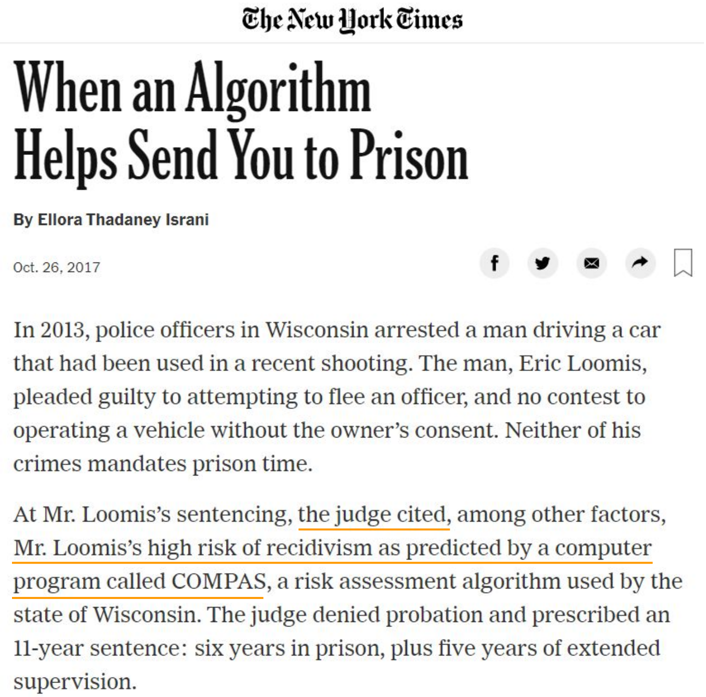
Special Topic: Fairness in Machine Learning

Chao Lan

Machine learning is increasingly applied in sensitive areas.

Fairness matters in recidivism prediction.



The New York Times

When an Algorithm Helps Send You to Prison

By Ellora Thadaney Israni

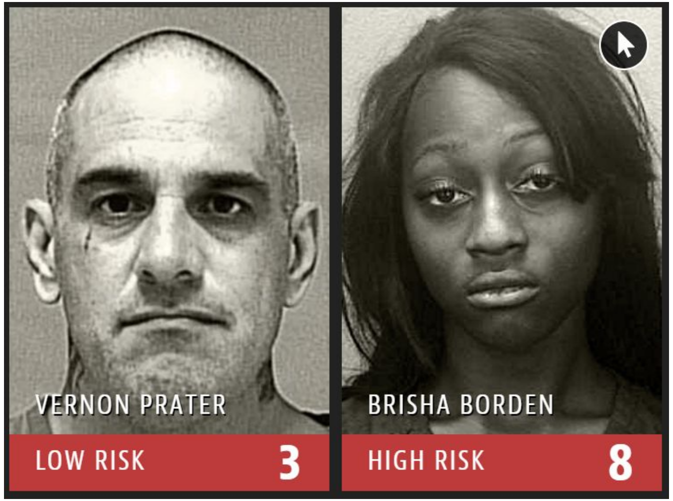
Oct. 26, 2017

In 2013, police officers in Wisconsin arrested a man driving a car that had been used in a recent shooting. The man, Eric Loomis, pleaded guilty to attempting to flee an officer, and no contest to operating a vehicle without the owner's consent. Neither of his crimes mandates prison time.

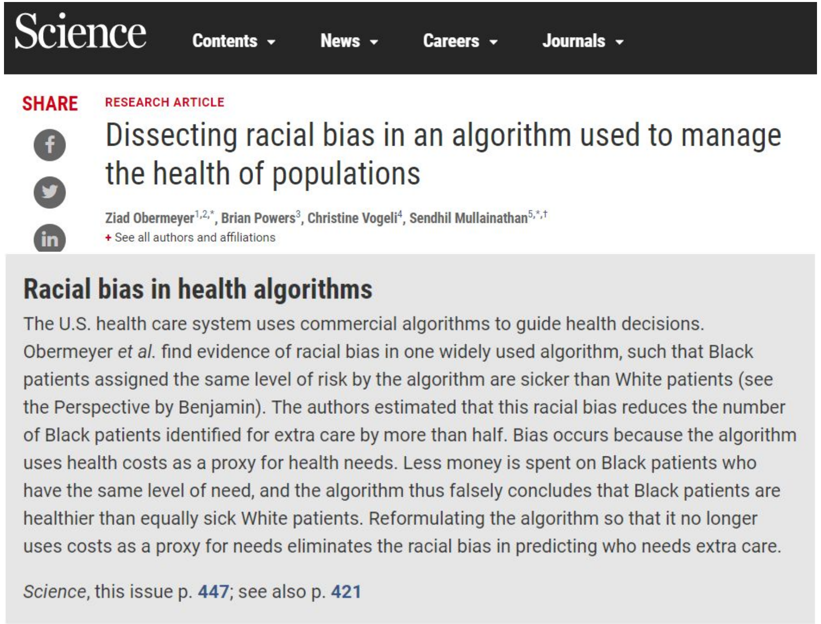
At Mr. Loomis's sentencing, the judge cited, among other factors, Mr. Loomis's high risk of recidivism as predicted by a computer program called COMPAS, a risk assessment algorithm used by the state of Wisconsin. The judge denied probation and prescribed an 11-year sentence: six years in prison, plus five years of extended supervision.

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

- *ProPublica*, 2016.



Fairness matters in auto health assessment.



RESEARCH ARTICLE

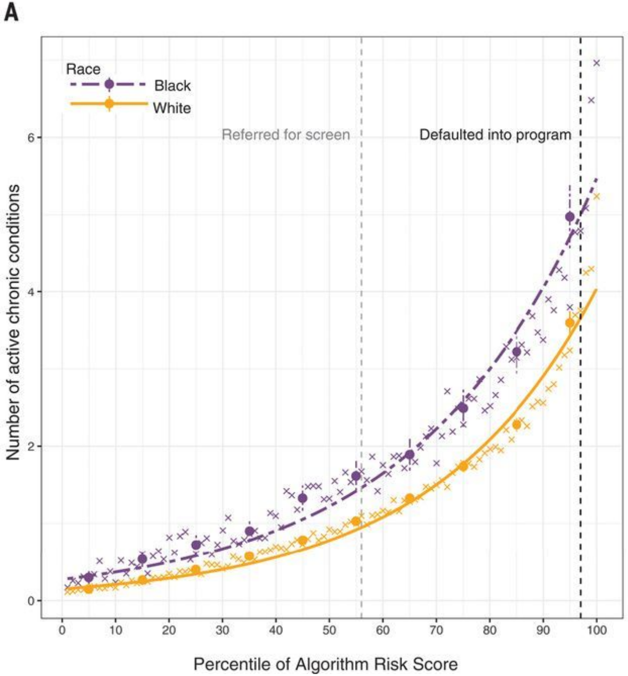
Dissecting racial bias in an algorithm used to manage the health of populations

Ziad Obermeyer[1,2,*], Brian Powers[3], Christine Vogeli[4], Sendhil Mullainathan[5,*,†]
+ See all authors and affiliations

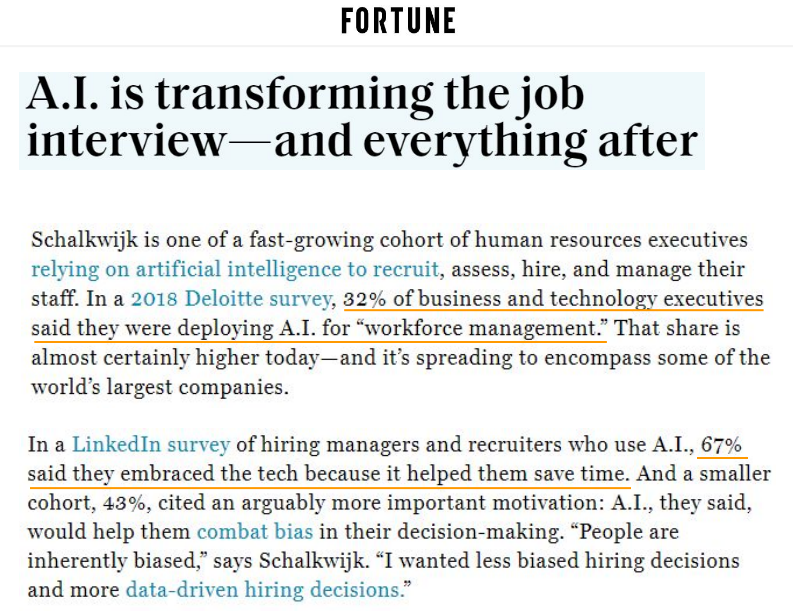
Racial bias in health algorithms

The U.S. health care system uses commercial algorithms to guide health decisions. Obermeyer *et al.* find evidence of racial bias in one widely used algorithm, such that Black patients assigned the same level of risk by the algorithm are sicker than White patients (see the Perspective by Benjamin). The authors estimated that this racial bias reduces the number of Black patients identified for extra care by more than half. Bias occurs because the algorithm uses health costs as a proxy for health needs. Less money is spent on Black patients who have the same level of need, and the algorithm thus falsely concludes that Black patients are healthier than equally sick White patients. Reformulating the algorithm so that it no longer uses costs as a proxy for needs eliminates the racial bias in predicting who needs extra care.

Science, this issue p. 447; see also p. 421



Fairness matters in auto job hiring.



FORTUNE

A.I. is transforming the job interview—and everything after

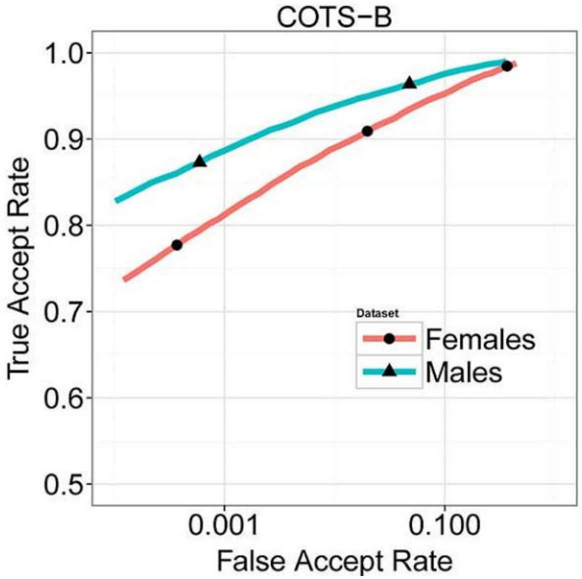
Schalkwijk is one of a fast-growing cohort of human resources executives relying on artificial intelligence to recruit, assess, hire, and manage their staff. In a 2018 Deloitte survey, 32% of business and technology executives said they were deploying A.I. for "workforce management." That share is almost certainly higher today—and it's spreading to encompass some of the world's largest companies.

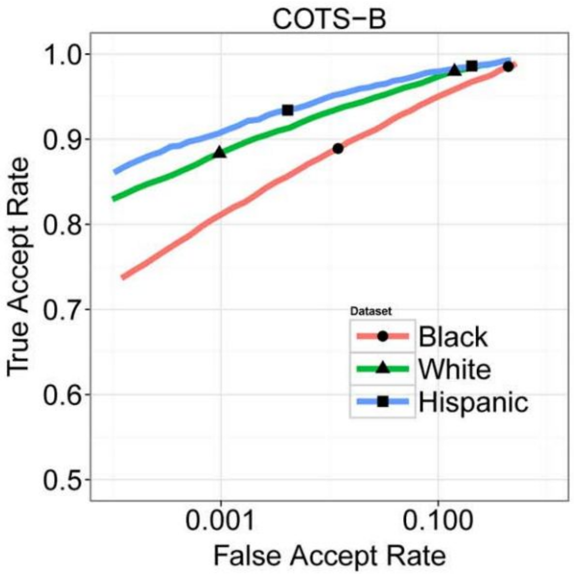
In a LinkedIn survey of hiring managers and recruiters who use A.I., 67% said they embraced the tech because it helped them save time. And a smaller cohort, 43%, cited an arguably more important motivation: A.I., they said, would help them combat bias in their decision-making. "People are inherently biased," says Schalkwijk. "I wanted less biased hiring decisions and more data-driven hiring decisions."

Amazon scraps secret AI recruiting tool that showed bias against women.

- *Reuters*, 2018.

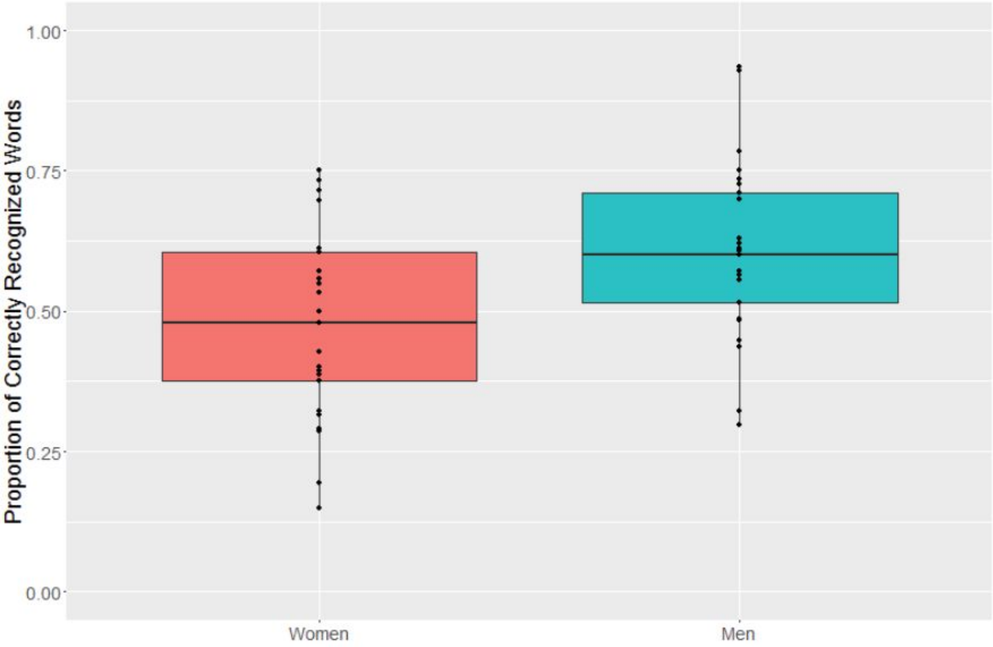
Other bias: commercial facial recognition.





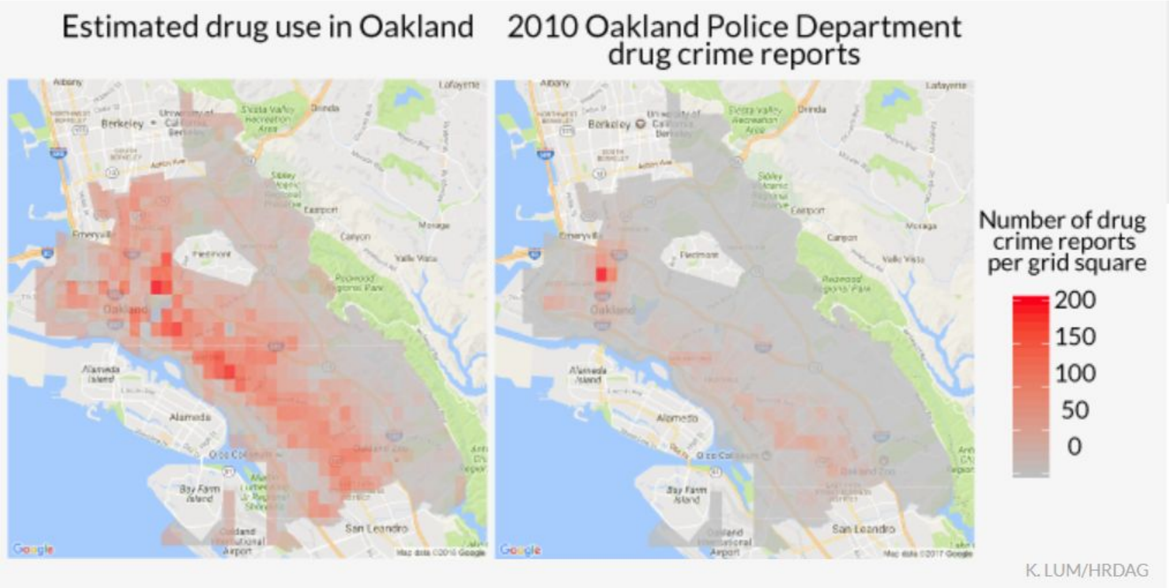
Klare et al. Face recognition performance: Role of demographic information. IEEE Trans. Information Forensics and Security, 2012.

Other bias: GoogleVoice.

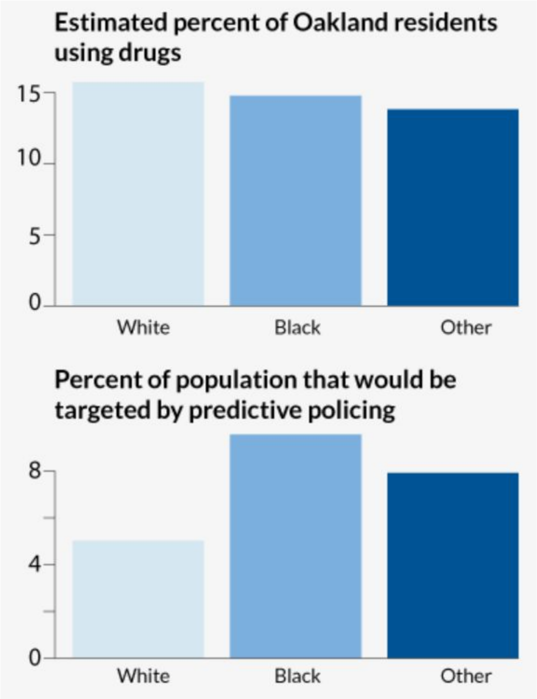


R. Tatman. Google’s speech recognition has a gender bias, 2016.

Other bias: drug use prediction.


Estimated drug use in Oakland
2010 Oakland Police Department drug crime reports
Number of drug crime reports per grid square
200
150
100
50
0
Berkeley
Emeryville
Piedmont
Oakland
Alameda
Alameda Island
Bay Farm Island
San Leandro
Oakland International Airport
Google

K. LUM/HRDAG



R. Ehrenberg. Data-driven crime prediction fails to erase human bias. Science News, 2017.

Other bias: word embedding techniques.

PRESIDENT
DIRECTOR
LEADER
CHIEF
EXECUTIVE
BOSS
SUBORDINATE
AIDE
HELPER
ASSISTANT
CO-WORKER
EMPLOYEE

W. Hutson. Even artificial intelligence can acquire biases against race and gender. Science, 2017.

Algorithmic fairness is a priority in the US AI R&D strategic plan (2016).

“ To avoid exacerbating biases by encoding them into technology systems, we need to develop a principle of ‘equal opportunity by design’ -- designing data systems that promote fairness and safeguard against discrimination from the first step of the engineering process and continuing throughout their lifespan. ”

Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights

Executive Office of the President

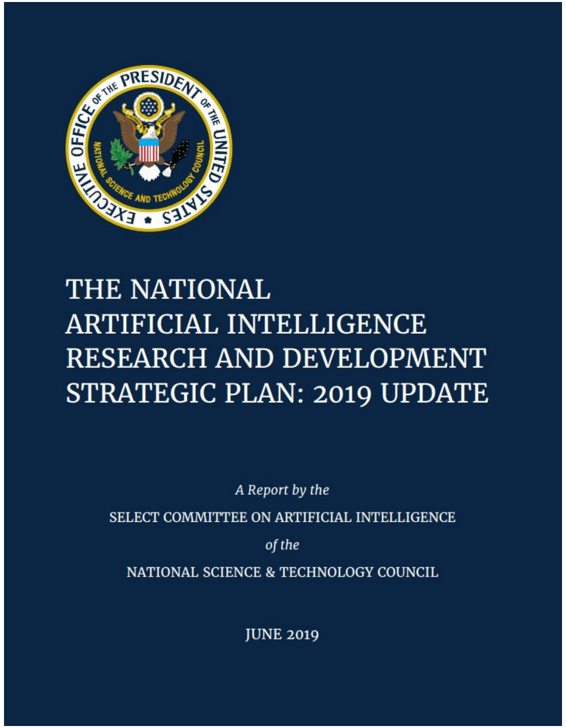
May 2016

And it remains a priority in the 2019 update.

“Beyond purely data-related issues, however, larger questions arise about the design of AI to be inherently just, fair, transparent, and accountable.

…

Scientists must also study to what extent justice and fairness considerations can be designed into the system, and how to accomplish this within the bounds of current engineering techniques.”



EXECUTIVE OFFICE OF THE PRESIDENT OF THE UNITED STATES

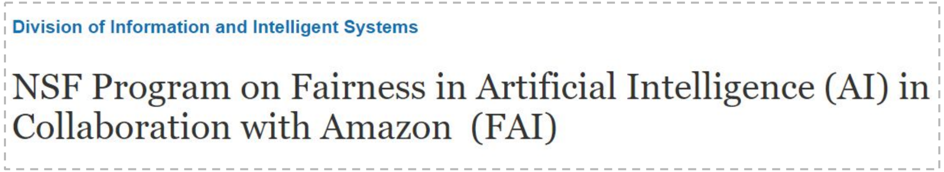
NATIONAL SCIENCE AND TECHNOLOGY COUNCIL

THE NATIONAL
ARTIFICIAL INTELLIGENCE
RESEARCH AND DEVELOPMENT
STRATEGIC PLAN: 2019 UPDATE

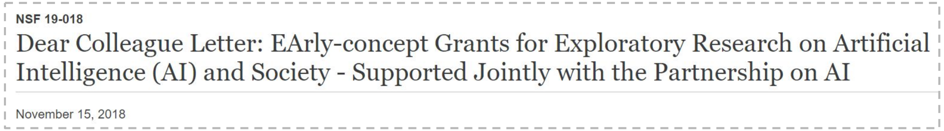
A Report by the
SELECT COMMITTEE ON ARTIFICIAL INTELLIGENCE
of the
NATIONAL SCIENCE & TECHNOLOGY COUNCIL

JUNE 2019

Many initiatives on fairness research.

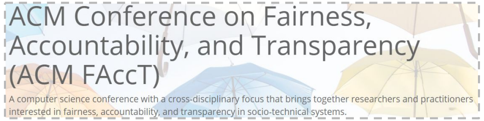

Division of Information and Intelligent Systems

NSF Program on Fairness in Artificial Intelligence (AI) in Collaboration with Amazon (FAI)

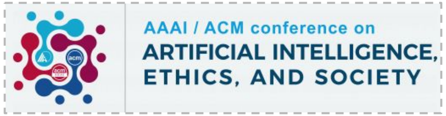

NSF 19-018

Dear Colleague Letter: EArly-concept Grants for Exploratory Research on Artificial Intelligence (AI) and Society - Supported Jointly with the Partnership on AI

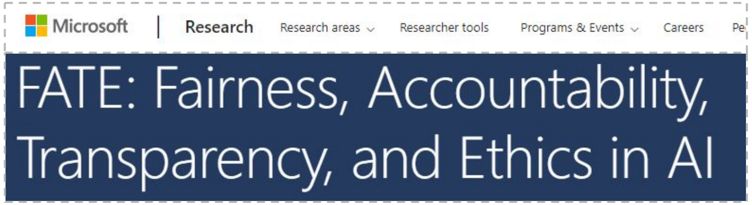
November 15, 2018


ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT)

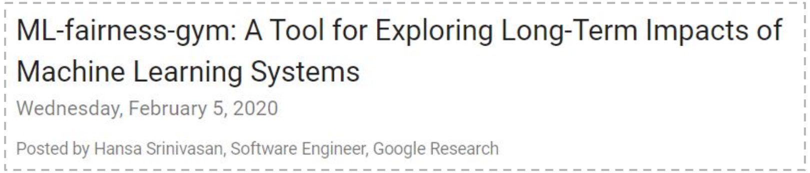
A computer science conference with a cross-disciplinary focus that brings together researchers and practitioners interested in fairness, accountability, and transparency in socio-technical systems.


AAAI / ACM conference on ARTIFICIAL INTELLIGENCE, ETHICS, AND SOCIETY

many workshops, papers, articles...

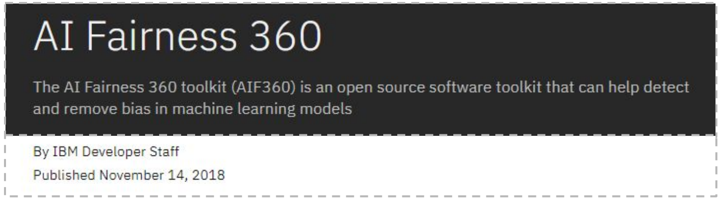


FATE: Fairness, Accountability, Transparency, and Ethics in AI


ML-fairness-gym: A Tool for Exploring Long-Term Impacts of Machine Learning Systems

Wednesday, February 5, 2020

Posted by Hansa Srinivasan, Software Engineer, Google Research


AI Fairness 360

The AI Fairness 360 toolkit (AIF360) is an open source software toolkit that can help detect and remove bias in machine learning models

By IBM Developer Staff
Published November 14, 2018

What is “fairness” in algorithmic prediction?

Statistical Disparity means big gap between

- Pr{ f(x) = hire | x is male}
- Pr{ f(x) = hire | x is female}

Equalized error rates means similar

- Pr{ f(x) = hire | x is male & y = not hired}
- Pr{ f(x) = hire | x is female & y = not hired}

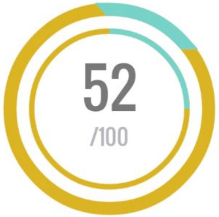
Individual Fairness means

- f(x) = f(z) if x and z are equally qualified

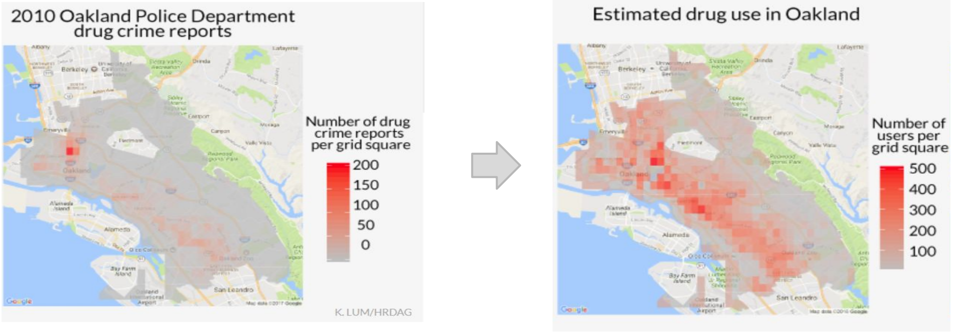
Amazon scraps secret AI recruiting tool that showed bias against women.

- *Reuters*, 2018.

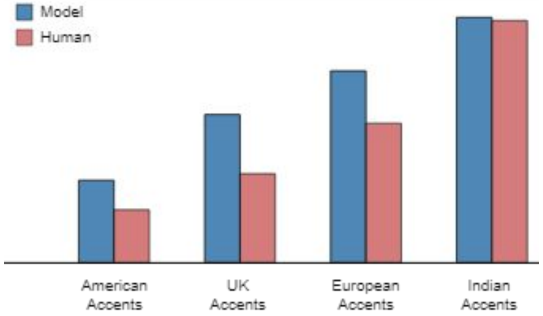
Where does algorithmic bias come from?



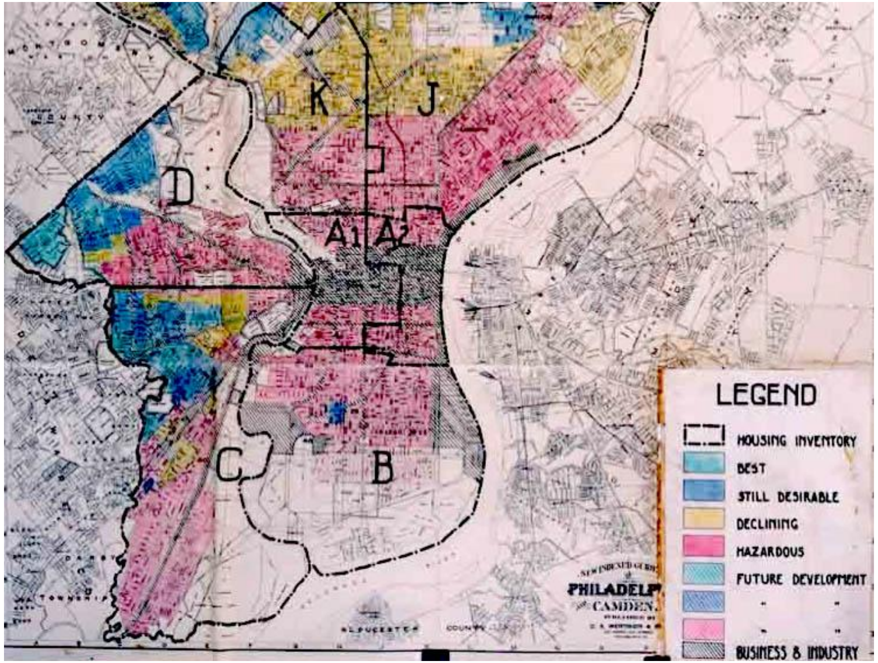
Women were 25% of the group and spoke 13% of the time.


2010 Oakland Police Department drug crime reports
Number of drug crime reports per grid square
200
150
100
50
0
Estimated drug use in Oakland
Number of users per grid square
500
400
300
200
100

K. LUM/HRDAG



Does hiding sensitive attribute help?



W. Norton. Cultural geography: Environments, landscapes, identities, inequalities. Oxford University Press, 2013.

Let’s design a fair learner for linear model.