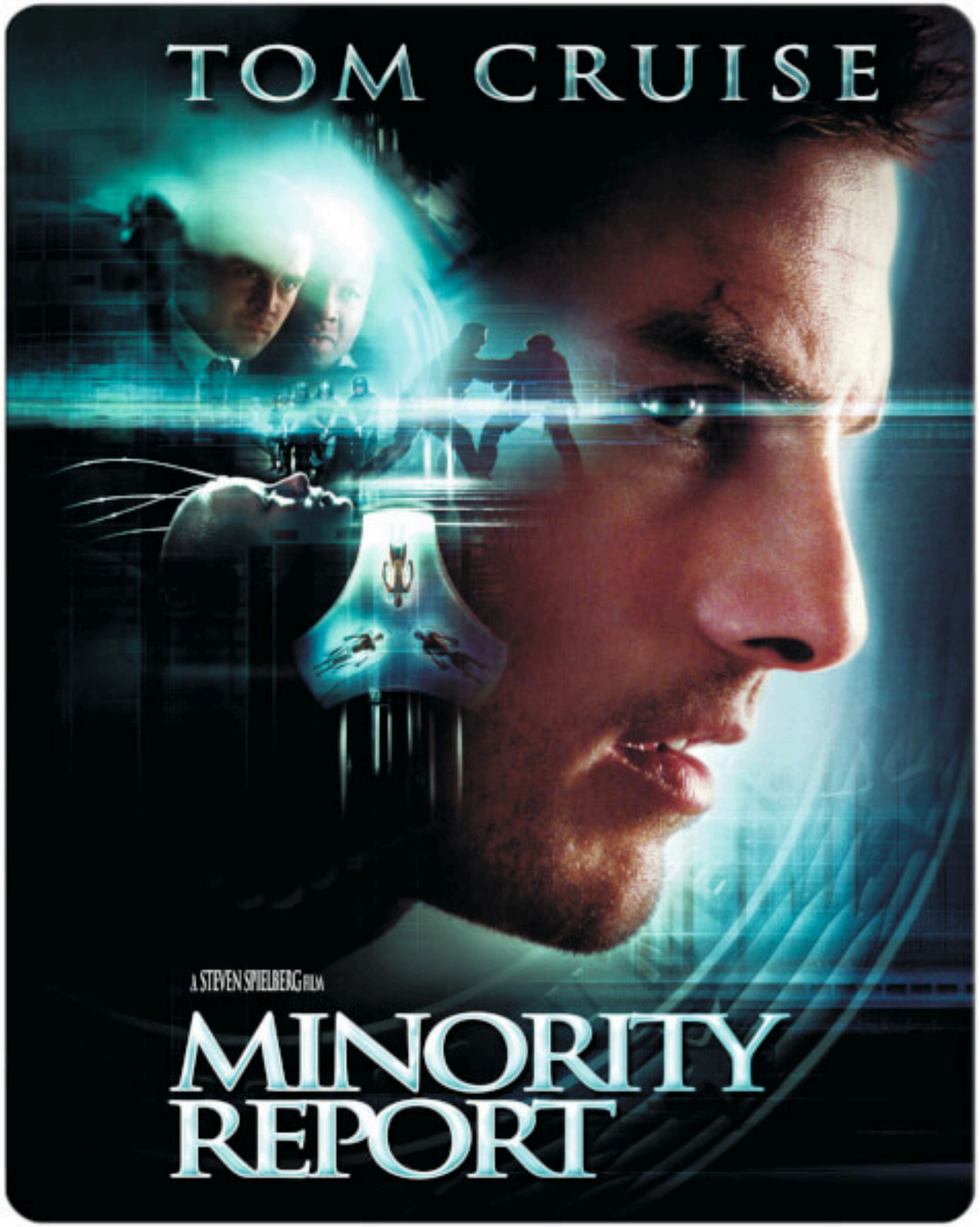
CS 520

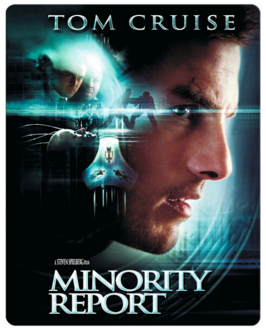
Theory and Practice of Software Engineering
Fall 2019

Software Fairness

September 5, 2019



TOM CRUISE
A STEVEN SPIELBERG FILM
MINORITY
REPORT



Resilient cities Cities

Predicting crime, LAPD-style

Cutting edge data-driven analysis directs Los Angeles patrol officers to likely future crime scenes – but critics worry that decision-making by machine will bring 'tyranny of the algorithm'

- Join our live Q&A with Homicide Watch this Friday

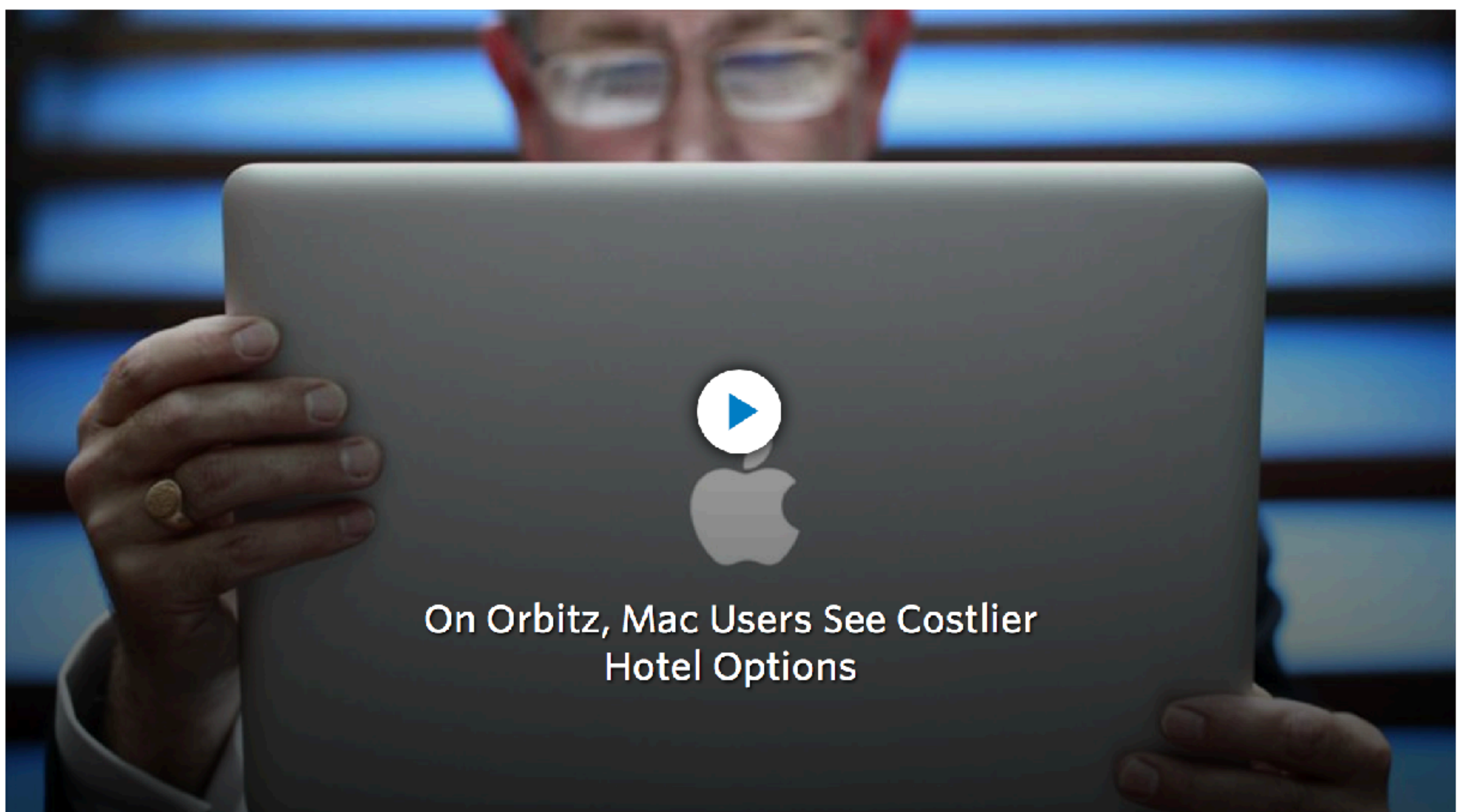
▲ PredPol co-developer P Jeffrey Brantingham at the Unified Command Post in Los Angeles. 'This is not Minority Report,' he said. Photograph: Damian Dovarganes/AP

https://www.theguardian.com/cities/2014/jun/25/predicting-crime-lapd-los-angeles-police-data-analysis-algorithm-minority-report

THE WALL STREET JOURNAL.

TECH

On Orbitz, Mac Users Steered to Pricier Hotels



Orbitz has found that Apple users spend as much as 30% more a night on hotels, so the online travel site is starting to show them different, and sometimes costlier, options than Windows visitors see. Dana Mattioli has details on The News Hub. Photo: Bloomberg.

By Dana Mattioli

MAR 15, 2011 @ 12:55 PM 16,947

The Algorithm That Beats Your Bank Manager

Minority homebuyers face widespread statistical lending discrimination, study finds

By Laura Counts | NOVEMBER 13, 2018

Face-to-face meetings between mortgage officers and homebuyers have been rapidly replaced by online applications and algorithms, but lending discrimination hasn't gone away.

A new University of California, Berkeley study has found that both online and face-to-face lenders charge higher interest rates to African American and Latino borrowers, earning 11 to 17 percent higher profits on such loans. All told, those homebuyers pay up to half a billion dollars more in interest every year than white borrowers with comparable credit scores do, researchers found.

The findings raise legal questions about the rise of statistical discrimination in the fintech era, and point to potentially widespread violations of U.S. fair lending laws, the researchers say. While lending discrimination has historically been caused by human prejudice, pricing disparities are increasingly the result of algorithms that use machine learning to target applicants who might shop around less for higher-priced loans.

"The mode of lending discrimination has shifted from human bias to algorithmic bias," said study co-author Adair Morse, a finance professor at UC Berkeley's Haas School of Business. "Even if the people writing the

EDITOR'S PICK

Wisconsin Supreme Court allows state to continue using computer program to assist in sentencing

KATELYN FERRAL | The Capital Times | kferral@madison.com | @katelynferral Jul 13, 2016

propublica.org

ProPublica

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

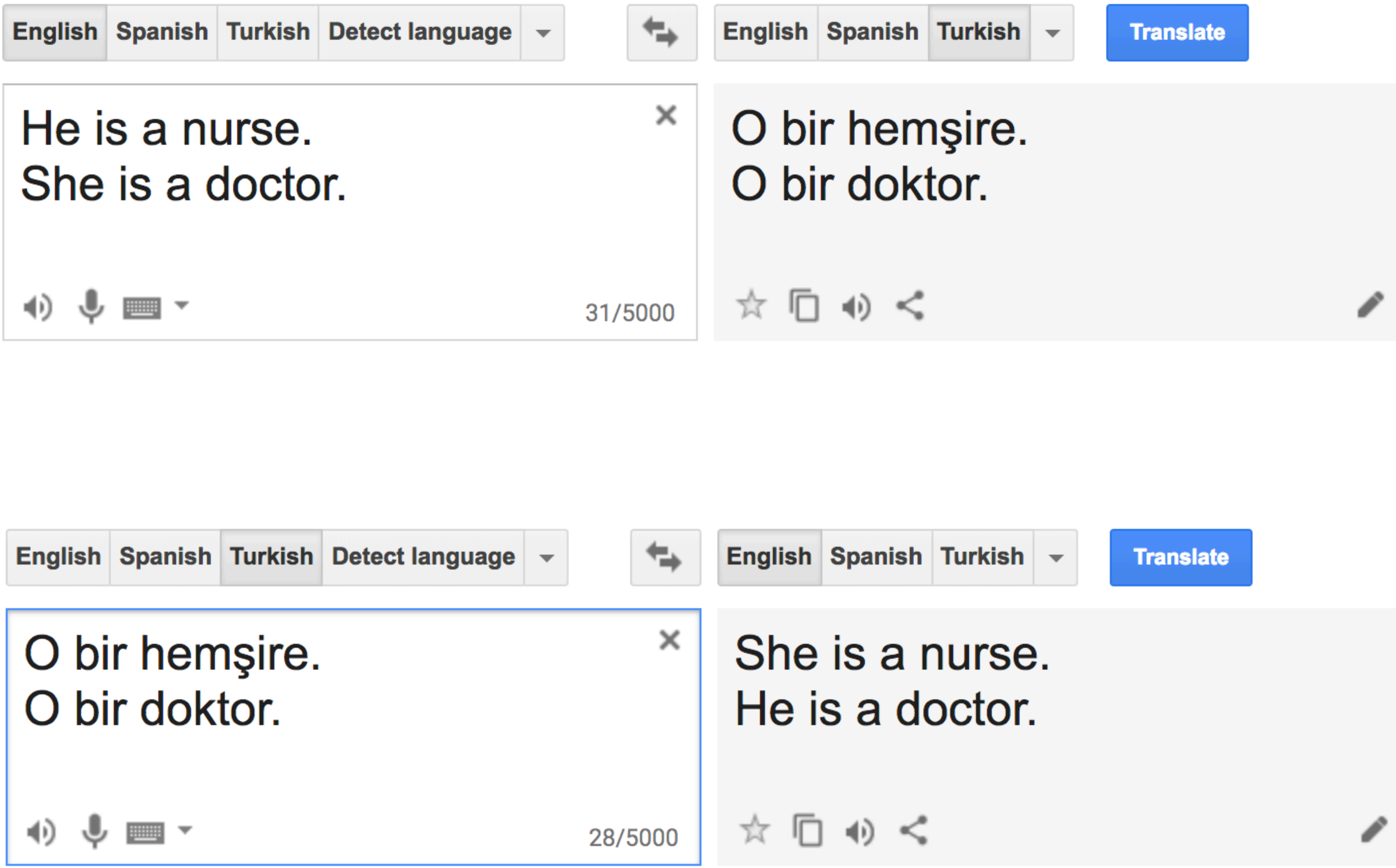
Software can make bad decisions.
Software can discriminate!

Just as the 18-year-old girls were realizing they were too big for the tiny conveyances — which belonged to a 6-year-old boy — a woman came running after them saying, "That's my kid's stuff." Borden and her friend immediately dropped the bike and scooter and walked away.

But it was too late — a neighbor who witnessed the heist had already called the police. Borden and her friend were arrested and charged with burglary and petty theft for the items, which were valued at a total of $80.

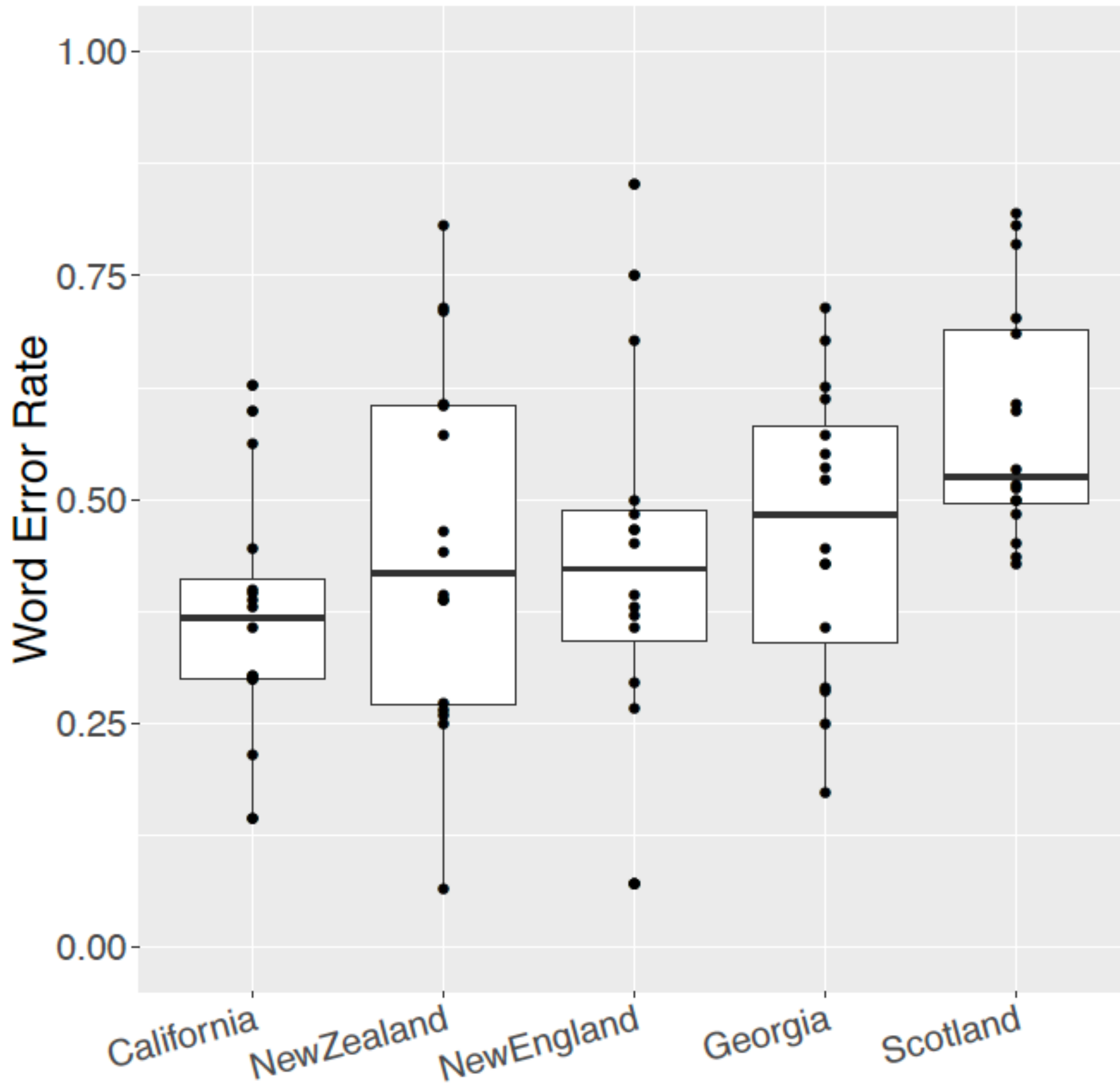
Compare their crime with a similar one: The

Subscribe to the Series


English Spanish Turkish Detect language
English Spanish Turkish
Translate
He is a nurse.
She is a doctor.
31/5000
O bir hemşire.
O bir doktor.
English Spanish Turkish Detect language
English Spanish Turkish
Translate
O bir hemşire.
O bir doktor.
28/5000
She is a nurse.
He is a doctor.

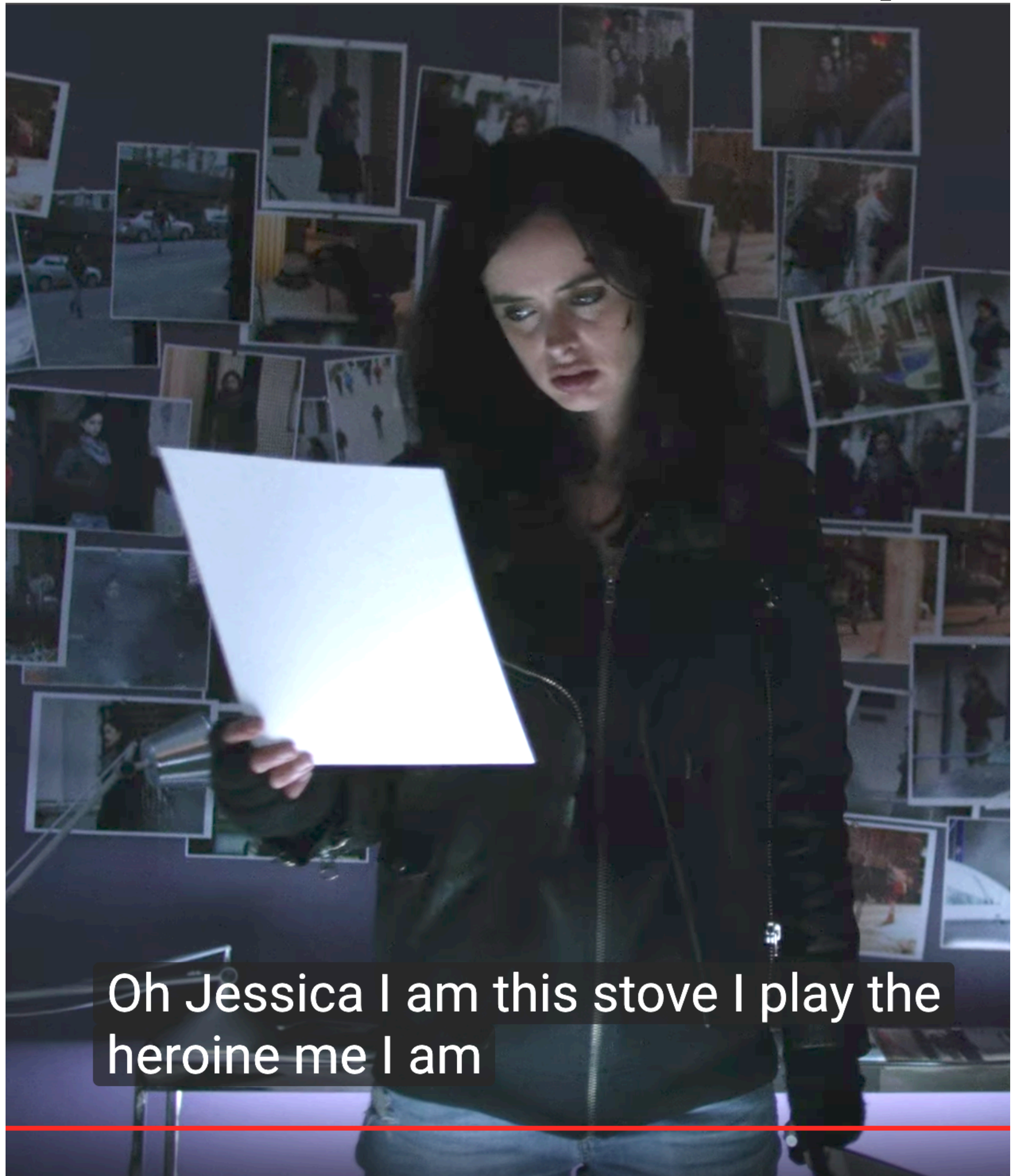
YouTube

YouTube automatic captions

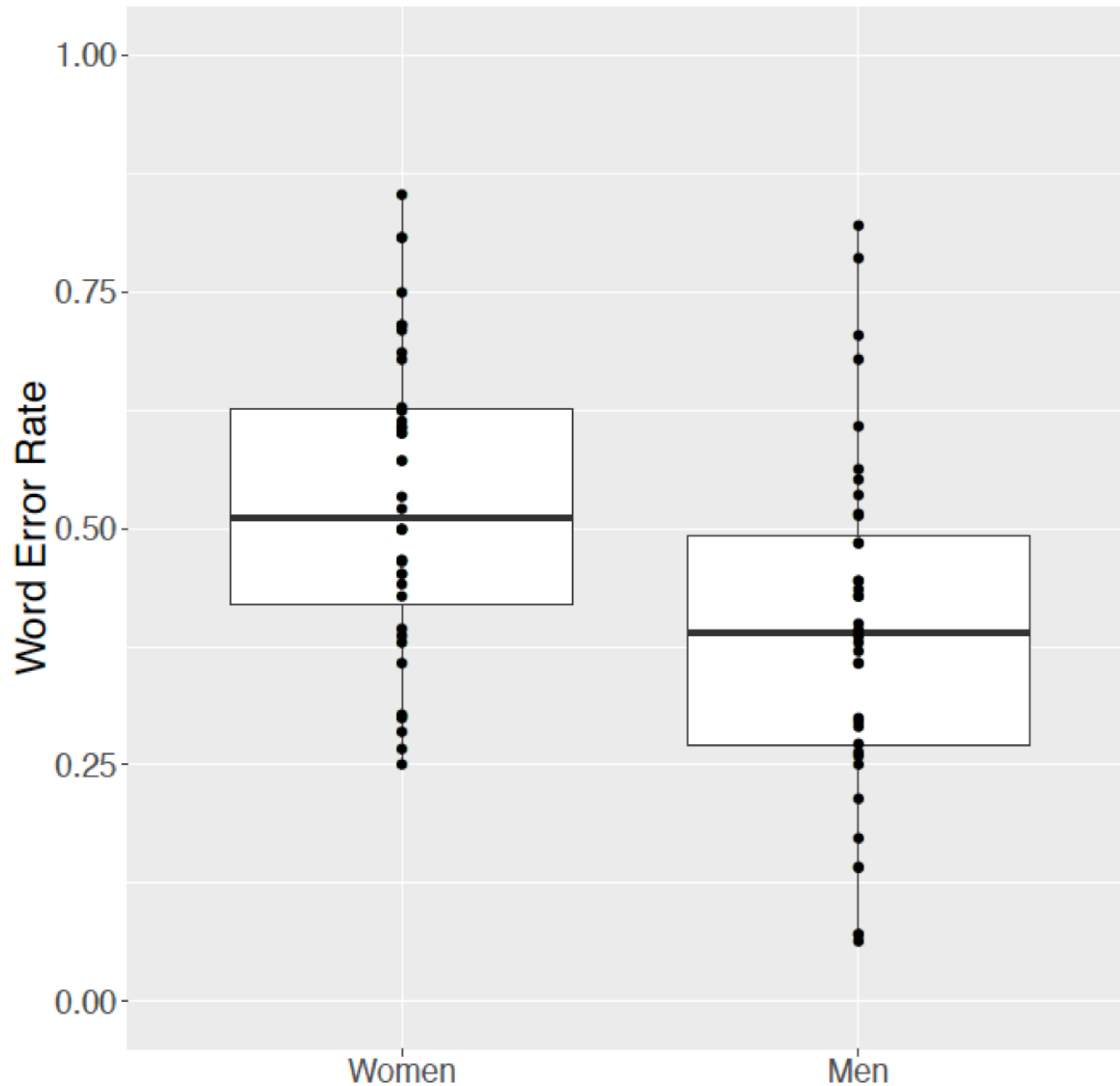


Rachael Tatman, "Gender and Dialect Bias in YouTube's Automatic Captions" in 2017 Workshop on Ethics in Natural Language Processing

YouTube automatic captions



YouTube automatic captions



Rachael Tatman, "Gender and Dialect Bias in YouTube's Automatic Captions" in 2017 Workshop on Ethics in Natural Language Processing

Joy Buolamwini
https://www.ted.com/talks/joy_buolamwini_how_i_m_fighting_bias_in_algorithms

how people want to use vision software

how people want to use vision software

Hire★Vue

SEE A HIREVUE ASSESSMENTS DEMO

WANT TO SEE HIREVUE IN ACTION?

SEE A DEMO

The predictive power of a traditional assessment in a video interview.

- Predictive assessments tied to higher quality hires
- Assess and interview in one step and hire faster
- Better candidate experience
- Validated [illegible]

JOIN 700+ CUSTOMERS THAT USE HIREV[illegible]E

vodafone TJX TIFFANY & CO.

GEICO DR PEPPER SNAPPLE GROUP intel

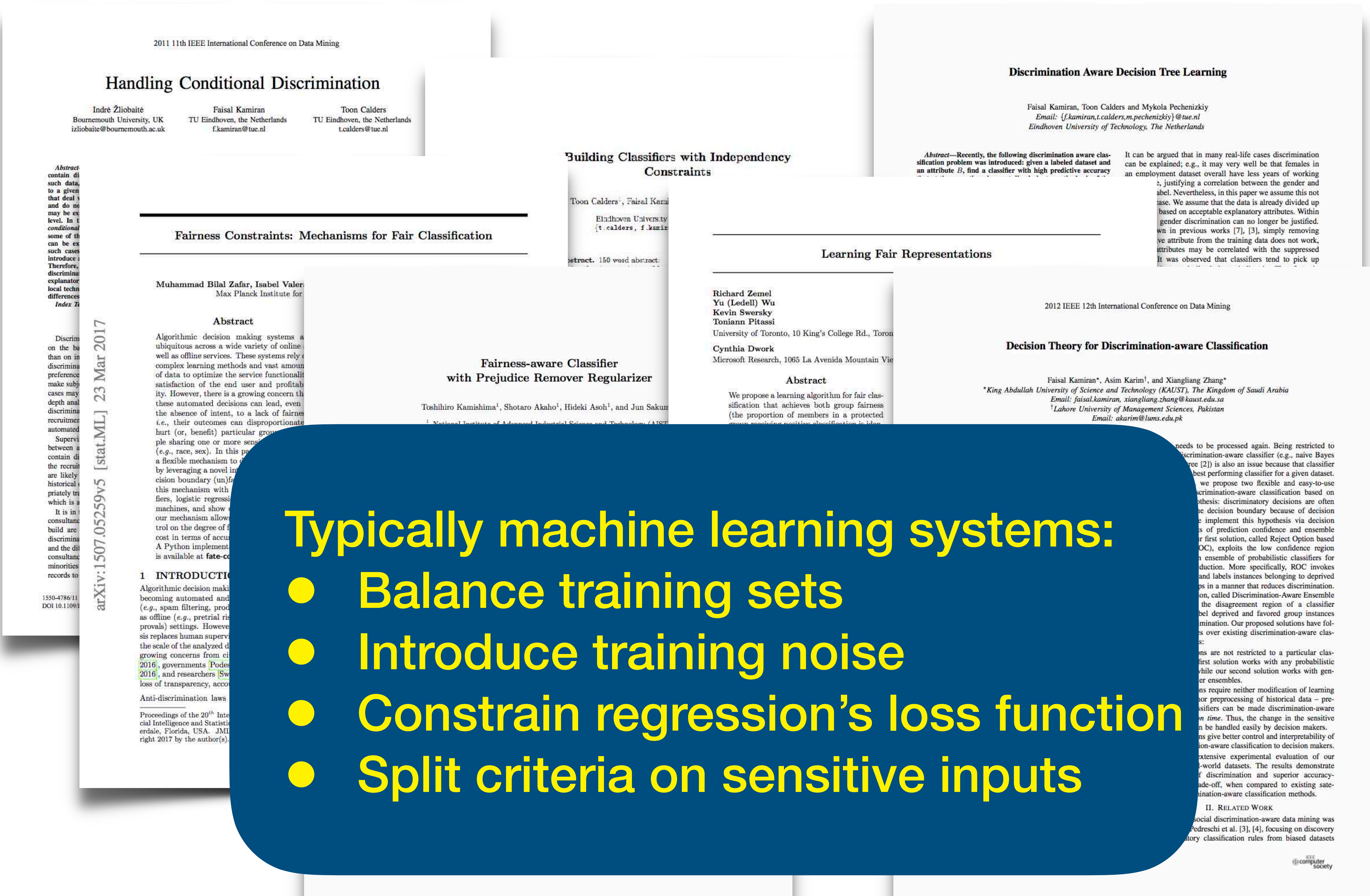
today's goals

Define software discrimination.

Operationalize measuring discrimination through causal software testing.

Provide provable fairness guarantees.

Design software to be fair



Typically machine learning systems:

- Balance training sets
- Introduce training noise
- Constrain regression's loss function
- Split criteria on sensitive inputs

Design alone is not enough

possible causes

biased data

poor design

implementation bugs

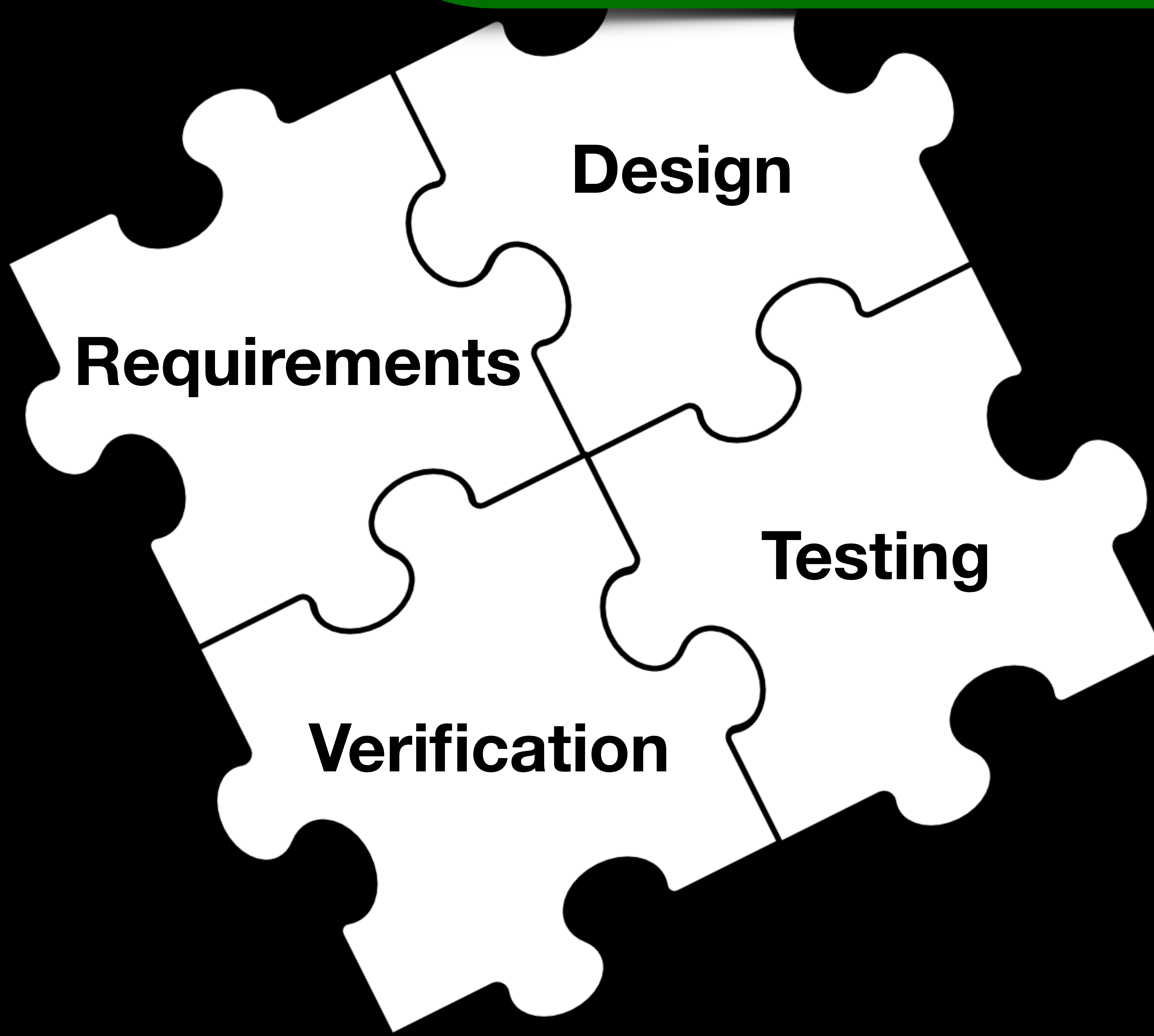
unintended interactions and mismatched components

Fairness is just like quality and security

Fairness must be part of the software engineering lifecycle

Call to Action!

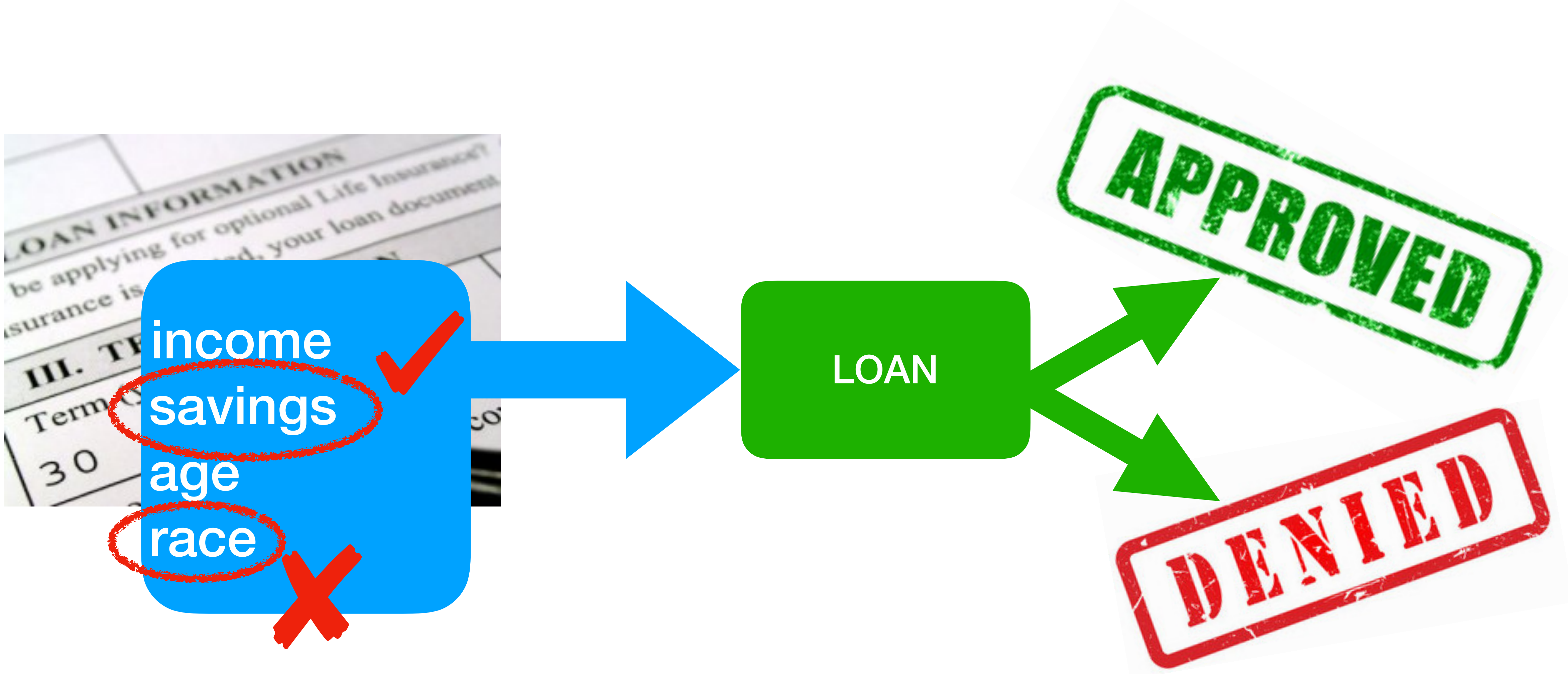
Fairness must be part of the software engineering lifecycle



Let’s talk about requirements.

What does it mean for software to discriminate?

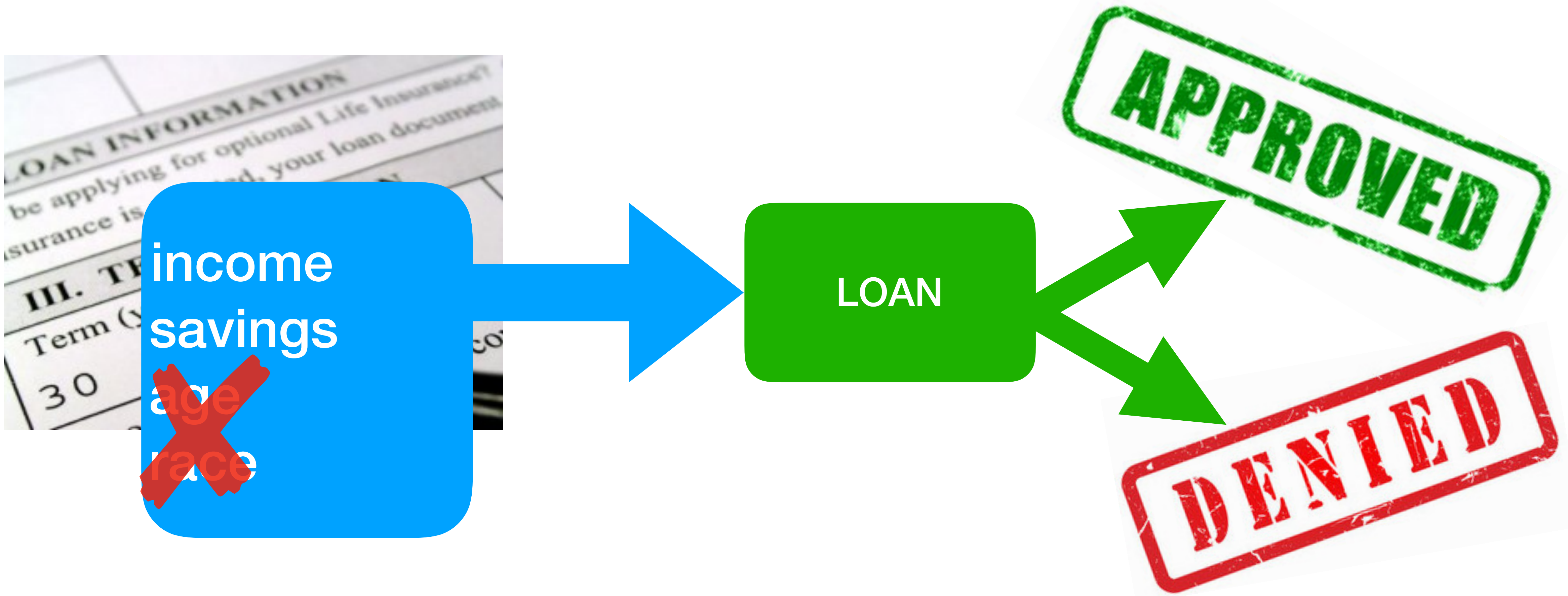
LOAN program



This talk is not about policy.

Fairness: Disparate Treatment

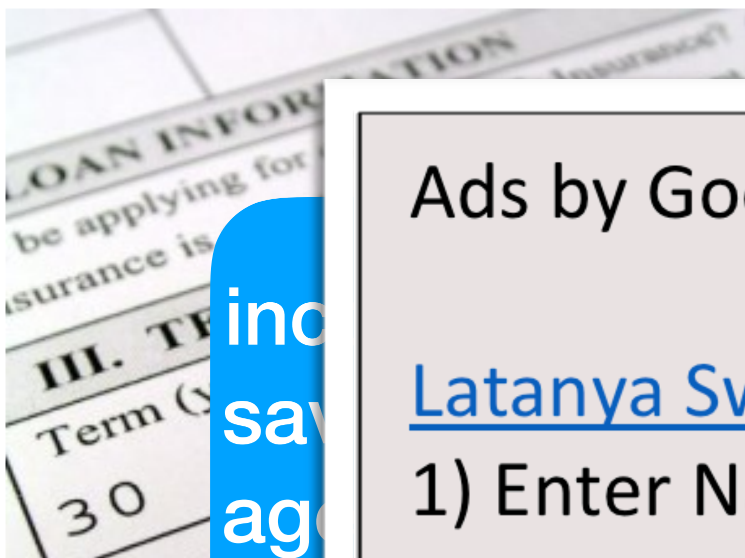
Hide the data

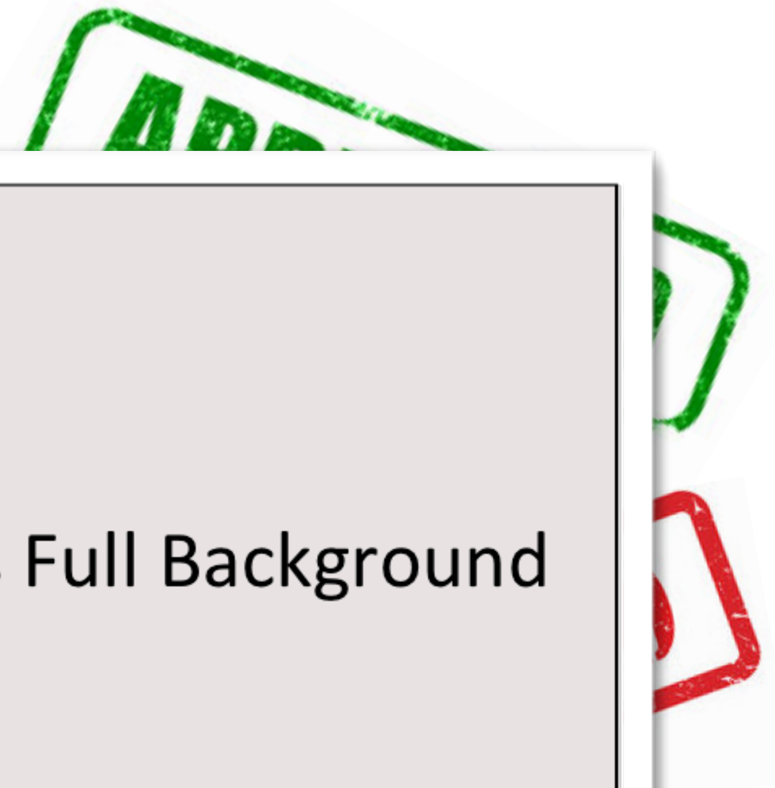


Zafar et al. Fairness constraints: Mechanisms for fair classification. AISTATS 2017.

Fairness: Disparate Treatment

Hide the data





Ads by Google

[Latanya Sweeney, Arrested?](www.instantcheckmate.com/)
1) Enter Name and State. 2) Access Full Background Checks Instantly.
www.instantcheckmate.com/

Ineffective because of data correlation.
[Latanya Sweeney. Discrimination in online ad delivery. CACM 2013]

businessinsider.com

BUSINESS INSIDER TECH NEWS

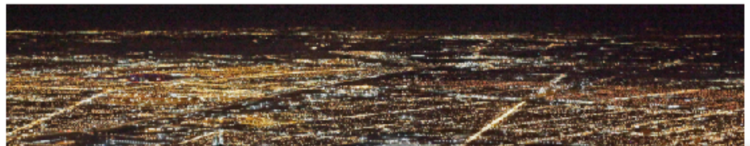
Amazon just showed us that 'unbiased' algorithms can be inadvertently racist

Rafi Letzter

Apr. 21, 2016, 4:50 PM 1,259

FACEBOOK LINKEDIN TWITTER

A Bloomberg report Thursday revealed that Amazon's same-day delivery service offered to Prime users around major US cities seems to routinely, if unintentionally, exclude black neighborhoods.



Chi
Blo

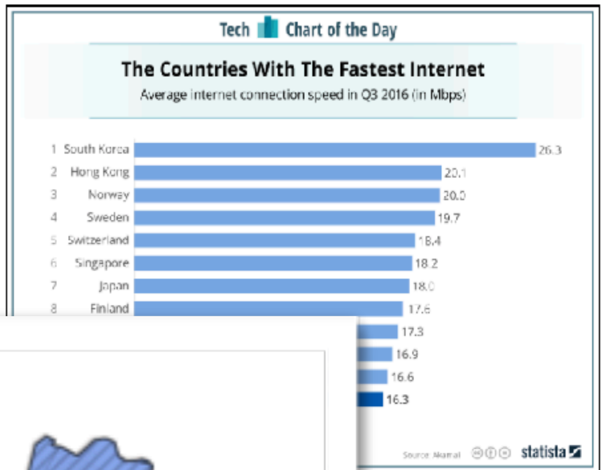
The maps, which you should check out on Bloomberg's site, show that in cities like Chicago, New York, and Atlanta, same-da
code at this point — except the m

But the thing is that Amazon's de
of PR Scott Stanzel wrote in an e

> There are a number of factor
> deliver same-day. Those incl
> center, local demand in an ar
> area, as well as the ability of
> to 9:00 pm every single day, even Sunday".

Recommended For You

'None of it makes much sense': Experts are baffled by Comey's use of a fake Russian document to skirt the DOJ



@BI_Enterprise
n Twitter, Steve
investing
Zj1STQJ
YHm9NbE
2 hours ago

@BI_Enterprise
Levie is taking a
Bezos' playbook as
ompany for th...
Vs9FYN
2 hours ago

Real-time market data. Get the latest on stocks, commodities, currencies, funds, rates, ETFs, and

BUSINESS INSIDER

TECH | FINANCE | POLITICS | STRATEGY | LIFE | ALL

PRIME | INTELLIGENCE

Amazon built an AI tool to hire people but had to shut it down because it was discriminating against women

Isobel Asher Hamilton 4h

- **Amazon tried building an artificial-intelligence tool to help with recruiting, but it showed a bias against women, Reuters reports.**
- **Engineers reportedly found the AI was unfavorable toward female candidates because it had combed through male-dominated résumés to accrue its data.**
- **Amazon reportedly abandoned the project at the beginning of 2017.**

Amazon wo

with hiring

disparate treatment: still not fair

https://www.businessinsider.com/amazon-built-ai-to-hire-people-discriminated-against-women-2018-10

Fairness: Demographic Parity

Compare subpopulation proportions

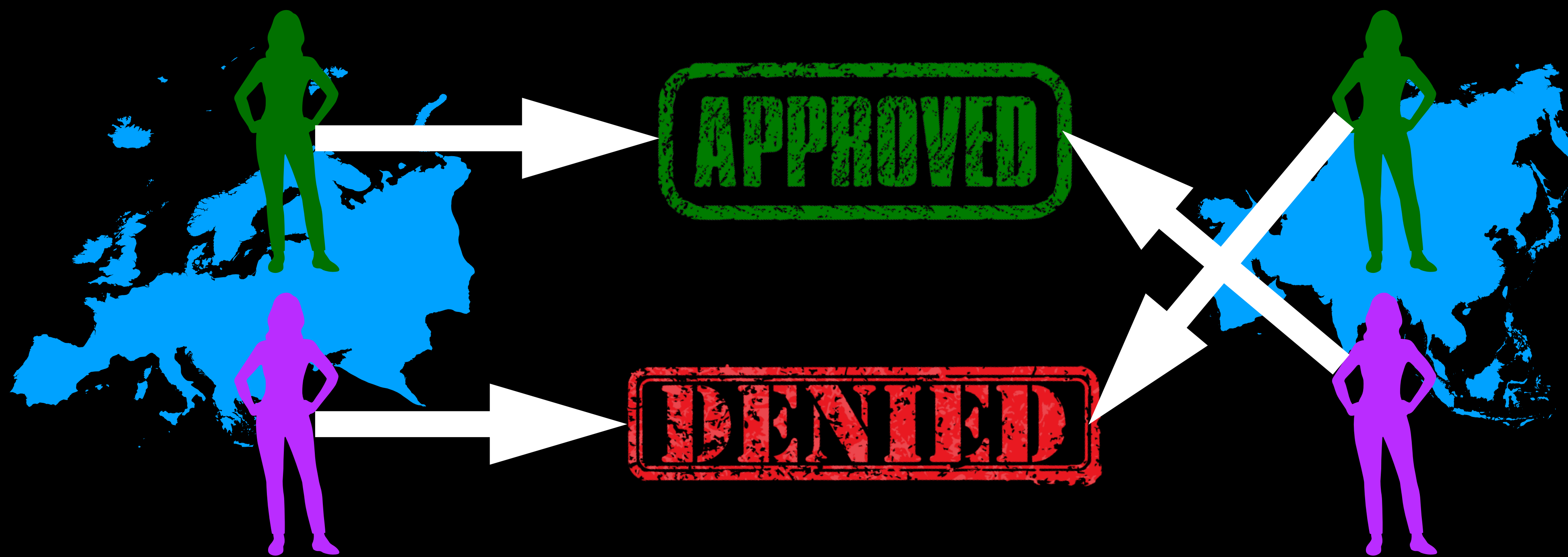
often called group discrimination

Fails to identify discrimination against individuals.

Dwork et al. Fairness through awareness. ITCS 2012.
Calders and Verwer. Three naive Bayes approaches for discrimination-free classification. DMKD 2010.

How group discrimination can fail

Europe Asia



approve loans to all **green** deny loans to all **purple** applicants

approve loans to all **purple** deny loans to all **green** applicants

European and Asian discriminations cancel each other out, and the group discrimination measure can be 0.

Fairness: Disparate Impact

Prohibits using a facially neutral practice that has an unjustified adverse impact on members of a protected class.

80% rule: Employer's hiring rates for protected groups may not differ by more than 80%.

Zafar et al. Fairness constraints: Mechanisms for fair classification. AISTATS 2017.

Fairness: Delayed Impact

Making seemingly fair decisions can (but shouldn't), in the long term, produce unfair consequences

Liu et al., Delayed impact of fair machine learning. ICML 2018

Fairness: Predictive Equality

False positive rates should not differ

Chouldechova. Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. FATML 2016
Corbett-Davies. Algorithmic decision making and the cost of fairness. KDD 2017

Fairness: Equal Opportunity

False negative rates should not differ

Hardt et al. Equality of Opportunity in Supervised Learning. NIPS 2016
Chouldechova. Fair prediction with disparate impact: A study of bias in recidivism prediction instruments FATML 2016

Fai rness: Equal dds

P quality tunity

earning. NIPS 2016

Fairn Equality

to

eg

Berk et al. Fairness in criminal justice rt. Sociolo 18

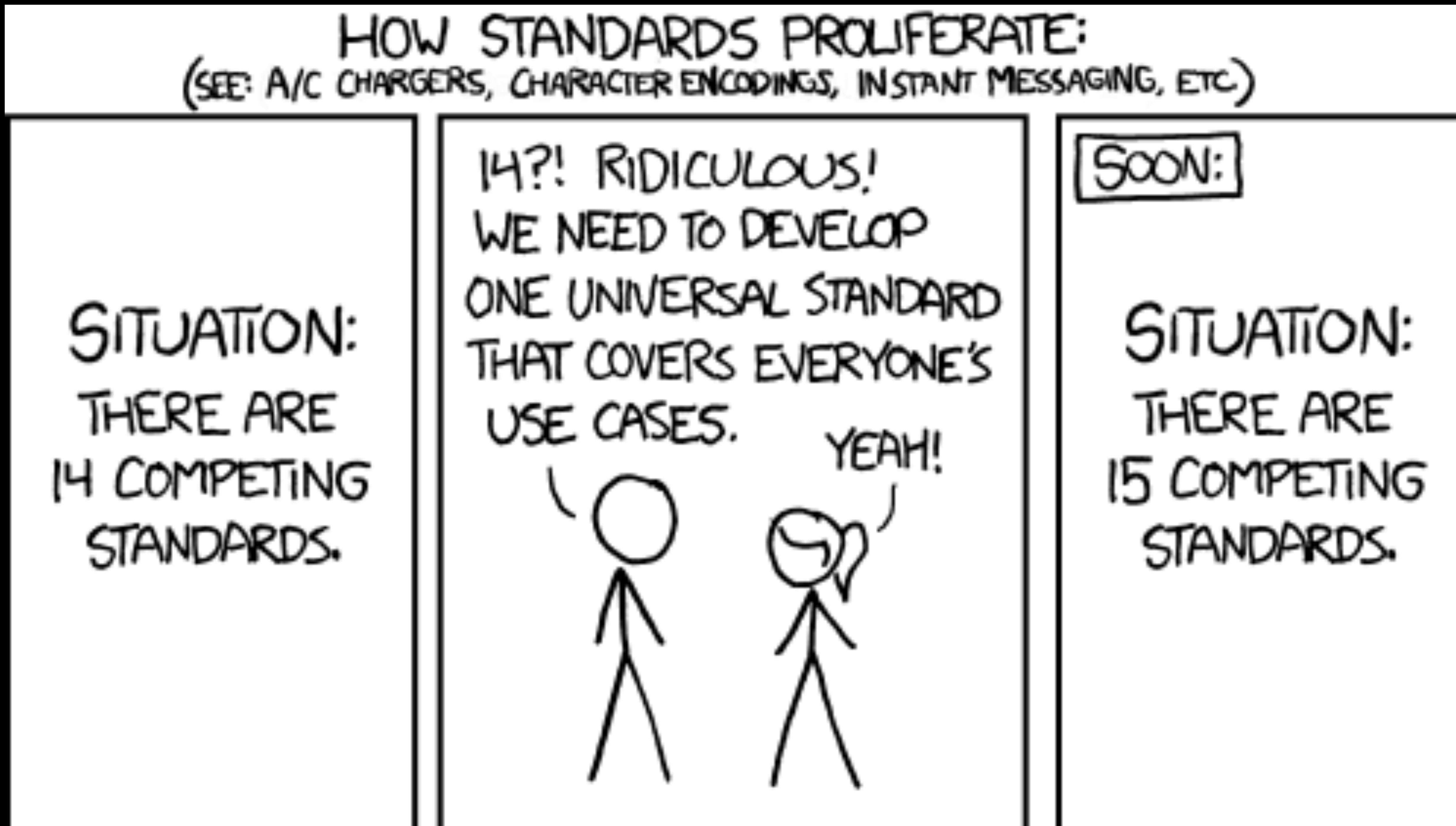
Use Accuracy

rness

ation Disparity

Accuracy Equality

Conditi


HOW STANDARDS PROLIFERATE:
(SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC.)
SITUATION: THERE ARE 14 COMPETING STANDARDS.
14?! RIDICULOUS! WE NEED TO DEVELOP ONE UNIVERSAL STANDARD THAT COVERS EVERYONE'S USE CASES.
YEAH!
SOON:
SITUATION: THERE ARE 15 COMPETING STANDARDS.

Fairness: Correlation

correlation(race, 

) = 0.8

mutual information(race, 

) = 0.6

Correlation does not measure causation

Atlidakis et al. FairTest: Discovering unwarranted associations in data-driven applications. EuroS&P 2017

What is fairness?

Sensitive inputs should not affect software behavior.

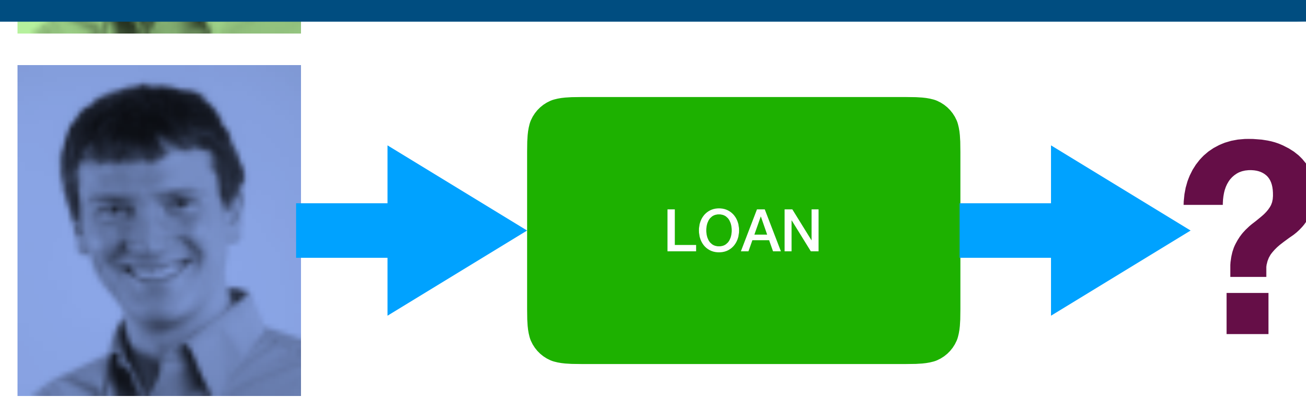
We want to measure causality!

Judea Pearl. Causal inference in statistics: An overview. Statistics Surveys 2009

causal testing

Sensitive inputs should not affect software behavior.

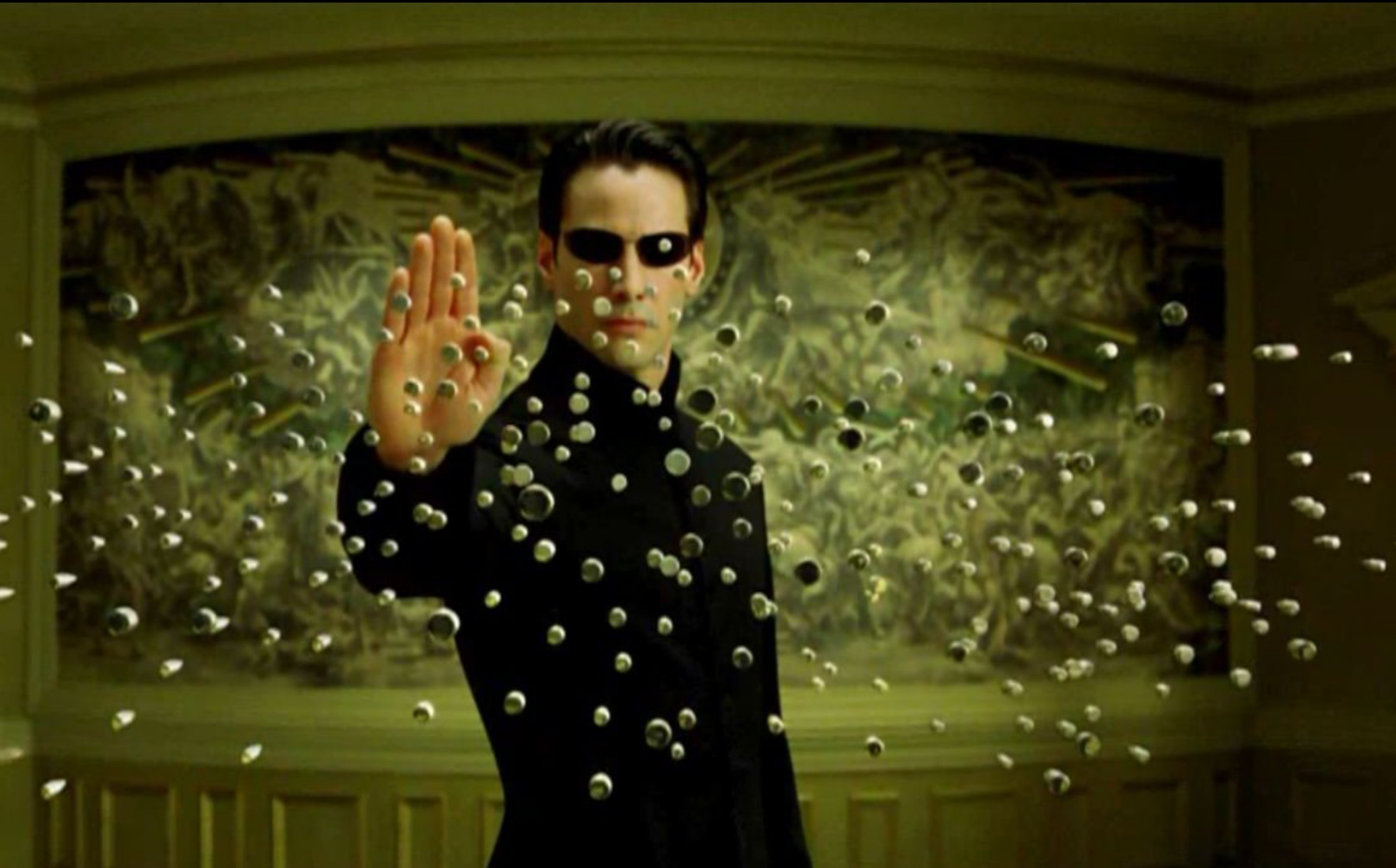
hypo
testing:



Galhotra, Brun, and Meliou, Fairness Testing: Testing Software for Discrimination. ESEC/FSE 2017

causal testing

causal testing



Themis

automated test-suite generator

How much does my software discriminate with respect to …?

Does my software discriminate more than 10% of the time, and against what?

Themis generates a test suite or can use a manually written one

http://fairness.cs.umass.edu

Angell, Johnson, Brun, and Meliou, Themis: Automatically Testing Software for Discrimination. ESEC/FSE 2018 Demo

discrimination measures

causal discrimination

LOAN() $\stackrel{?}{=}$ LOAN ()

group discrimination

How does Themis work?

adaptive, confidence-driven sampling

input schema

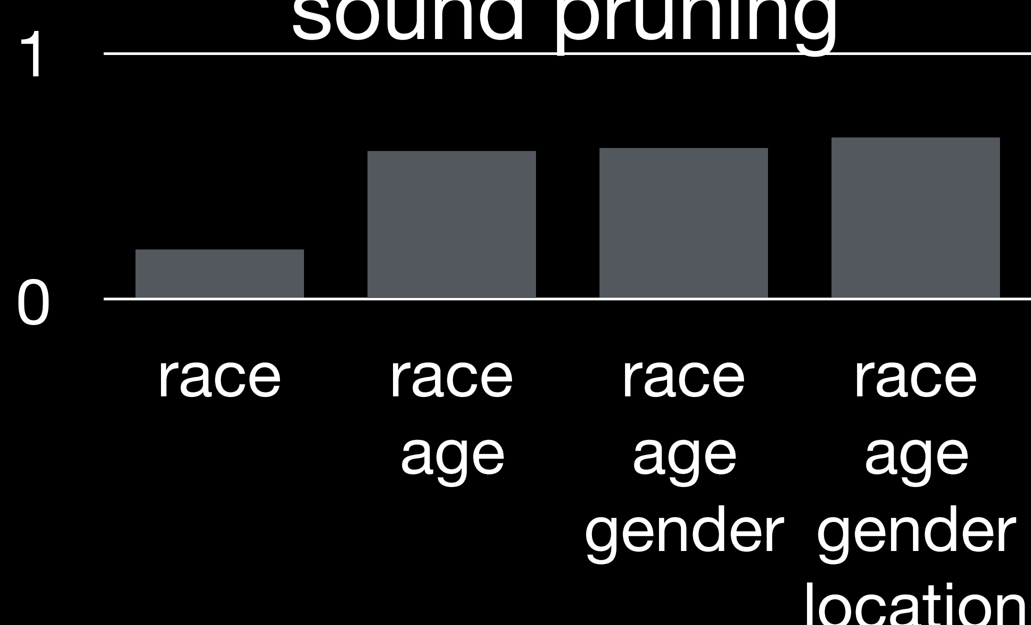
confidence

error bound

Themis

$$error = z^* \sqrt{\frac{p(1-p)}{r}}$$

sound pruning



Evaluation

Eight open-source decision systems trained on two public data sets

discrimination-aware logistic regression	[88]
discrimination-aware decision tree	[40]
discrimination-aware naive Bayes	[18]
discrimination-aware decision tree	[91]
naive Bayes	scikit-learn
decision tree	
logistic regression	
SVM	

- Census income dataset:
  financial data
  45K people
  income > $50K?
- Statlog German credit dataset:
  credit data
  1K people
  "good" or "bad" credit?

findings

Group discrimination is not enough.

More than 11% of the individuals had the output flipped just by altering the individual's gender.

Decision tree trained not to group discriminate against gender causal discriminated against gender: 0.11.

findings

Trying to avoid group discrimination may introduce other discrimination.

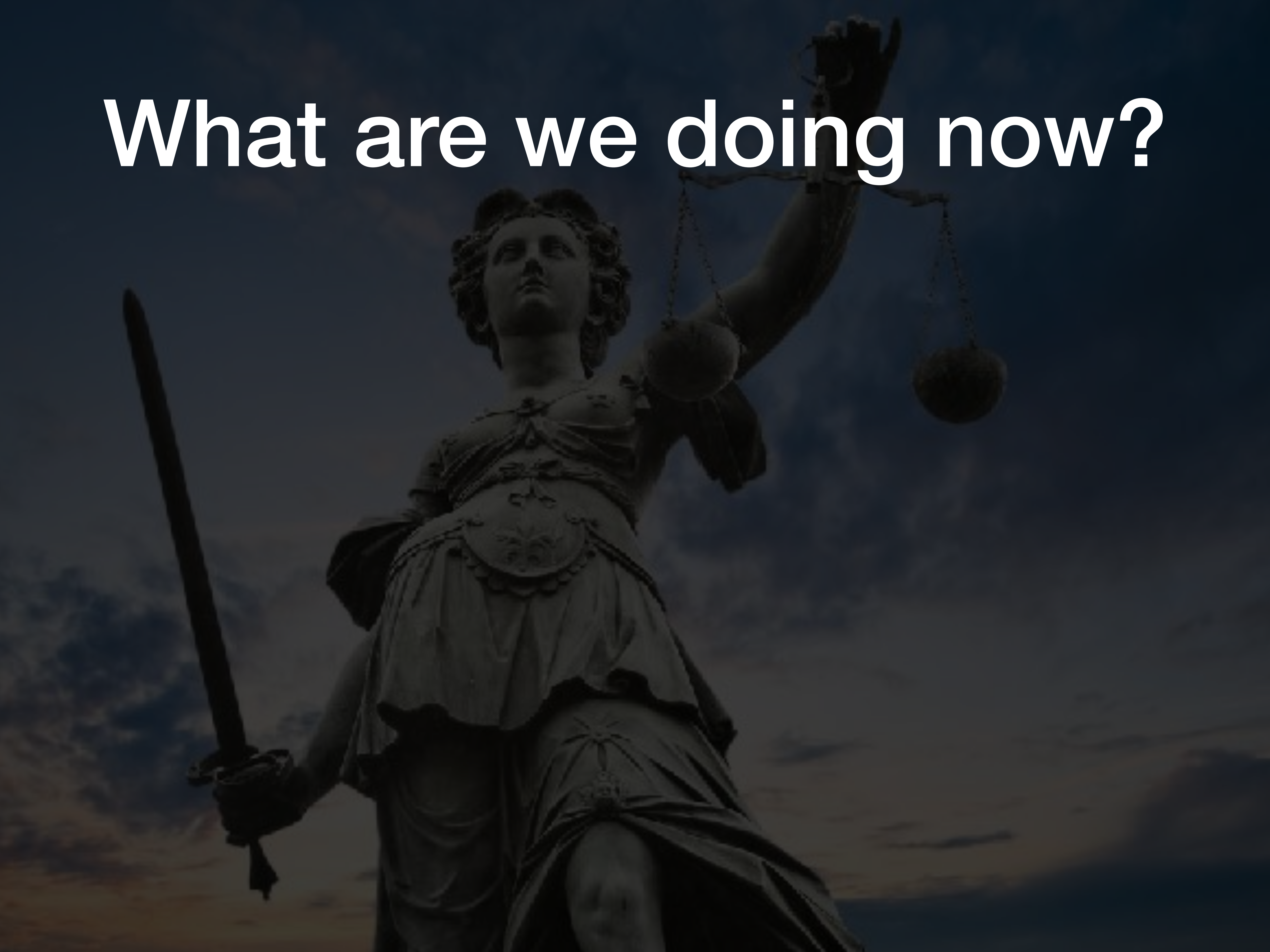
Training a decision tree not to discriminate against gender made it discriminate against race 38.4% of the time.

findings

Pruning is highly effective.

- The more a system discriminates, the more efficient Themis is.
- On average, pruning reduced test suites by 148× for causal and 2,849× for group discrimination. Best improvement was 13,000×.

What are we doing now?



ACLU GET UPDATES / DONATE

Amazon's Face Recognition Falsely Matched 28 Members of Congress With Mugshots

By Jacob Snow, Technology & Civil Liberties Attorney, ACLU of Northern California
JULY 26, 2018 | 8:00 AM

TAGS: Face Recognition Technology, Surveillance Technologies, Privacy & Technology

"The false matches were disproportionately of people of color, including six members of the Congressional Black Caucus, among them civil rights legend Rep. John Lewis (D-Ga.)."

nationwide, and today, there are 28 more causes for concern. In a test the ACLU recently conducted of the facial recognition tool, called "Rekognition," the software incorrectly matched 28 members of Congress, identifying them as other people who have been arrested for a crime.

The members of Congress who were falsely matched with the mugshot

https://www.aclu.org/blog/privacy-technology/surveillance-technologies/amazons-face-recognition-falsely-matched-28

What are we doing now?


ACLU
GET UPDATES / DONATE

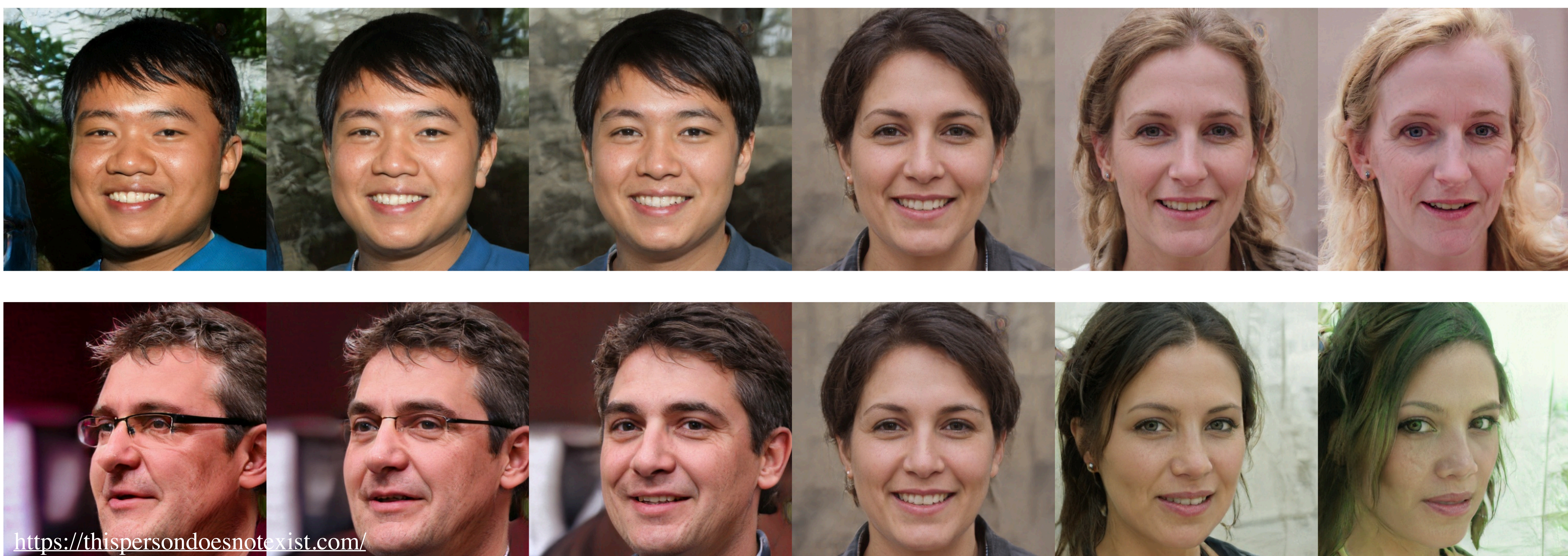
Amazon's Face Recognition Falsely Matched 28 Members of Congress With Mugshots

By Jacob Snow, Technology & Civil Liberties Attorney, ACLU of Northern California
JULY 26, 2018 | 8:00 AM

TAGS: Face Recognition Technology, Surveillance Technologies, Privacy & Technology

Amazon's face surveillance technology is the target of growing opposition nationwide, and today, there are 28 more causes for concern. In a test the ACLU recently conducted of the facial recognition tool, called "Rekognition,"

Fair computer vision



https://thispersondoesnotexist.com/

What are we doing now?

ACLU GET UPDATES / DONATE

Amazon's Face Recognition Falsely Matched 28 Members of Congress With Mugshots

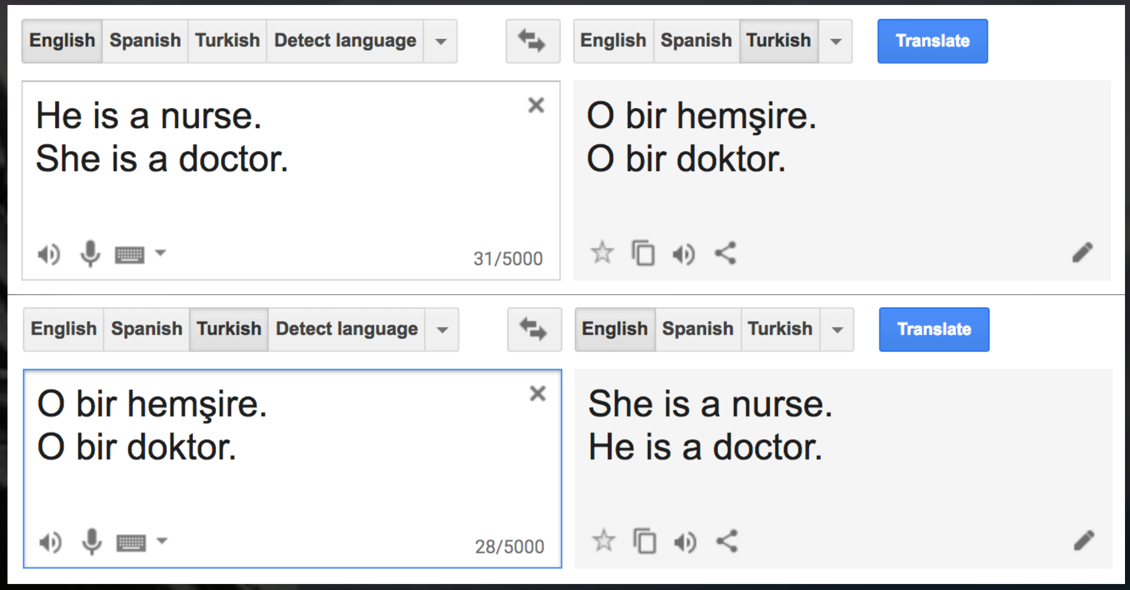
By Jacob Snow, Technology & Civil Liberties Attorney, ACLU of Northern California
JULY 26, 2018 | 8:00 AM

TAGS: Face Recognition Technology, Surveillance Technologies, Privacy & Technology

Amazon's face surveillance technology is the target of growing opposition nationwide, and today, there are 28 more causes for concern. In a test the ACLU recently conducted of the facial recognition tool, called "Rekognition," the software incorrectly matched 28 members of Congress, identifying them as other people who have been arrested for a crime.

The members of Congress who were falsely matched with the mugshot

Fair computer vision



Fair natural language processing

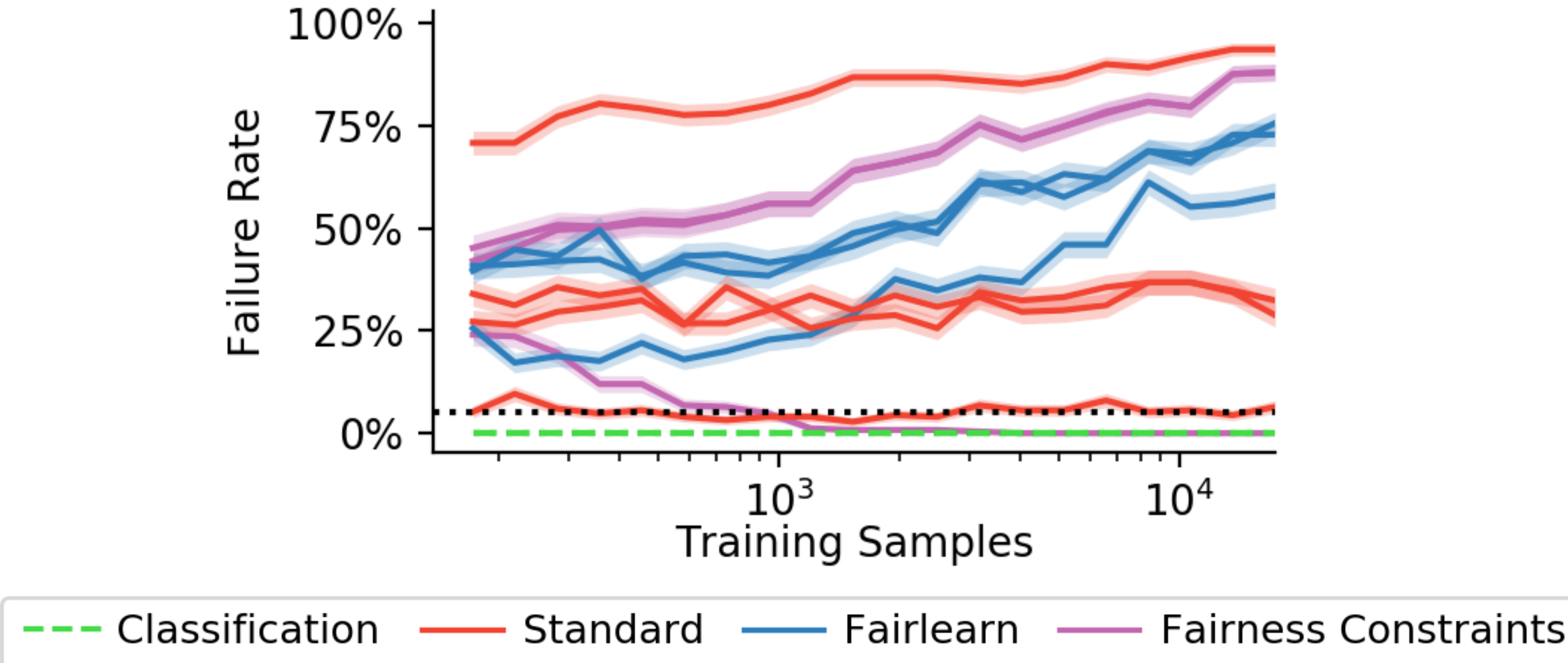
applied

But what's the holy grail?

Provably fair machine learning:

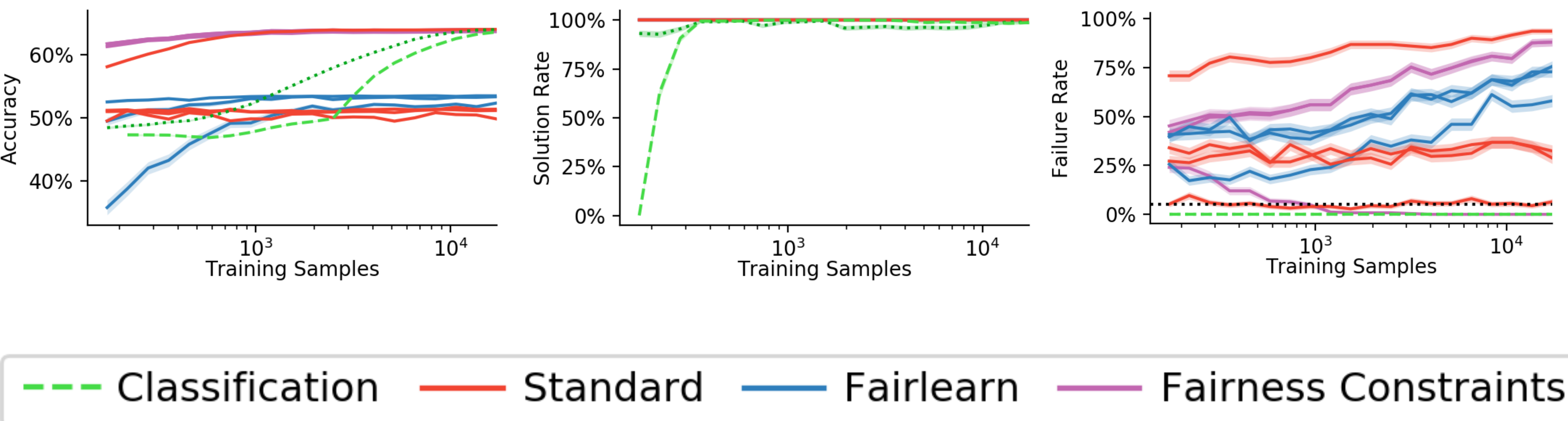
Provide (high-probability) guarantees that the classifier is fair on unseen data.

Disparate Impact



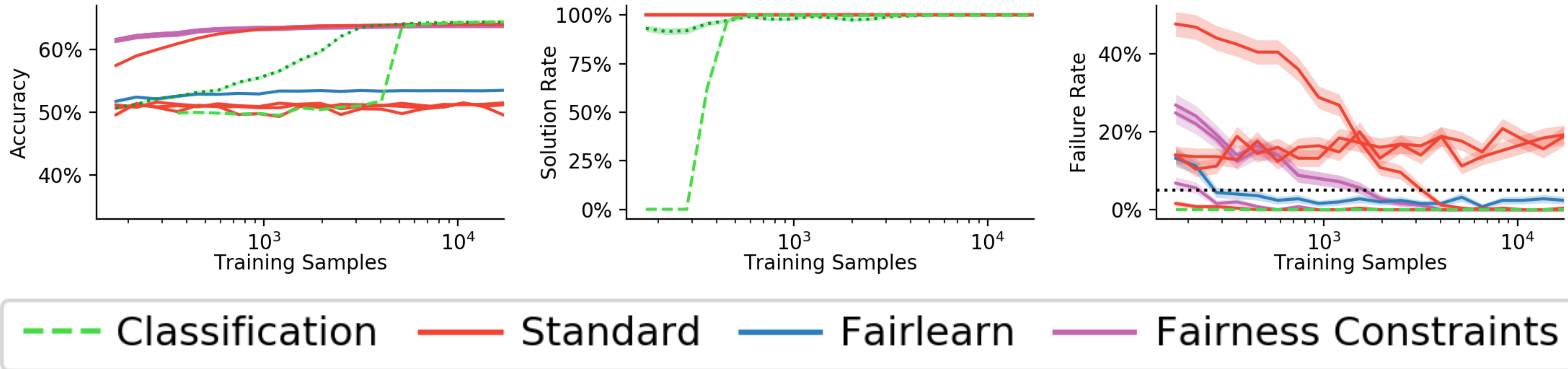
Fairlearn: Agarwal et al. A reductions approach to fair classification. ICML 2018.
Fairness Constraints: Zafar et al., Fairness Constraints: A Mechanism for Fair Classification. FATML 2015.

Disparate Impact

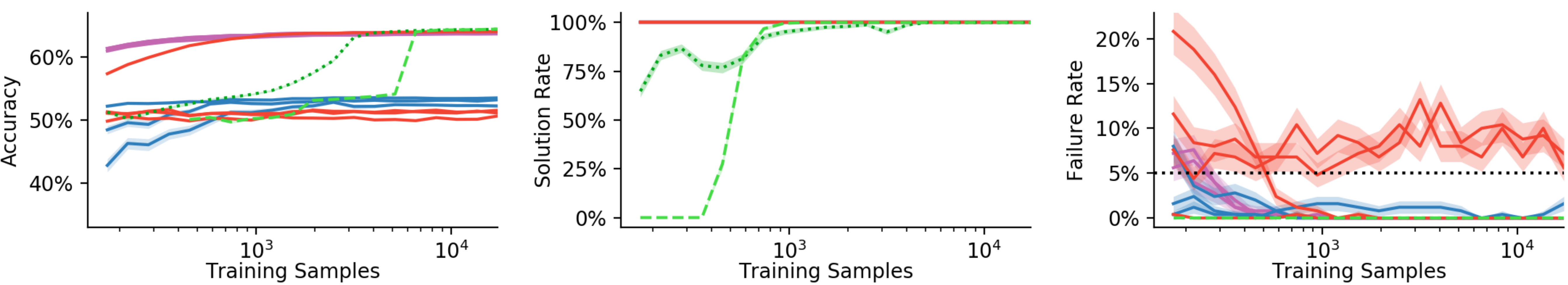


Fairlearn: Agarwal et al. A reductions approach to fair classification. ICML 2018.
Fairness Constraints: Zafar et al., Fairness Constraints: A Mechanism for Fair Classification. FATML 2015.

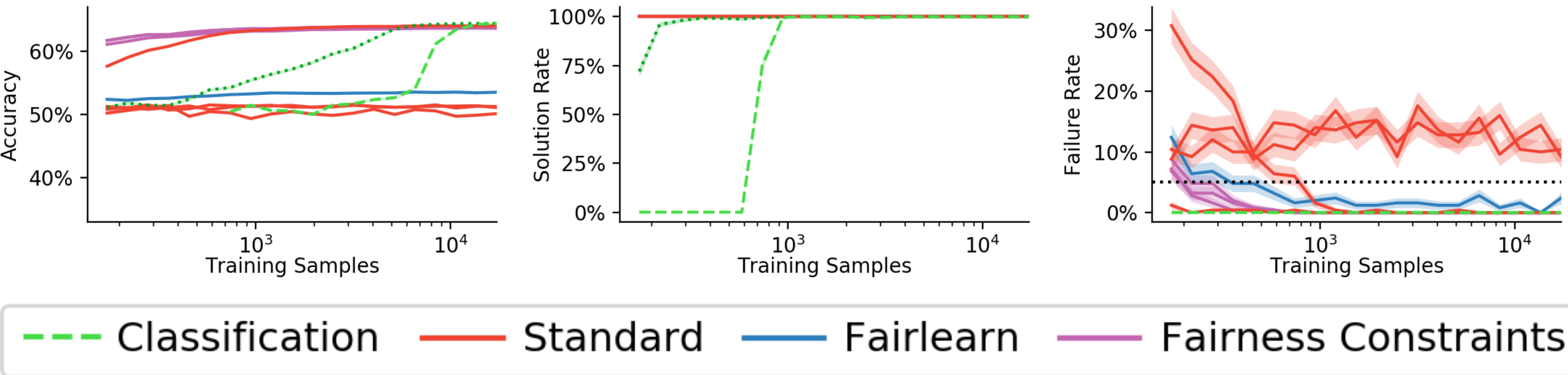
Demographic Parity



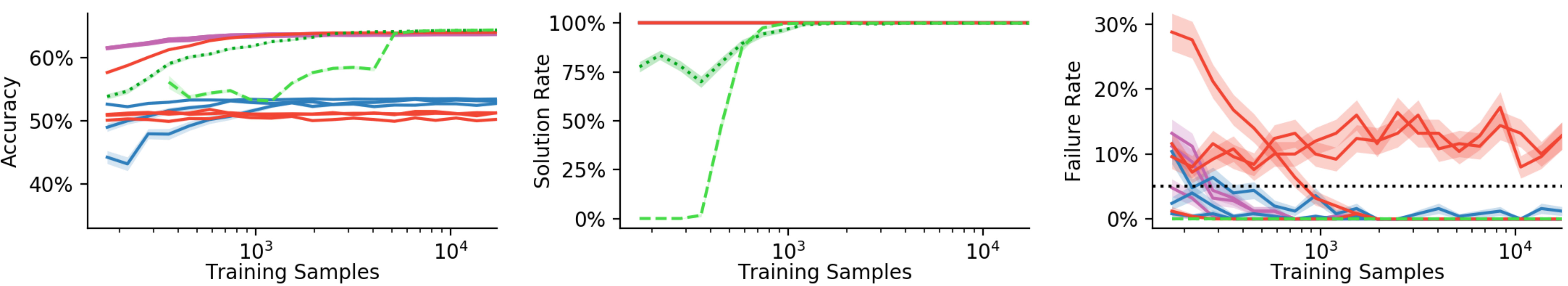
Equal Opportunity



Equalized Odds



Predictive Equality

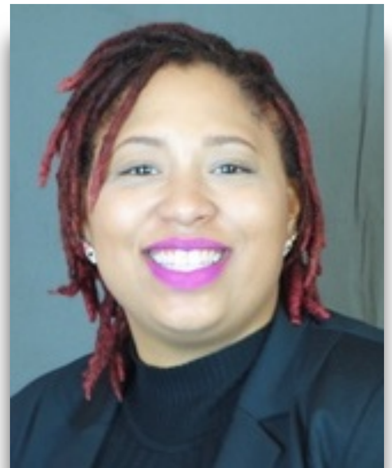


Contributions

http://fairness.cs.umass.edu

- Causality-based definition and method for measuring software fairness
- Themis, an automated test-suite generator for fairness testing
- Evaluation on real-world software, demonstrating software is biased and our methods can catch it
- Provable guarantees on fairness in machine learning


Rico Angell


Brittany Johnson


Stephen Giguere


Sarah Brockman


Blossom Metevier


Sainyam Galhotra


Alexandra Meliou


Andy Barto


Bruno Castro da Silva


Emma Brunskill


Philip Thomas


Yuriy Brun

http://fairness.cs.umass.edu

https://tinyurl.com/FairnessPaper

NSF

UMass Amherst

ORACLE

Contributions

http://fairness.cs.umass.edu

- Causality-based definition and method for measuring software fairness
- Themis, an automated test-suite generator for fairness testing
- Evaluation on real-world software, demonstrating software is biased and our methods can catch it
- Provable guarantees on fairness in machine learning

Homework 1

- Due September 17, 9AM
- Will be posted shortly (you'll get an email)
- Learn some machine learning! Learn to use tools that help evaluate and mitigate bias in machine learning.
- Requires downloading a 5GB file, so do that early.