Alexandra Meliou

University of Massachusetts Amherst


Modern software
influences critical decisions
Forbes
YOUR READING LIST
Trusting Robots
host.madison.com
MENU
THE CAP TIMES
Madison, Wisconsin
News
Opinion
Leisure
Photos
Podcasts
Reviews
Philanthropy
EDITOR'S PICK
Wisconsin Supreme Court allows
KATELYN FERRAL
PARENT.
LEARN ABOUT ADOPTION
#1 DAD
AdoptUSKids
STOP THE TEXTS.
host.madison.com/content/tncms/live/#1
is one with me," Stifelman says as his mechanized appendages
pull tight another knot.

businessinsider.com

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

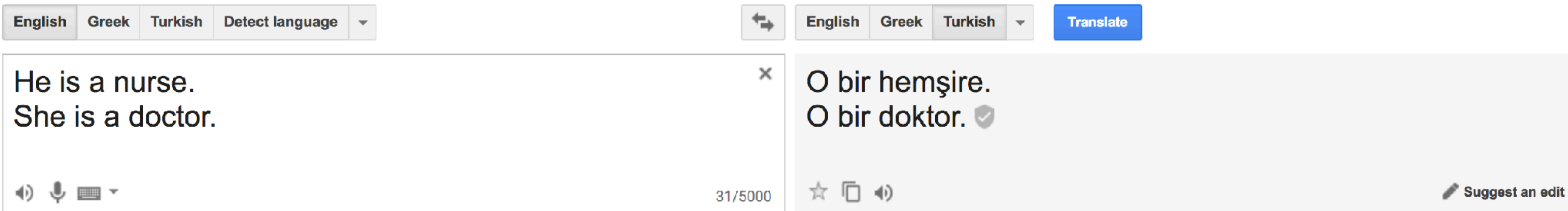
Software can make bad decisions. Software can discriminate!

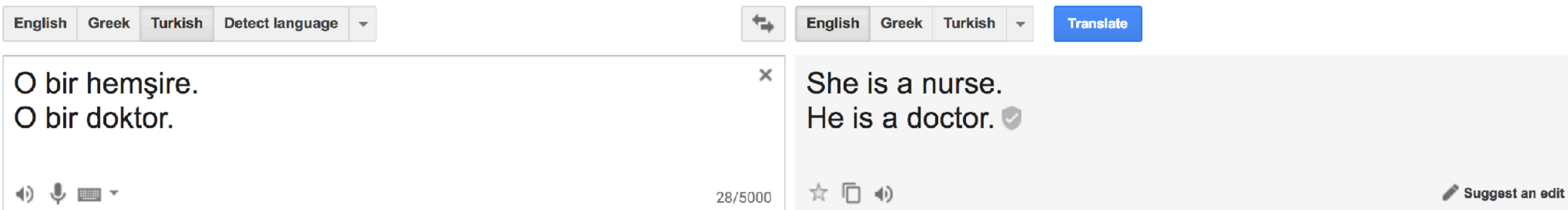
Just as the 18-year-old girls were realizing they were too big for the tiny conveyances — which belonged to a 6-year-old boy — a woman came running after them saying, "That's my kid's stuff." Borden and her friend immediately dropped the bike and scooter and walked away.

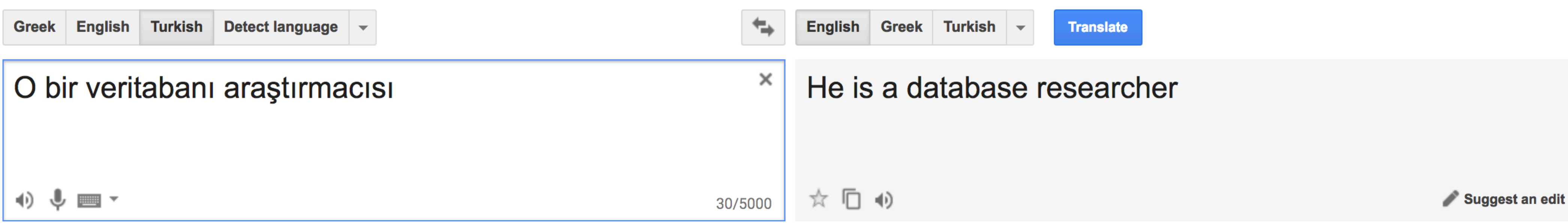
But it was too late — a neighbor who witnessed the heist had already called the police. Borden and her friend were arrested and charged with burglary and petty theft for the items, which were valued at a total of $80.

area, as well as the ability of our various carrier partners to deliver up to 9:00 pm every single day, even Sunday .

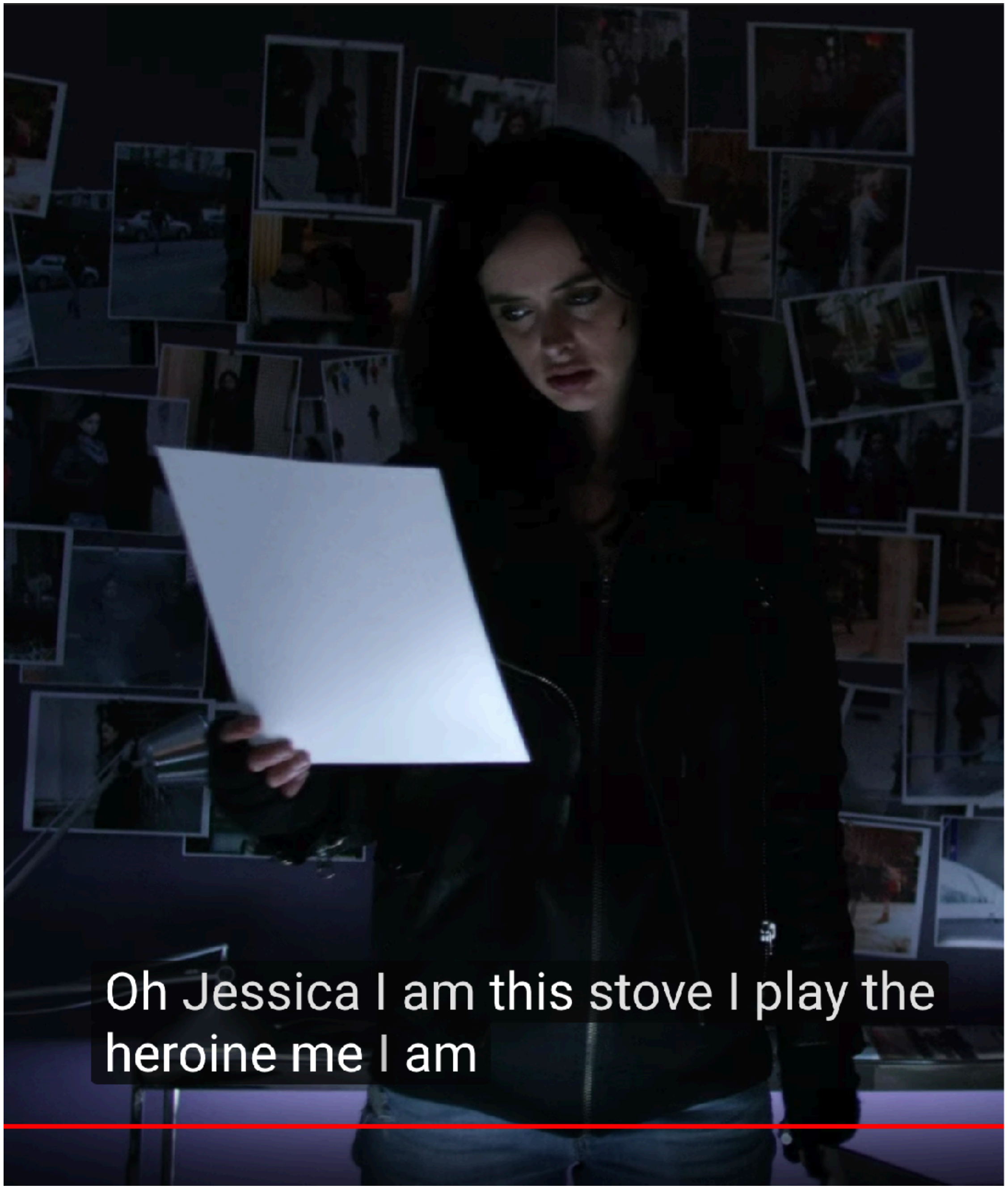
algorithms can exacerbate societal biases





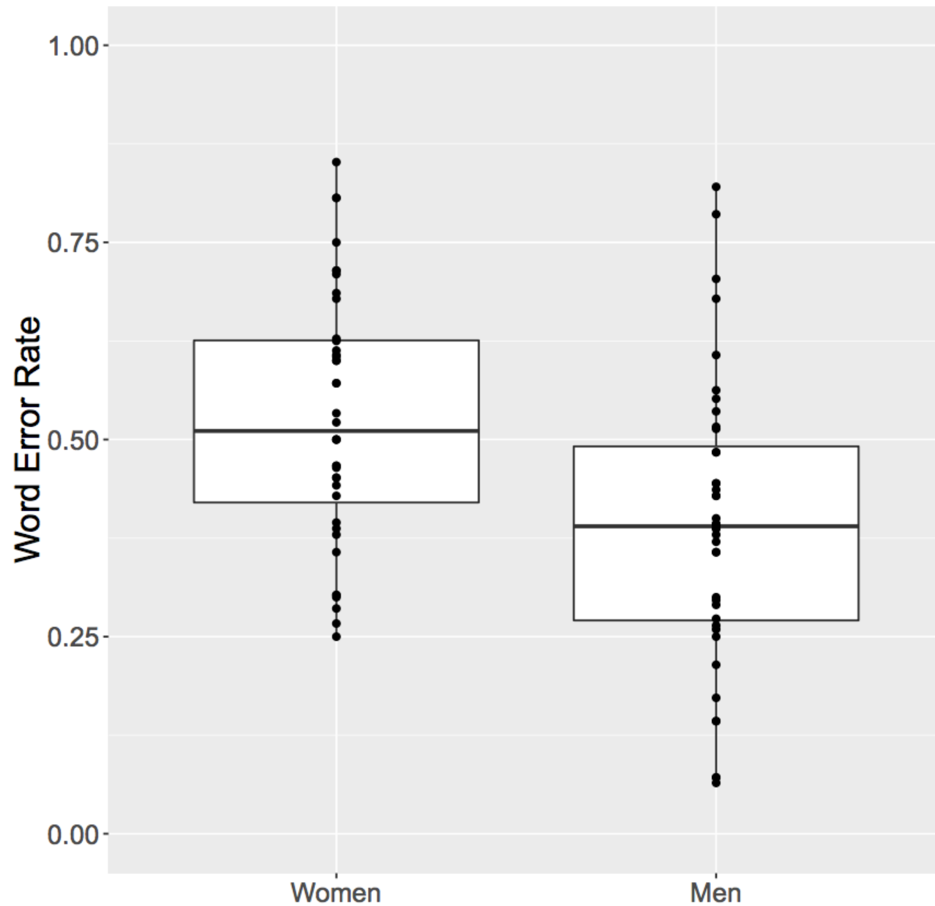


algorithms don't provide the same service to all



automatic captions

algorithms don't provide the same service to all

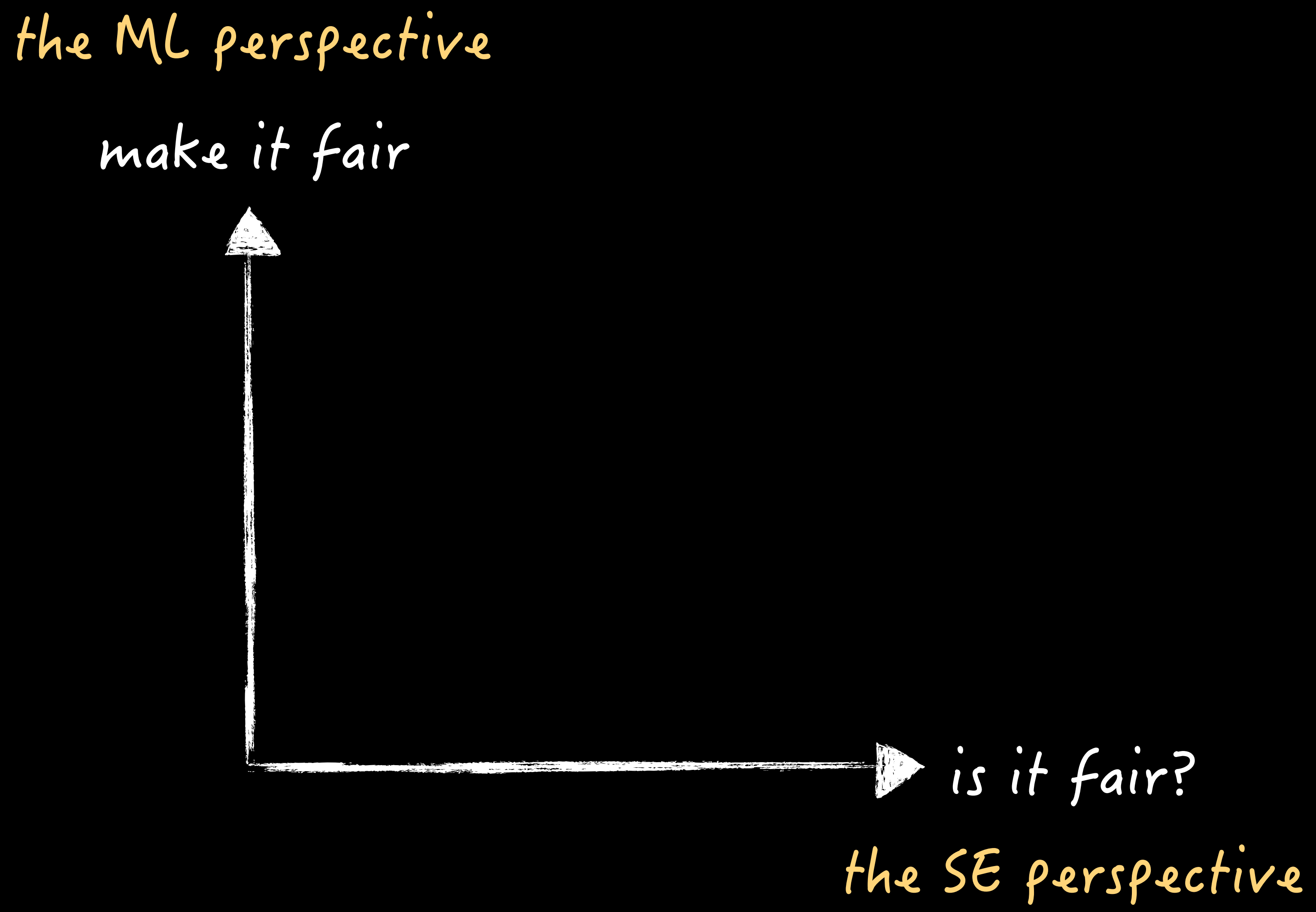


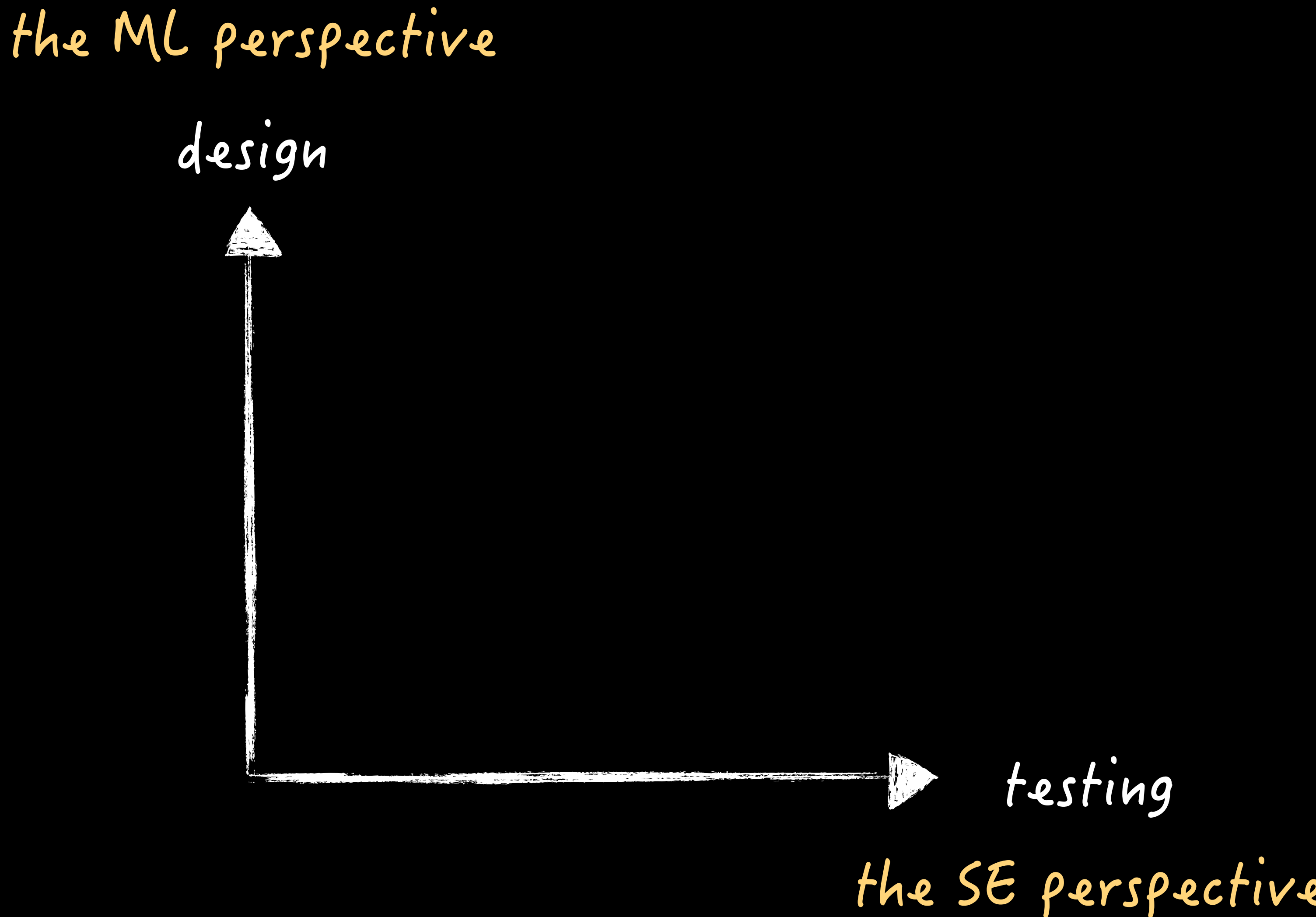
Rachael Tatman, "Gender and Dialect Bias in YouTube's Automatic Captions" in 2017 Workshop on Ethics in Natural Language Processing

algorithms don’t provide the same service to all

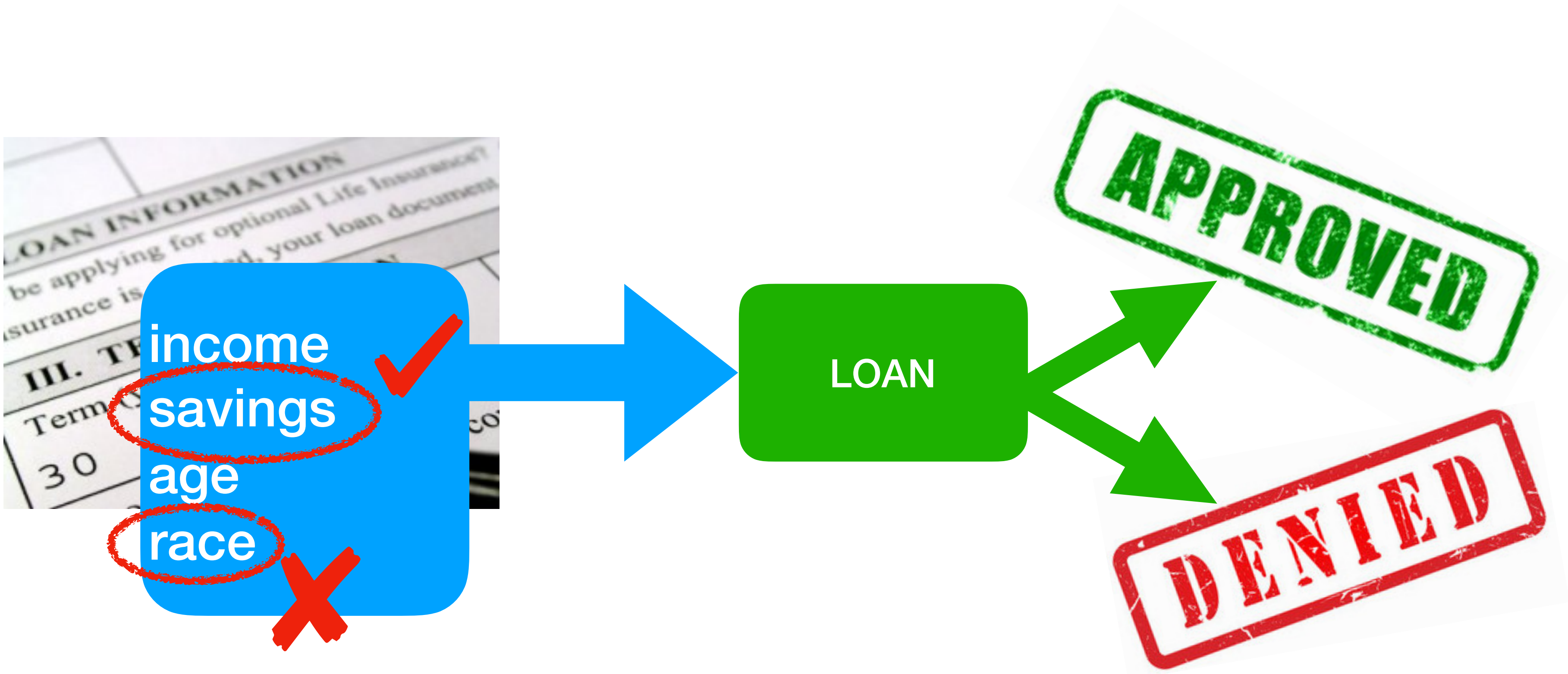
Joy Buolamwini
https://www.ted.com/talks/joy_buolamwini_how_i_m_fighting_bias_in_algorithms

data


the ML perspective
make it fair
is it fair?
the SE perspective


the ML perspective
design
testing
the SE perspective

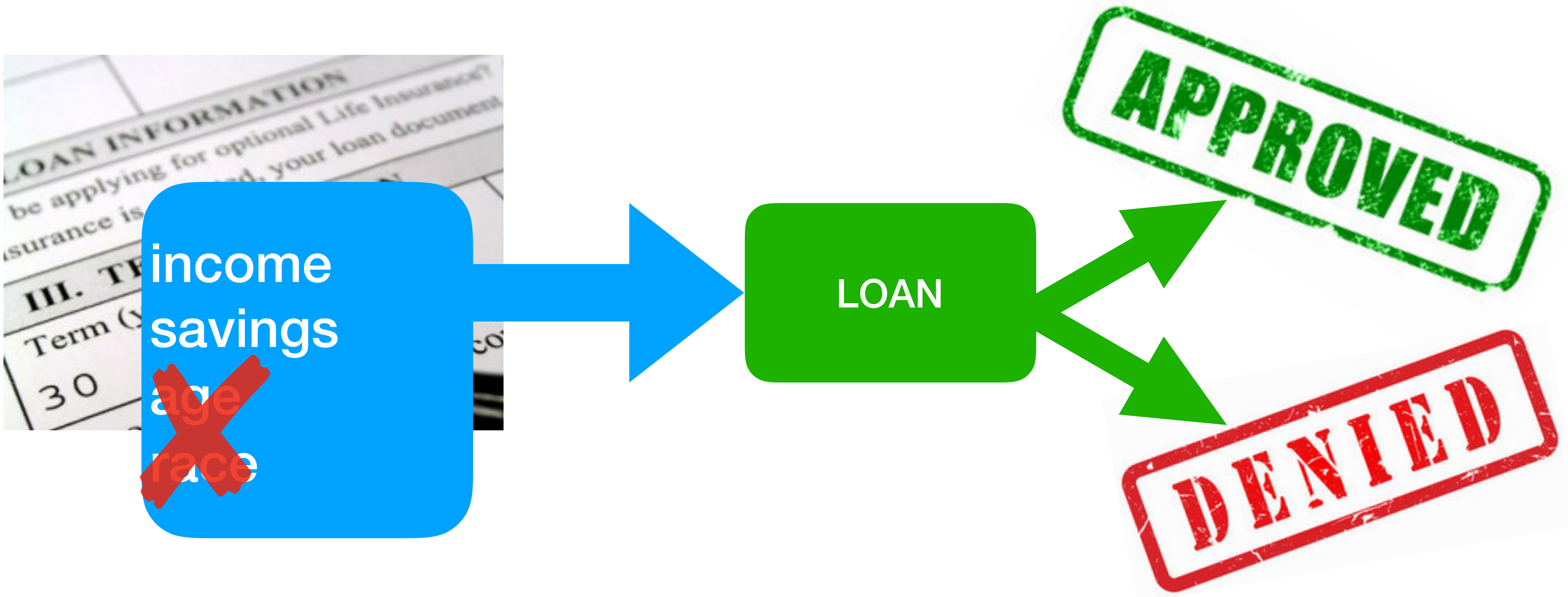
LOAN program



this is not about policy

approaches to fairness

1. Hide the data



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

approaches to fairness

2. Compare subpopulation proportions

1. Ineffective if race or age correlate with savings or income
2. Fails to identify discrimination against individuals

[Calders and Verwer. Three naive Bayes approaches for discrimination-free classification. Data Mining and Knowledge Discovery, 2010.]

how it can be unfair to individuals

country A country B

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

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

Country A and country B discriminations cancel each other out, and the group discrimination measure can be 0.

approaches to fairness

3. Measure differences for individuals

Sensitive inputs should not affect software behavior.

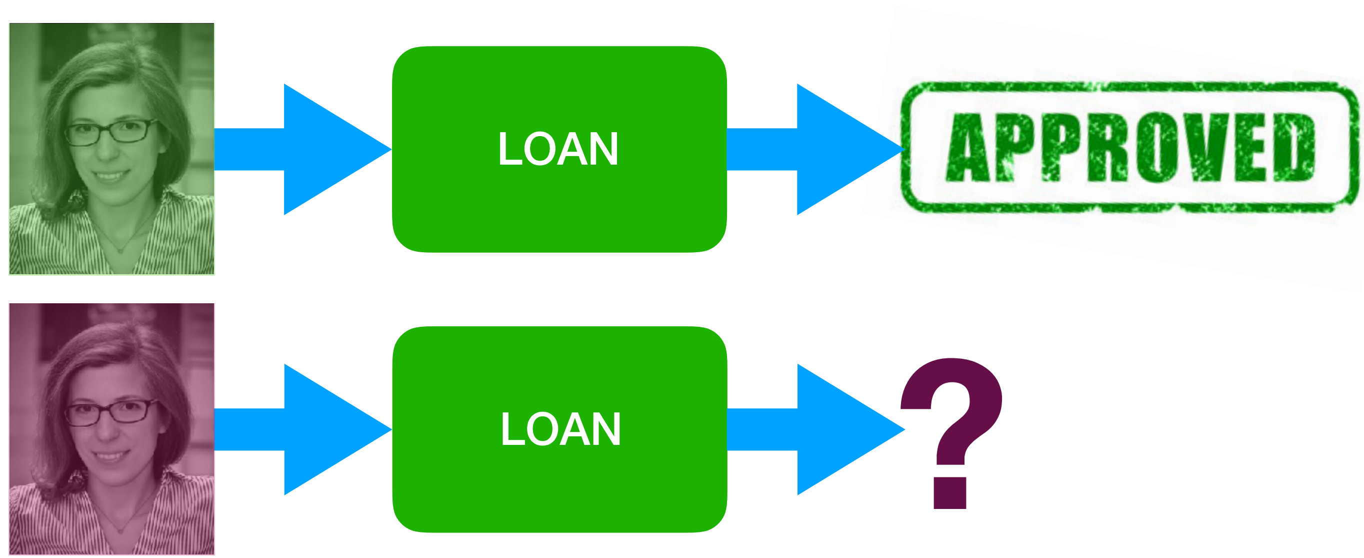
We want to measure causality!

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

causal testing

Sensitive inputs should not affect software behavior.

hypothesis testing:



- Why different definitions?
    - Systems designed to be fair under one definition may be unfair under another

- Why testing?
    - Systems designed to not discriminate may still have discrimination bugs

Fairness Testing: Testing Software for Discrimination
Sainyam Galhotra, Yuriy Brun, Alexandra Meliou

ESEC/FSE 2017
ACM Distinguished Paper

Themis

automated test-suite generator

How much does my software discriminate with respect to ...?

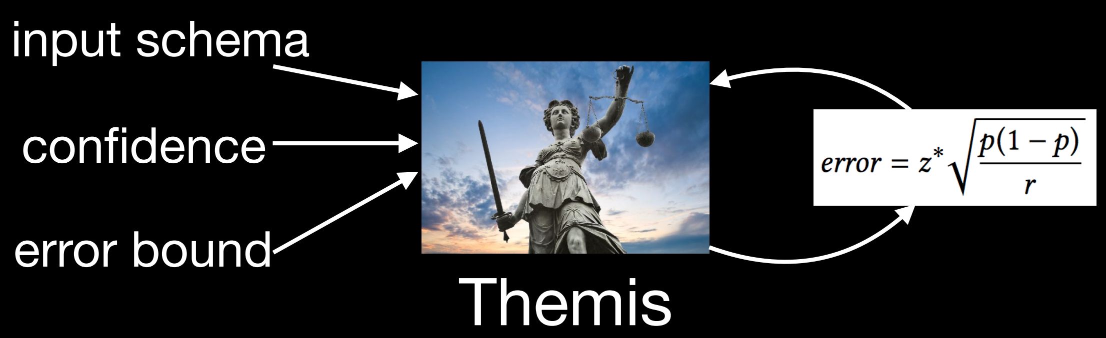
Does my software discriminate more than 10% of the time, and against what?

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

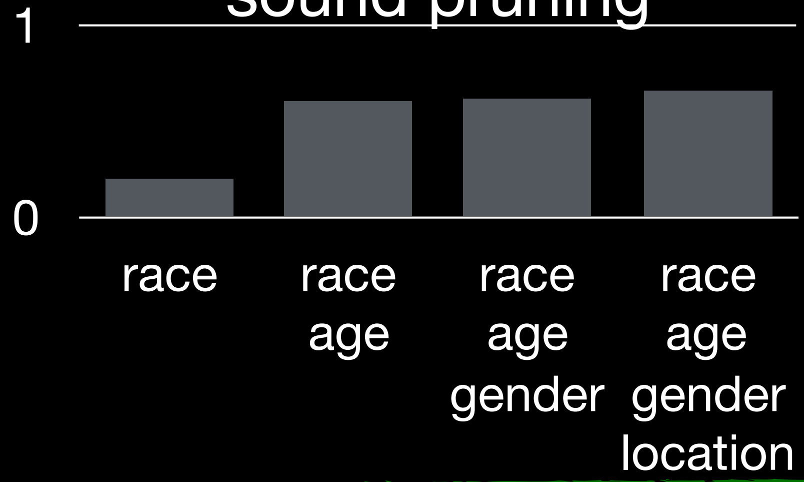
http://fairness.cs.umass.edu

How does Themis work?

adaptive, confidence-driven sampling



sound pruning



findings

Group discrimination is not enough.

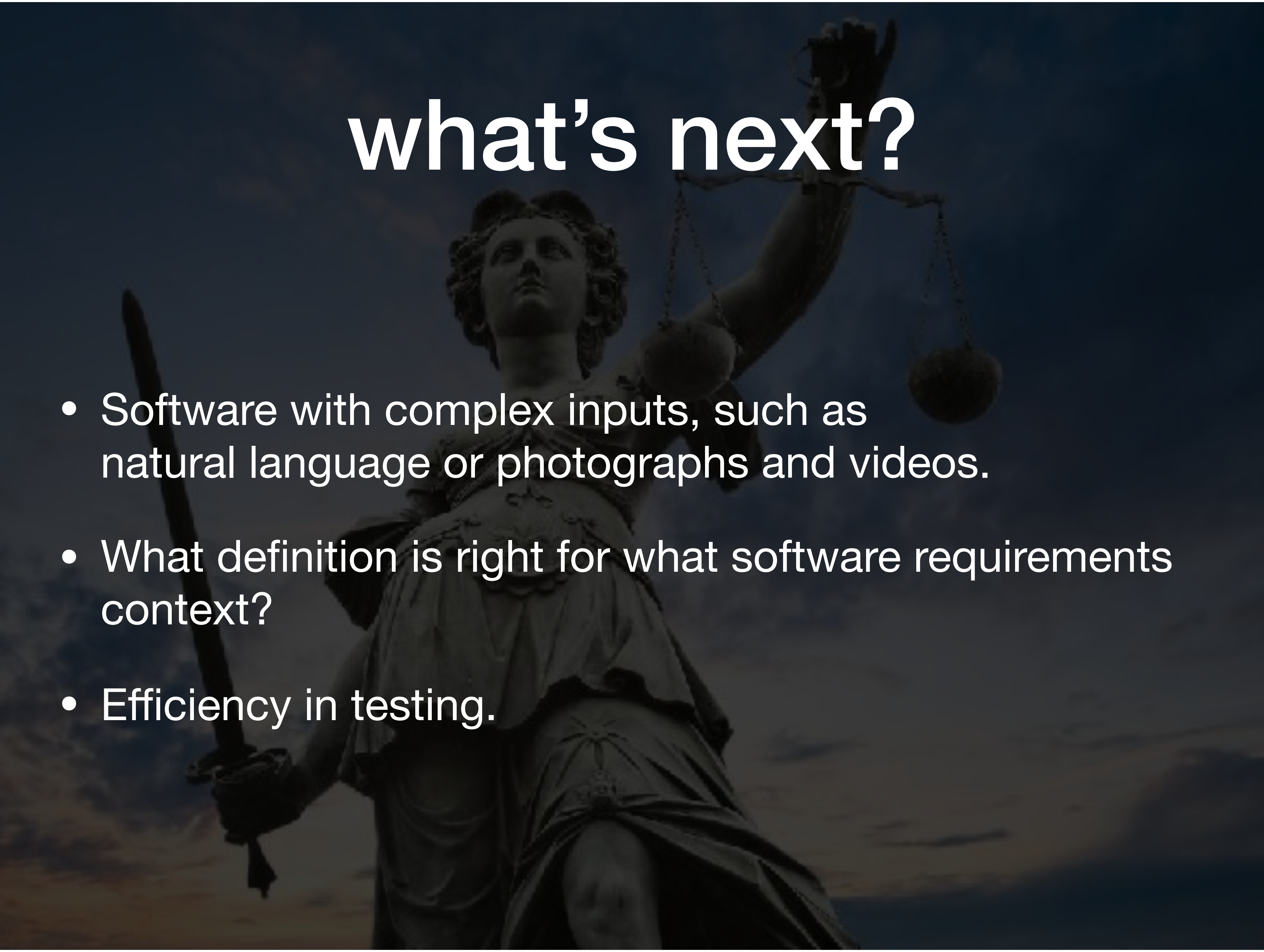
More than 11% of the individuals had the output flipped just by altering the individual's gender.

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.

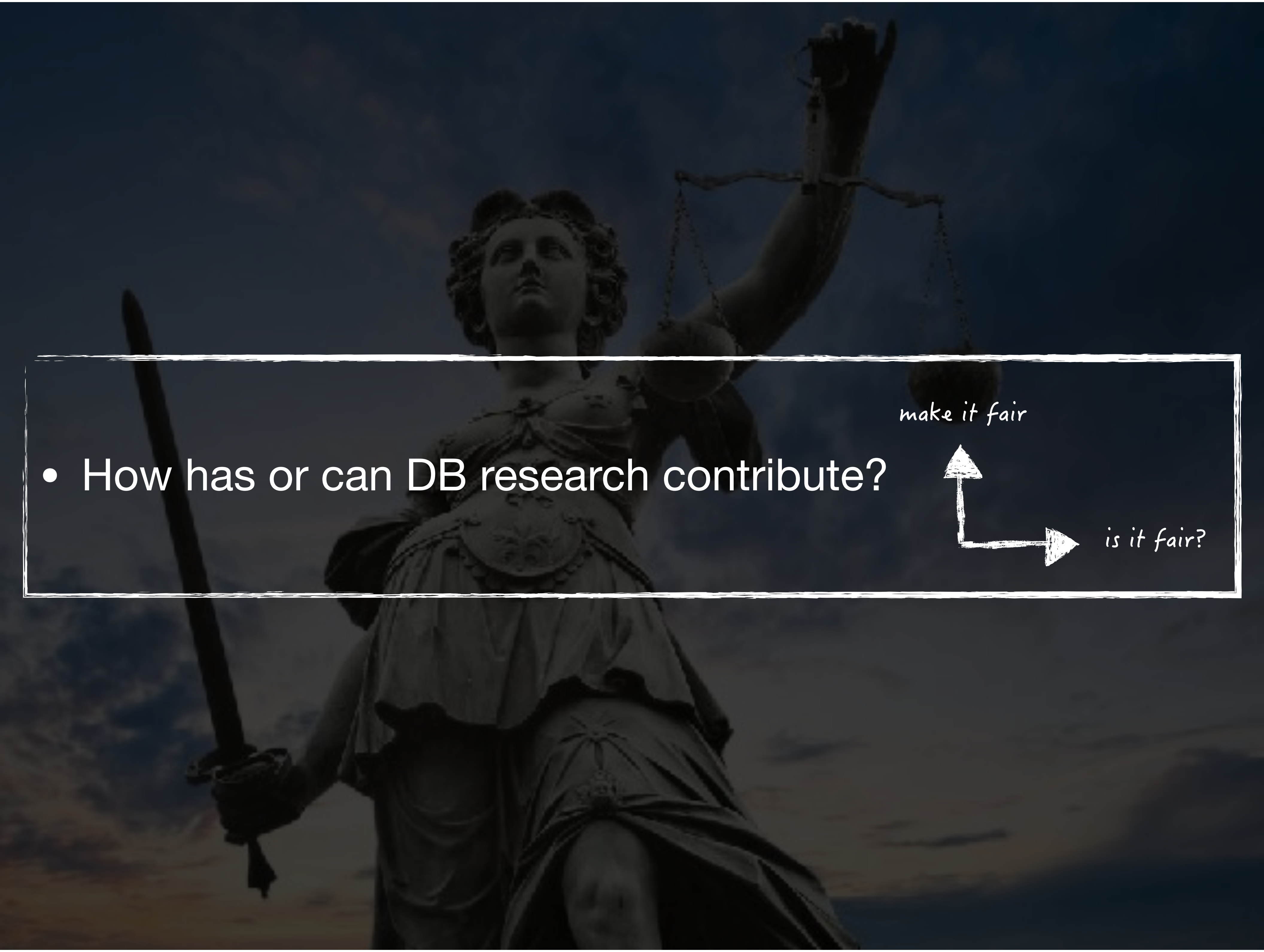


what’s next?

- Software with complex inputs, such as natural language or photographs and videos.
- What definition is right for what software requirements context?
- Efficiency in testing.

what's next?

- Software with complex inputs, such as natural language or photographs and videos.
    - How to do causal testing?
- What definition is right for what software requirements context?
    - Infrastructure that adjusts to new definitions
- Efficiency in testing.
    - Comparative behavior explodes search space



- How has or can DB research contribute?

make it fair

is it fair?

data

"What if data is biased?"

"What if my view of the data is skewed?"

"What if the data is sensitive?"

"What if data is missing?"

"What if data is dirty?"

"What if I don't understand the data or results?"

"What if data is biased?" diversification

skew

visualization

"What if my view of the data is skewed?"

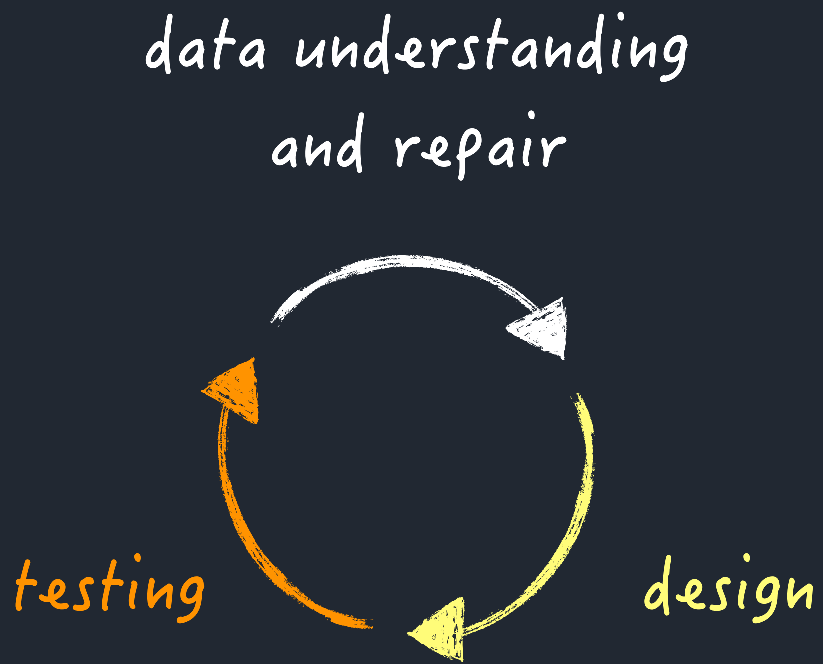
"What if the data is sensitive?" privacy

"What if data is missing?" completeness

"What if data is dirty?" quality repair

"What if I don't understand the data or results?"

explanations provenance


data understanding
and repair
design
testing