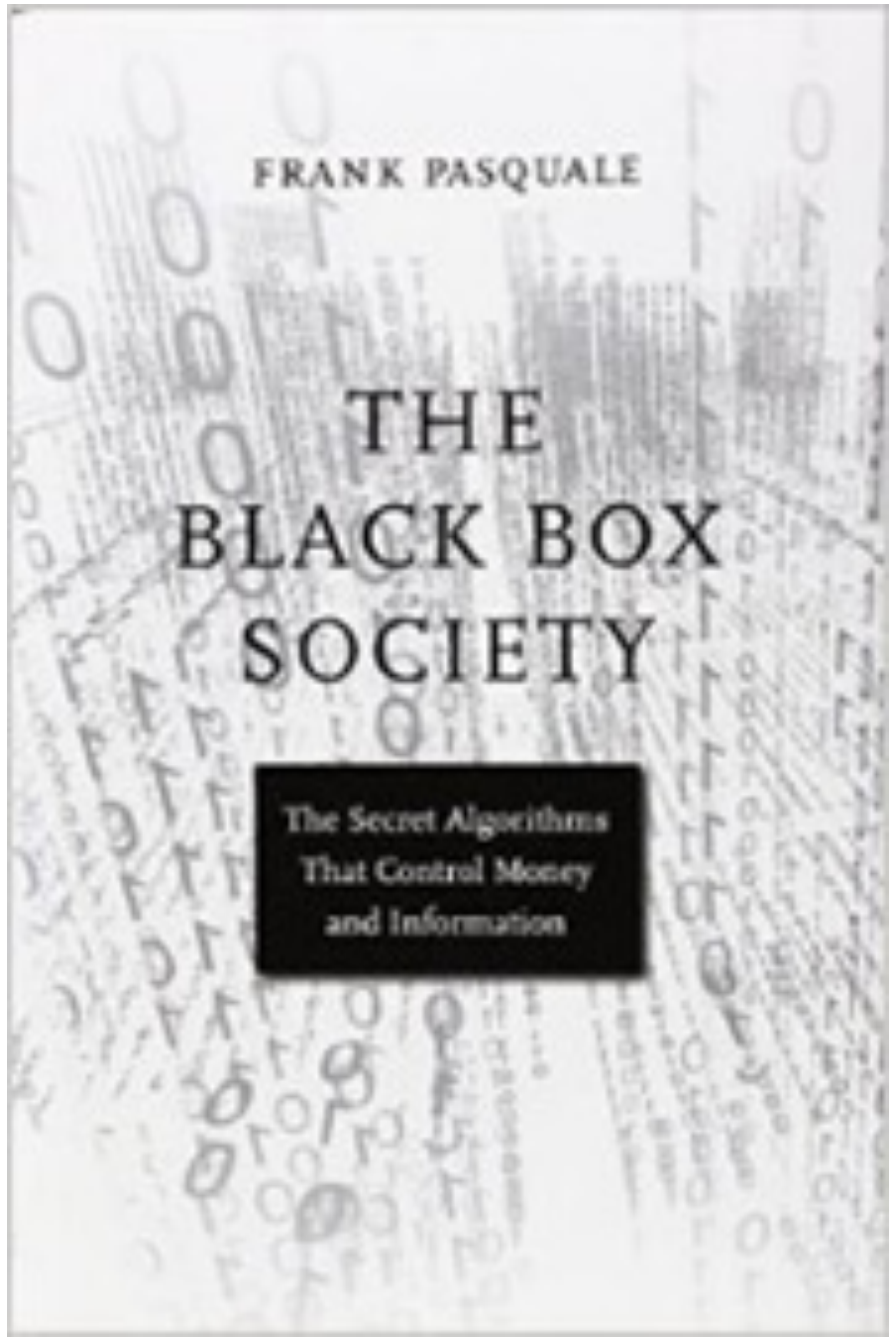
Slides from KDD 2016 Tutorial given by Sara Hajian, Francesco Bonchi, Carlos Castillo,
ICML 2019 Tutorial by Silvia Chiappa & Jan Leike
and slides on Algorithmic Fairness from Julia Stoyanovich incorporated.

Algorithms Embedded in Society

Algorithms can be "black boxes" protected by

- Industrial secrecy
- Legal protections
- Intentional obfuscation

Discrimination becomes invisible
Mitigation becomes impossible



F. Pasquale (2015): The Black Box Society. Harvard University Press.

Artificial Intelligence

“We define AI as the study of agents that receive percepts from the environment and perform actions. The main unifying theme is the idea of an intelligent agent.”

<u>AI in business management</u>

- spam filters
- smart email categorisation
- voice to text features
- smart personal assistants
- automated customer support
- process automation
- sales and business forecasting

<u>AI in e-commerce</u>

- smart searches
- personalisation as a service
- product recommendations
- fraud detection and prevention
- dynamic price optimisation

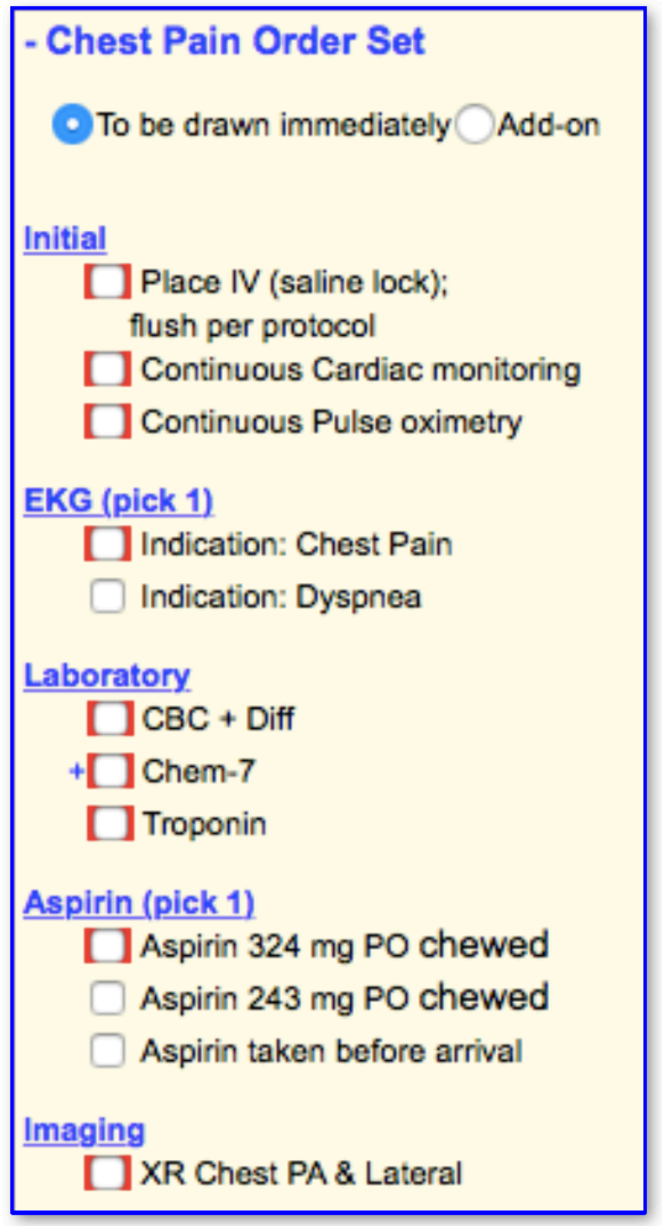
<u>AI in marketing</u>

- personalisation of news feeds
- pattern and image recognition
- ad targeting
- customer segmentation
- predictive customer service

Russell, Stuart J., and Peter Norvig. *Artificial intelligence: a modern approach*. Malaysia; Pearson Education Limited,, 2016.

AI in Health

- Healthcare data analysis and prediction
- Automating repetitive tasks
- Treatment decisions
- Drug creation
- Health monitoring
- Health systems optimization


- Chest Pain Order Set

To be drawn immediately Add-on

Initial
Place IV (saline lock); flush per protocol
Continuous Cardiac monitoring
Continuous Pulse oximetry

EKG (pick 1)
Indication: Chest Pain
Indication: Dyspnea

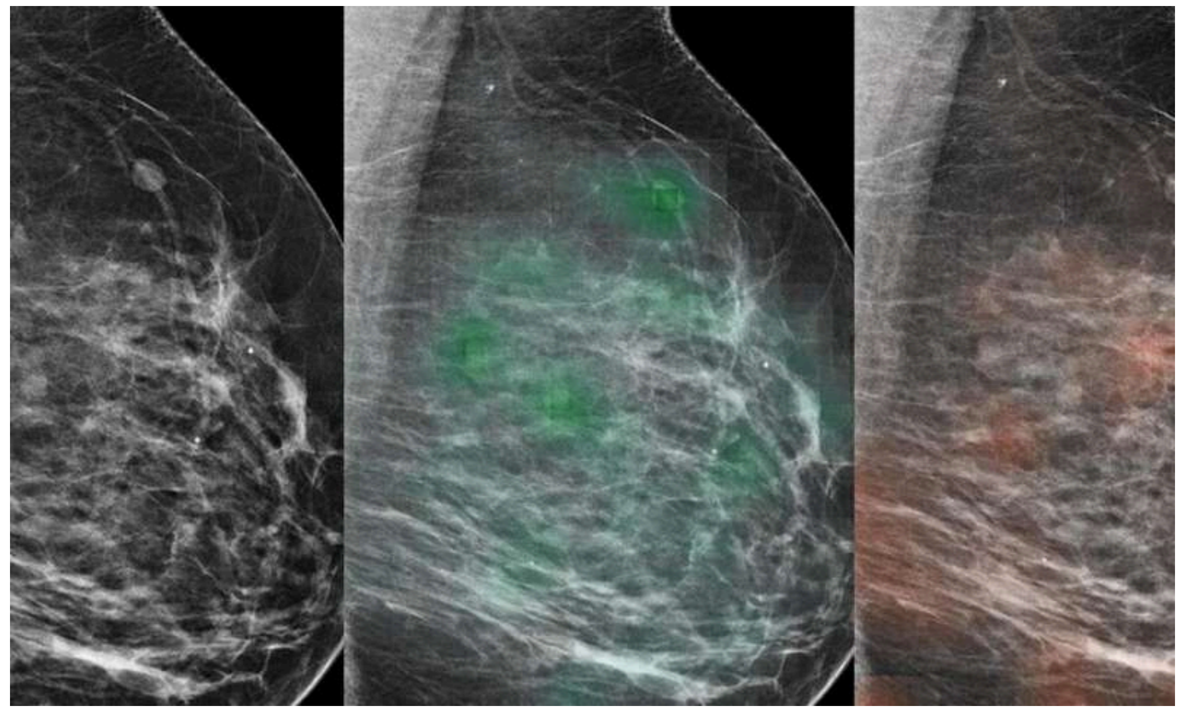
Laboratory
CBC + Diff
Chem-7
Troponin

Aspirin (pick 1)
Aspirin 324 mg PO chewed
Aspirin 243 mg PO chewed
Aspirin taken before arrival

Imaging
XR Chest PA & Lateral

[Rotter et al.
BMC Health Services 2008]

[Saria et al 2010]



[Wu et al IEEE Trans Med Imag 2019]

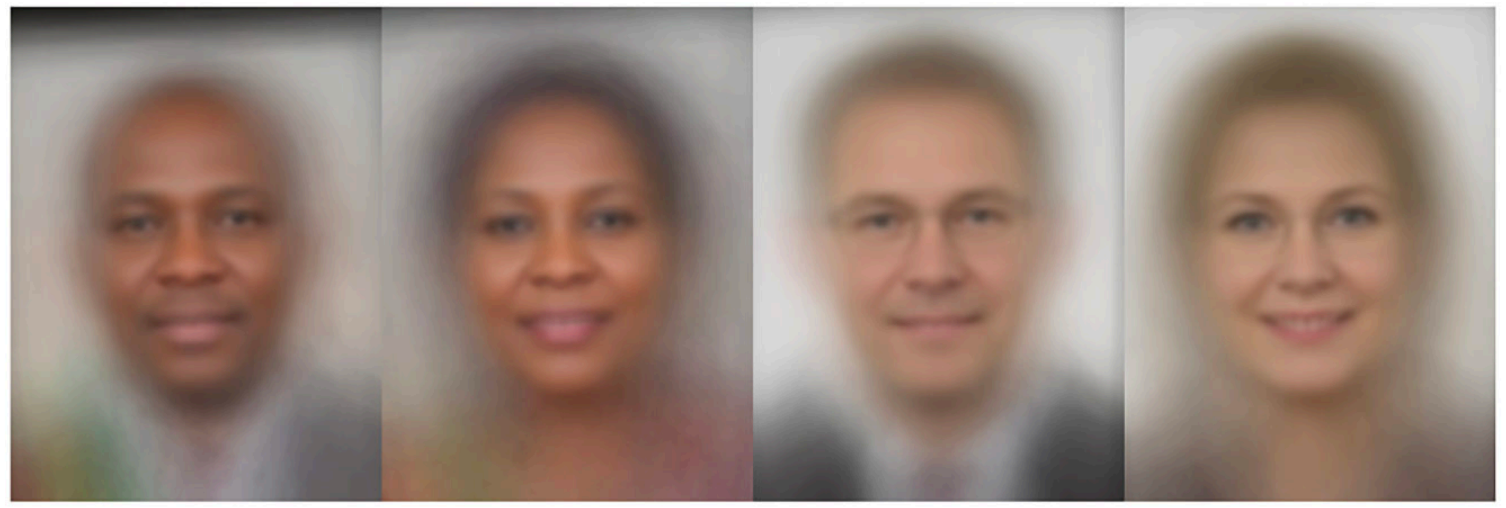
Bias in Facial Recognition Algorithms

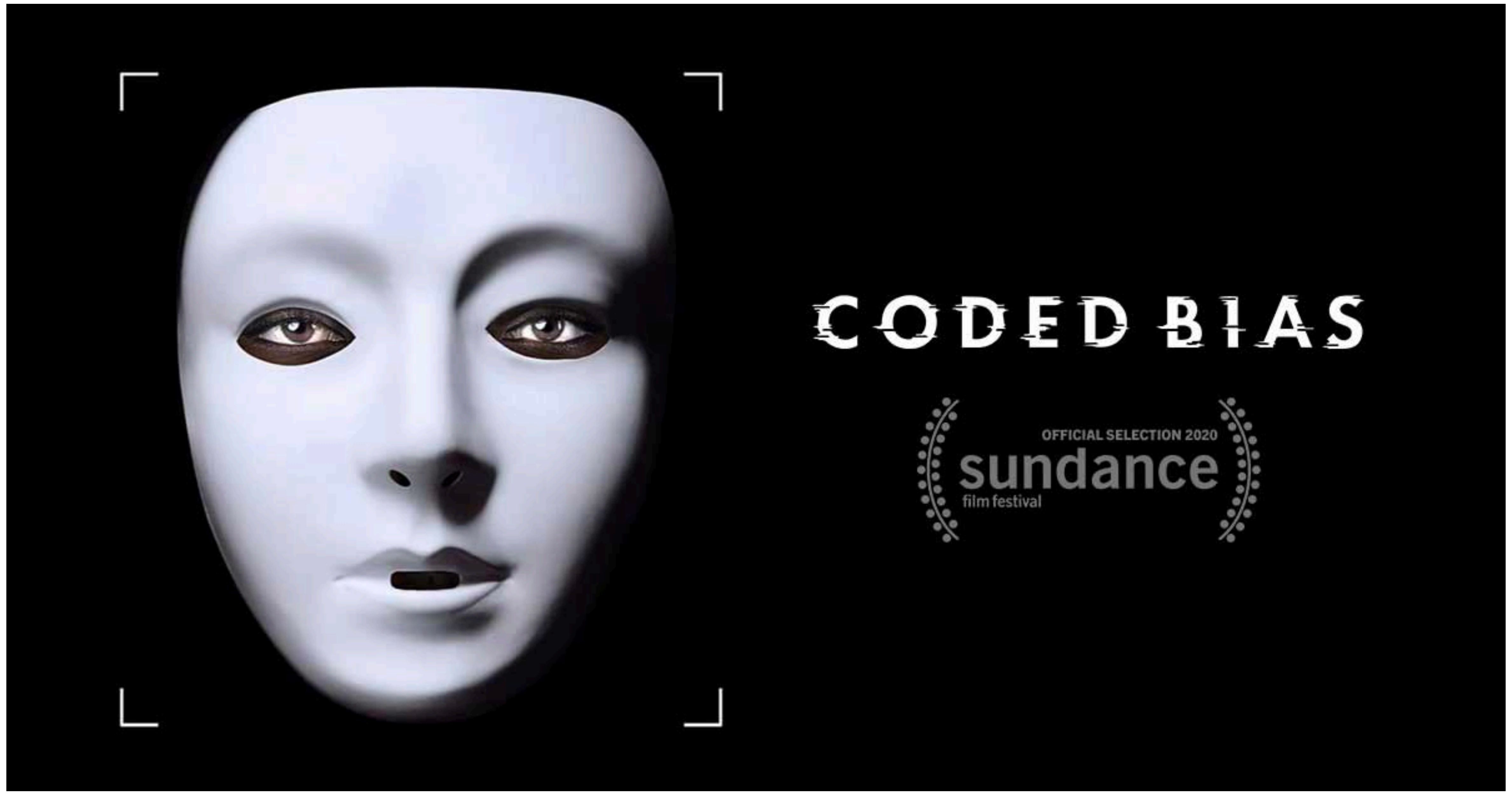
Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*

Joy Buolamwini JOYAB@MIT.EDU
MIT Media Lab 75 Amherst St. Cambridge, MA 02139

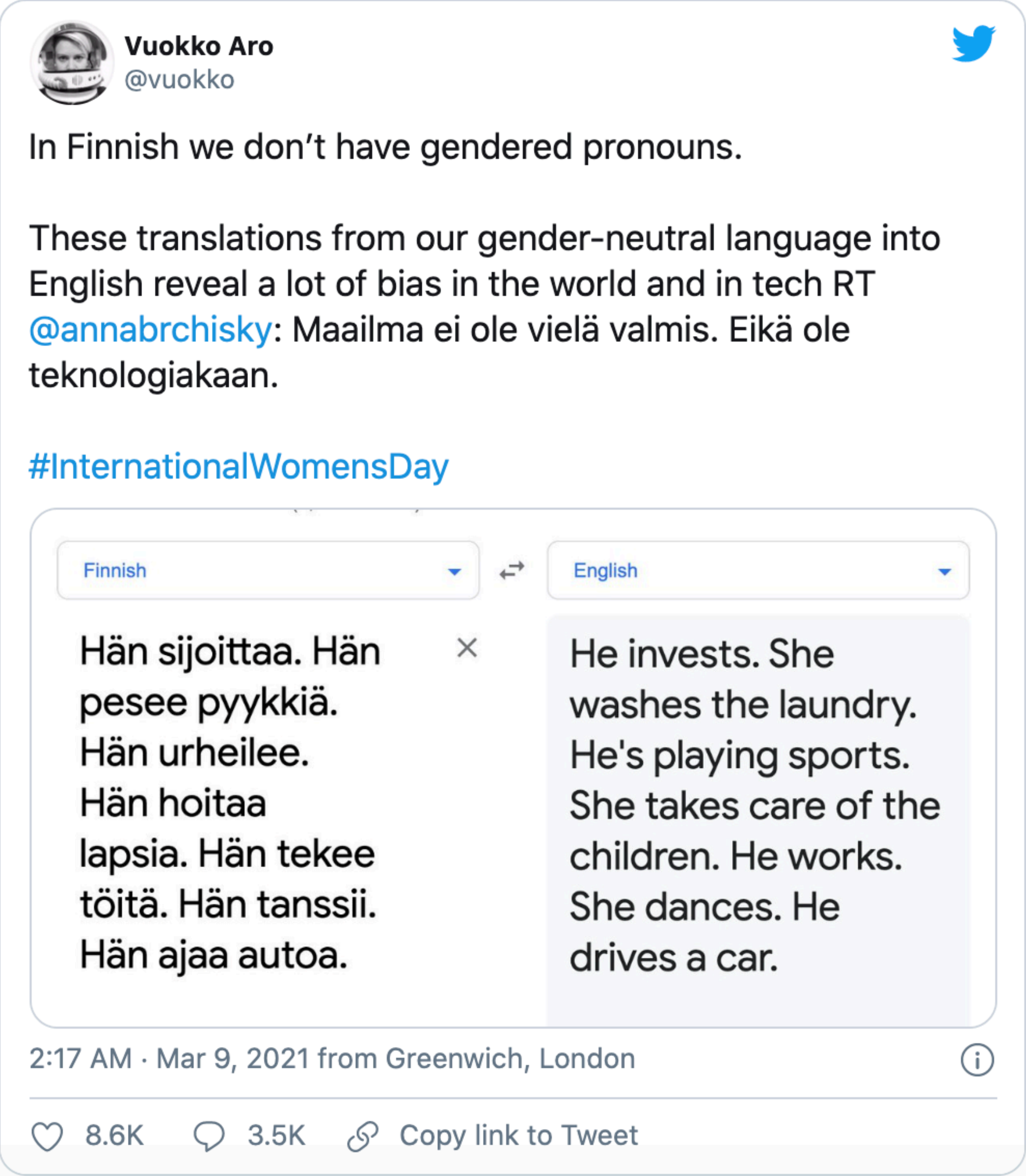
Timnit Gebru TIMNIT.GEBRU@MICROSOFT.COM
Microsoft Research 641 Avenue of the Americas, New York, NY 10011

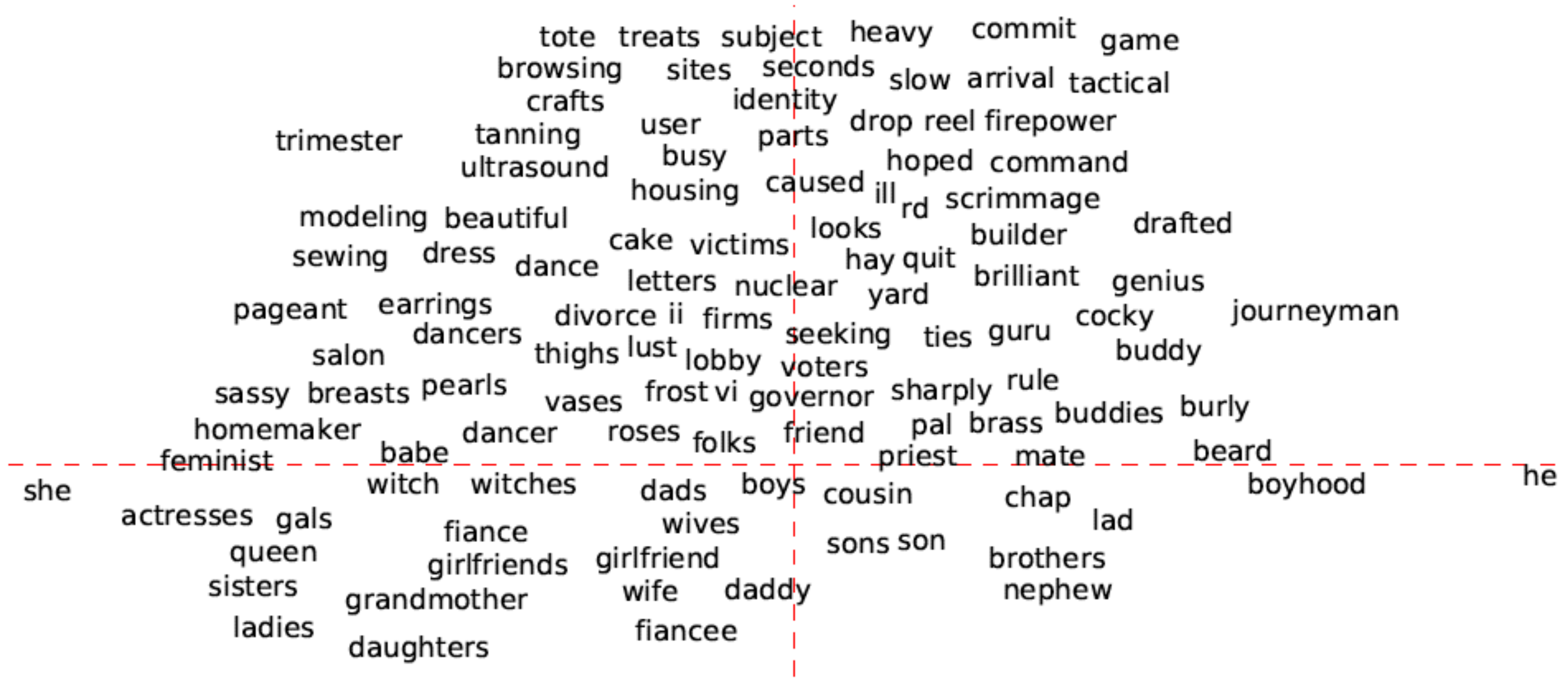
Gender Classifier	Darker Male	Darker Female	Lighter Male	Lighter Female	Largest Gap
Microsoft	94.0%	79.2%	100%	98.3%	20.8%
FACE++	99.3%	65.5%	99.2%	94.0%	33.8%
IBM	88.0%	65.3%	99.7%	92.9%	34.4%





Bias in Language Models





$$\overrightarrow{\mathrm{man}} - \overrightarrow{\mathrm{woman}} \approx \overrightarrow{\mathrm{computer\ programmer}} - \overrightarrow{\mathrm{homemaker}}.$$

Bolukbasi, Tolga, et al. "Man is to computer programmer as woman is to homemaker? debiasing word embeddings." *arXiv preprint arXiv:1607.06520* (2016).

Bias in Training Data

- On the web: race and gender stereotypes reinforced
    - Results for "CEO" in Google Images: 11% female, US 27% female CEOs
    - Also in Google Images, "doctors" are mostly male, "nurses" are mostly female
    - Google search results for professional vs. unprofessional hairstyles for work
-

Image results: "Unprofessional hair for work"

Image results: "Professional hair for work"

M. Kay, C. Matuszek, S. Munson (2015): [Unequal Representation and Gender Stereotypes in Image Search Results for Occupations](). CHI'15.

Discrimination in Data

Data as a social mirror

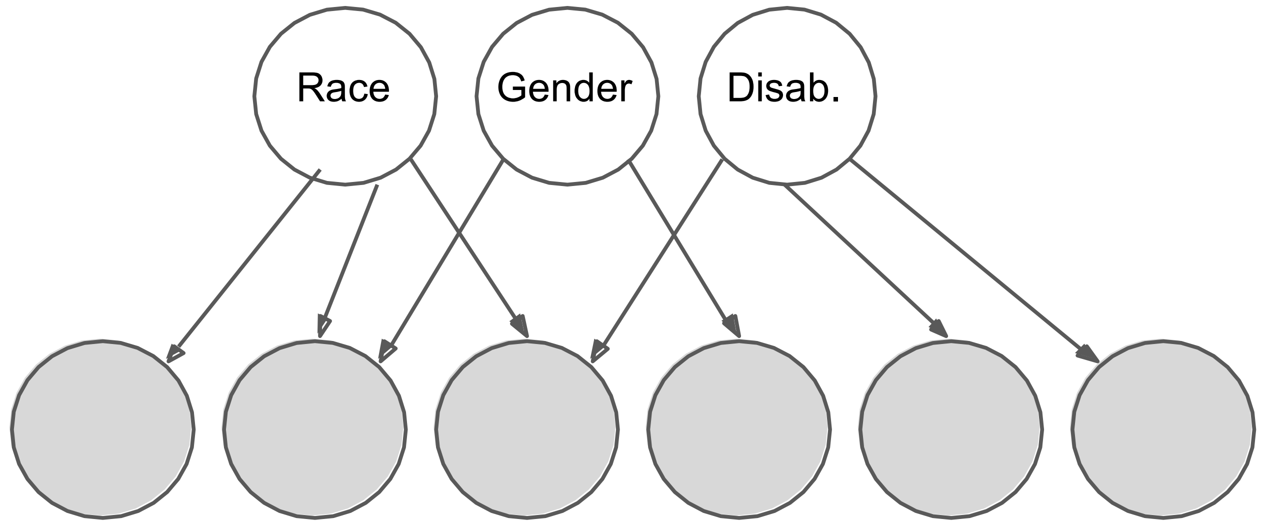
- Protected attributes redundantly encoded in observables
- Gender, race, social constructs; proxies used

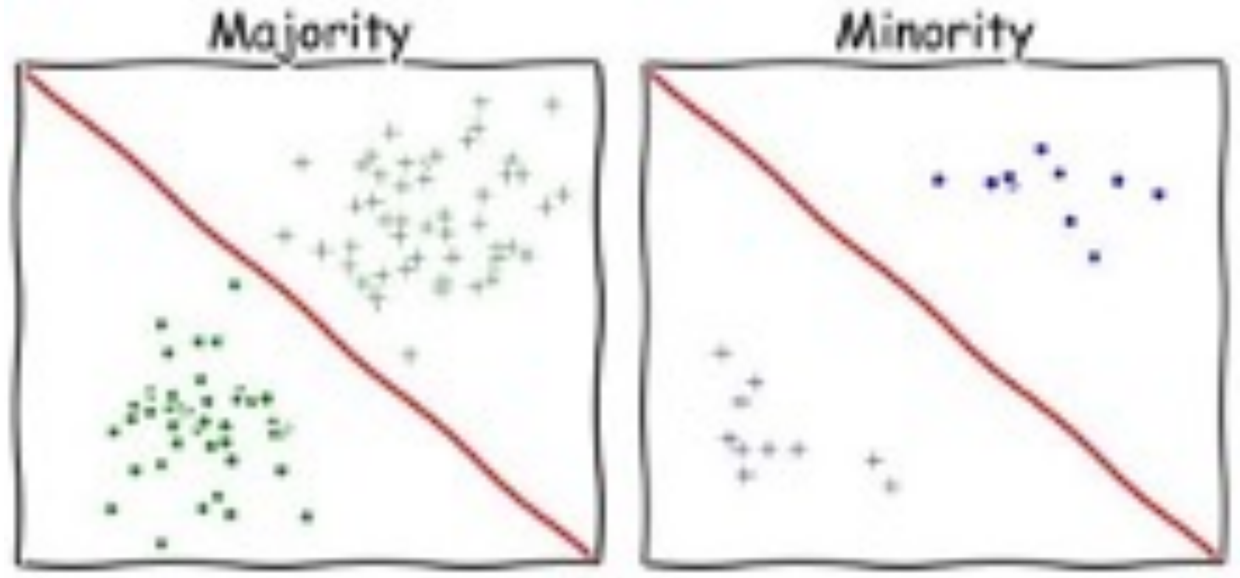
Correctness and completeness

Garbage in, garbage out (GIGO)

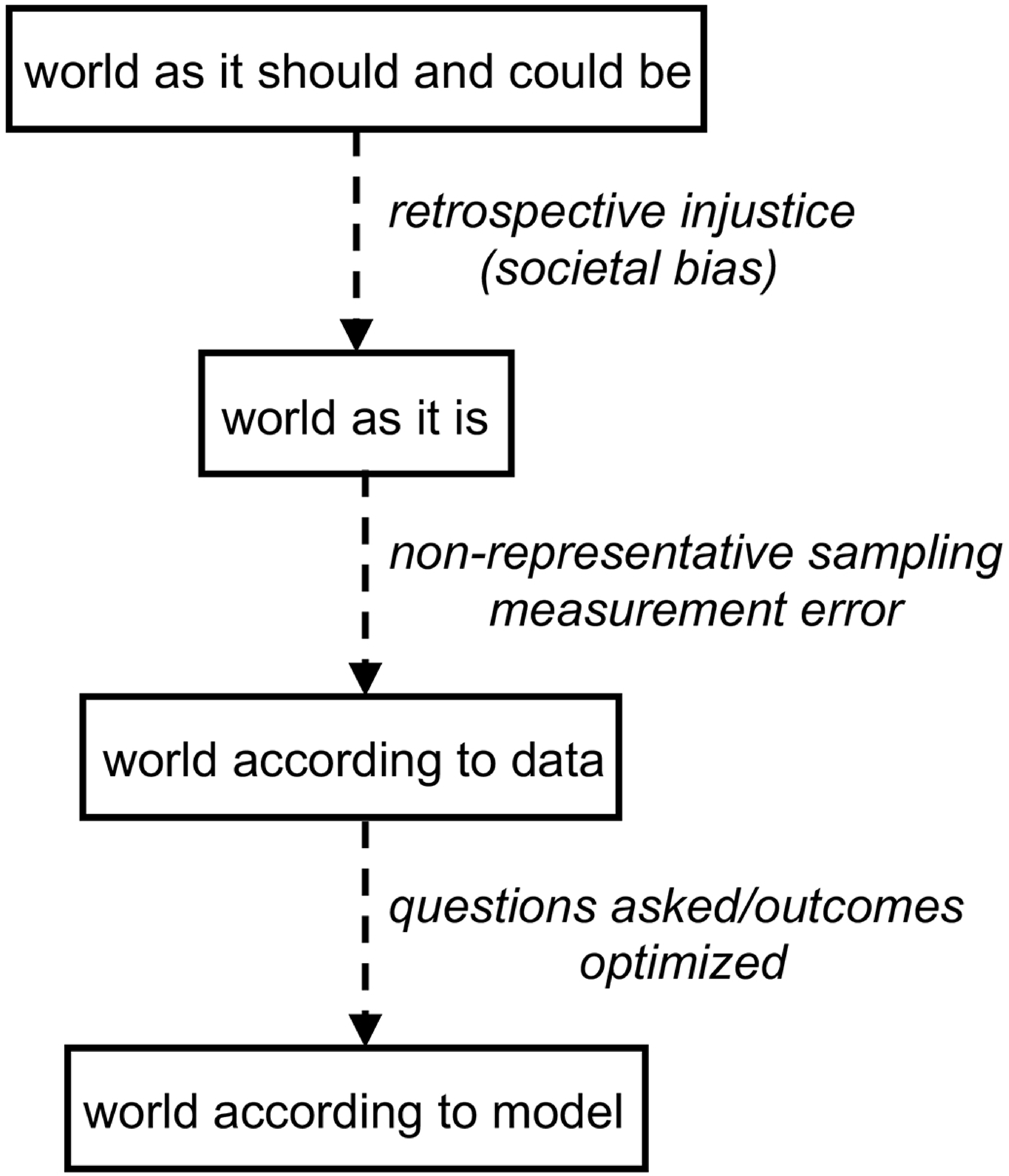
Sample size disparity: learn on majority

Errors concentrated in the minority class



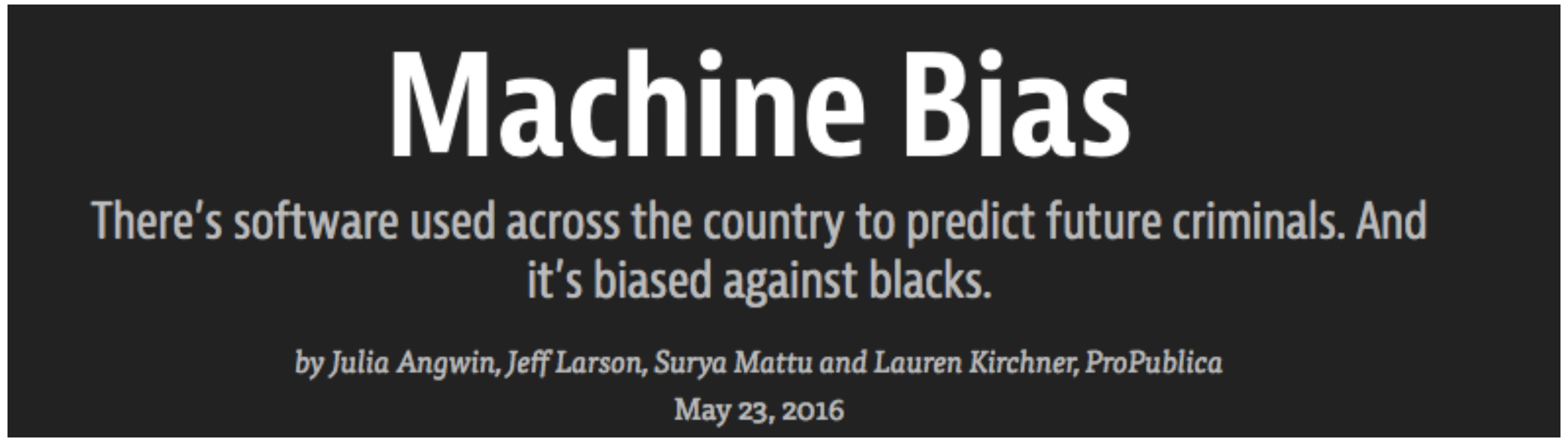


When data is about people, bias in data and algorithms can lead to discrimination



Mhasawade, V., Zhao, Y. and Chunara, R., 2021. Machine learning and algorithmic fairness in public and population health. *Nature Machine Intelligence*, *3*(8), pp.659-666.

Racial bias in criminal sentencing



Bernard Parker, left, was rated high risk; Dylan Fugett was rated low risk. (Josh Ritchie for ProPublica)

A commercial tool COMPAS automatically predicts some categories of future crime to assist in bail and sentencing decisions. It is used in courts in the US.

The tool correctly predicts **61% of the time.**

Black people are almost twice as likely as white people to be labeled a higher risk but not re-offend.

The tool makes **the opposite mistake on white people:** They are much more likely than black people to be labeled lower risk but go on to commit other crimes.

https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Racial bias in criminal sentencing

Machine Bias

There's software used across the country to predict future criminals. And it's biased against blacks.

by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica

May 23, 2016

Prediction Fails Differently for Black Defendants

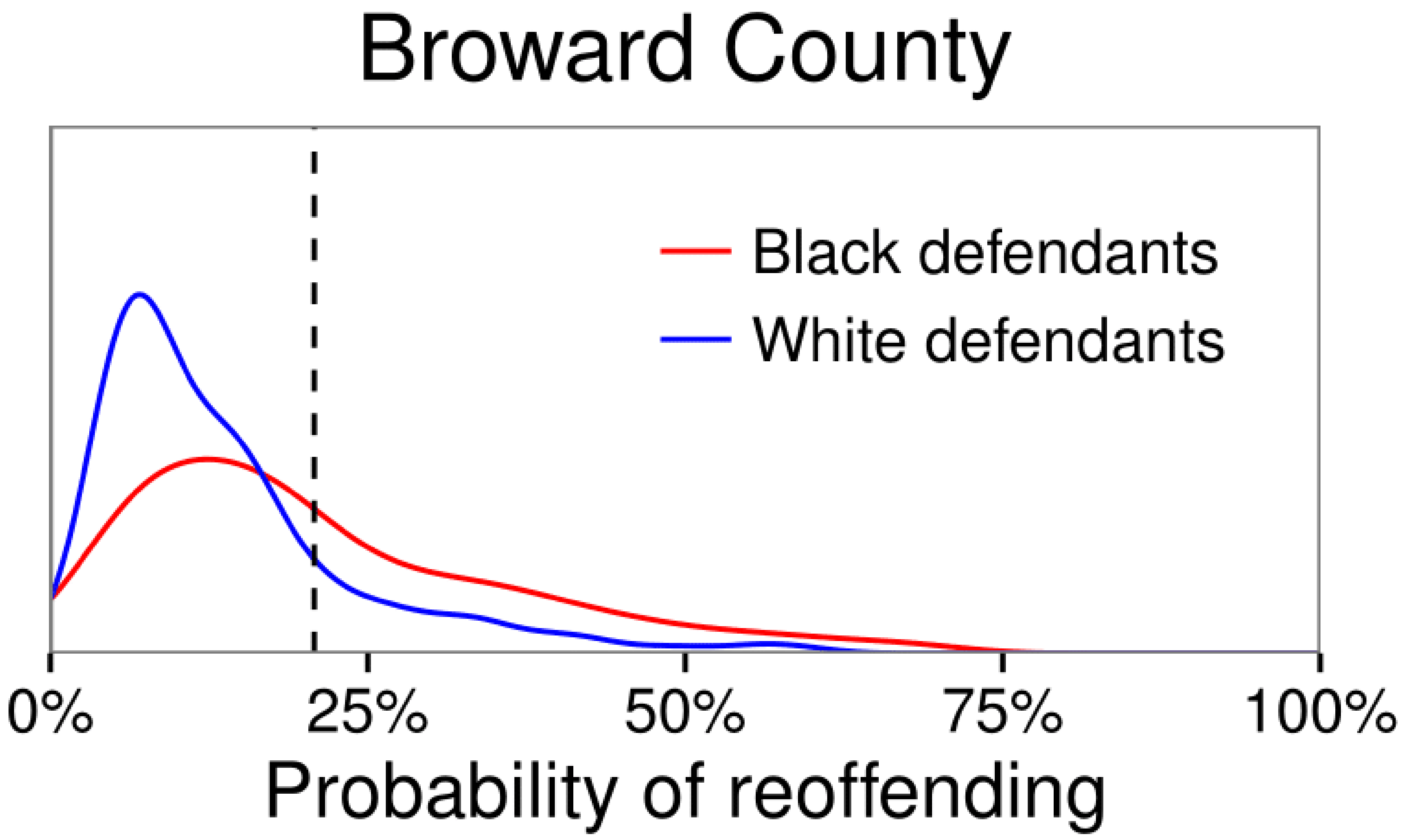
	WHITE	AFRICAN AMERICAN
Labeled Higher Risk, But Didn't Re-Offend	23.5%	44.9%
Labeled Lower Risk, Yet Did Re-Offend	47.7%	28.0%

Overall, Northpointe's assessment tool correctly predicts recidivism 61 percent of the time. But blacks are almost twice as likely as whites to be labeled a higher risk but not actually re-offend. It makes the opposite mistake among whites: They are much more likely than blacks to be labeled lower risk but go on to commit other crimes. (Source: ProPublica analysis of data from Broward County, Fla.)

https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

COMPAS as a predictive instrument

COMPAS is **well-calibrated**: in the window around 40%, the fraction of defendants who were re-arrested is ~40%, both over-all and per group.



Corbett-Davies, Sam, et al. "Algorithmic decision making and the cost of fairness." Proceedings of the 23rd acm sigkdd international conference on knowledge discovery and data mining. 2017.

theguardian

Samuel Gibbs

Wednesday 8 July 2015 11.29 BST

Women less likely to be shown ads for high-paid jobs on Google, study shows

Automated testing and analysis of company's advertising system reveals male job seekers are shown far more adverts for high-paying executive jobs

One experiment showed that Google displayed adverts for a career coaching service for executive jobs 1,852 times to the male group and only 318 times to the female group. Photograph: Alamy

The AdFisher tool simulated job seekers that did not differ in browsing behavior, preferences or demographic characteristics, except in gender.

One experiment showed that Google displayed ads for a career coaching service for "$200k+" executive jobs **1,852 times to the male group and only 318 times to the female group**. Another experiment, in July 2014, showed a similar trend but was not statistically significant.

https://www.theguardian.com/technology/2015/jul/08/women-less-likely-ads-high-paid-jobs-google-study