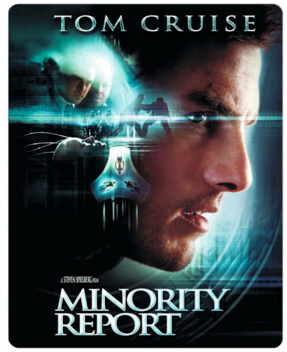
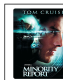
Software Fairness

April 11, 2023




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'

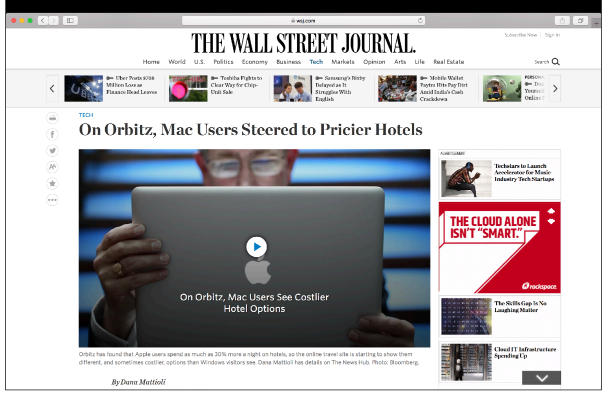
● Join our live Q&A with Homicide Watch this Friday

▲ 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

Modern software influences critical decisions

https://www.aclu.org/blog/national-security/discriminatory-profiling/government-blacklisting-people-based-predictions



HAAS NEWS > NEWS CATEGORIES > RESEARCH NEWS

Minority homebuyers face widespread statistical lending discrimination, study finds

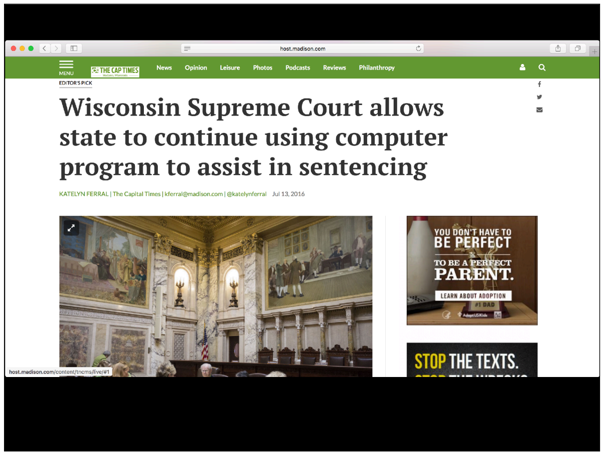
By Laura Counts | NOVEMBER 13, 2018

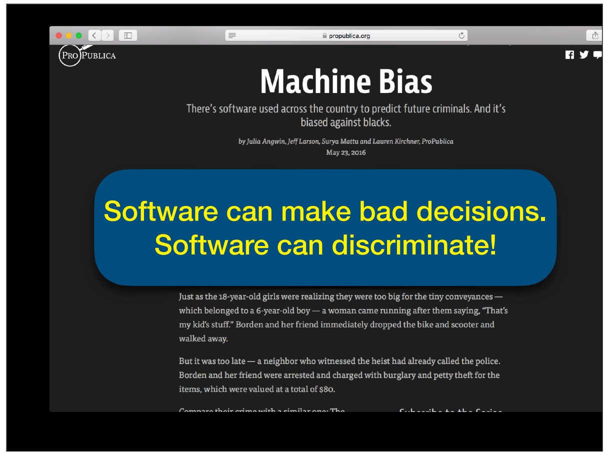
Face-to-face meetings between mortgage officers and homebuyers have been rapidly replaced by online applications and algorithms, but lending discrimination hasn't gone away.

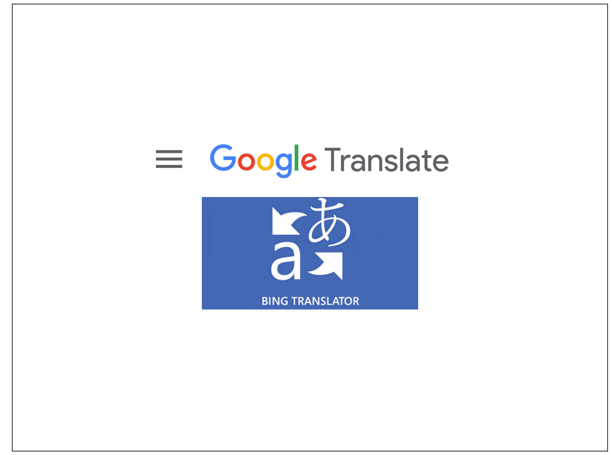
A new University of California, Berkeley study has found that both online and face-to-face lenders charge higher interest rates to African American and Latino borrowers, earning 11 to 17 percent higher profits on such loans. All told, those homebuyers pay up to half a billion dollars more in interest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

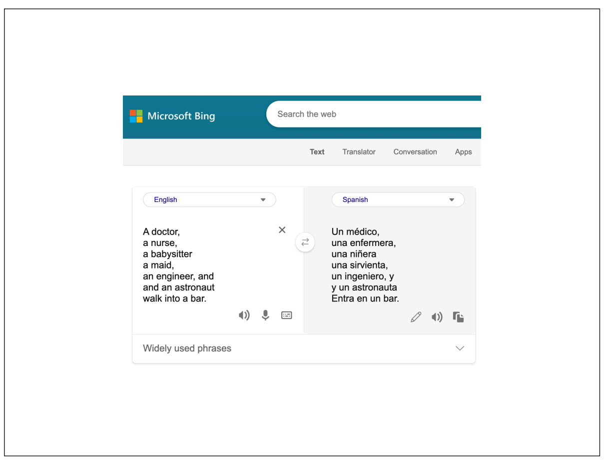
The findings raise legal questions about the rise of statistical discrimination in the fintech era, and point to potentially widespread violations of U.S. fair lending laws, the researchers say. While lending discrimination has historically been caused by human prejudice, pricing disparities are increasingly 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

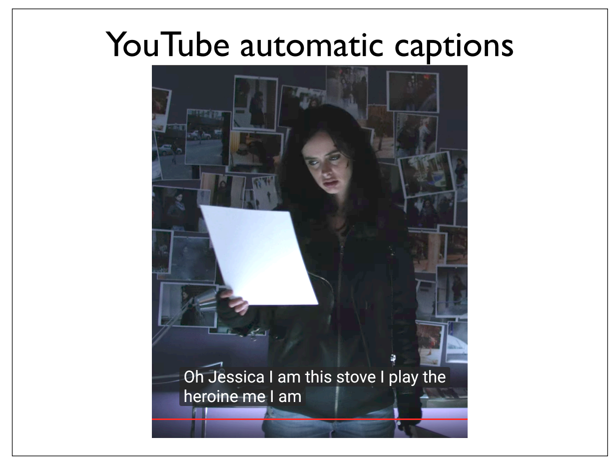

News Opinion Leisure Photos Podcasts Reviews Philanthropy
EDITOR'S PICK
Wisconsin Supreme Court allows state to continue using computer program to assist in sentencing
KATELYN FERRAL | The Capital Times | kferral@madison.com | @katelynferral Jul 13, 2016
YOU DON'T HAVE TO BE PERFECT TO BE A PERFECT PARENT.
LEARN ABOUT ADOPTION
STOP THE TEXTS.

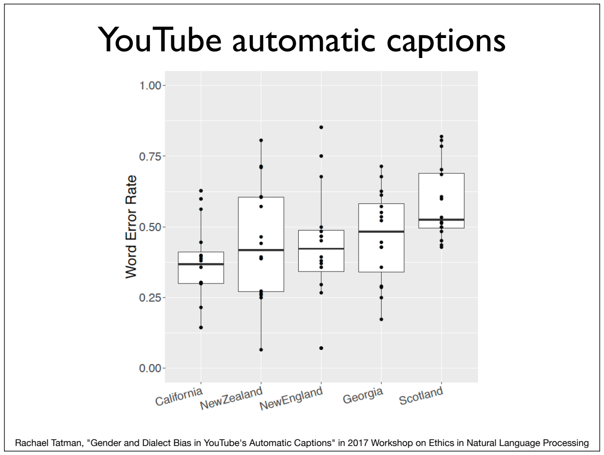

ProPublica
Machine Bias
There's software used across the country to predict future criminals. And it's biased against blacks.
by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica
May 23, 2016
Software can make bad decisions.
Software can discriminate!
Just as the 18-year-old girls 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.

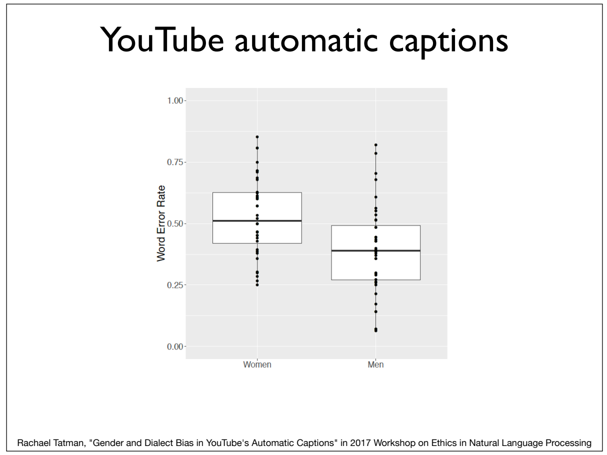

Google Translate
BING TRANSLATOR

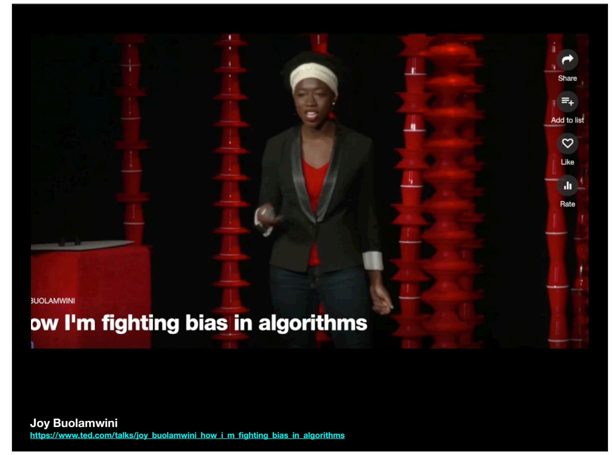

Microsoft Bing
Search the web
Text Translator Conversation Apps
English
Spanish
A doctor,
a nurse,
a babysitter
a maid,
an engineer, and
and an astronaut
walk into a bar.
Un médico,
una enfermera,
una niñera
una sirvienta,
un ingeniero, y
y un astronauta
Entra en un bar.
Widely used phrases

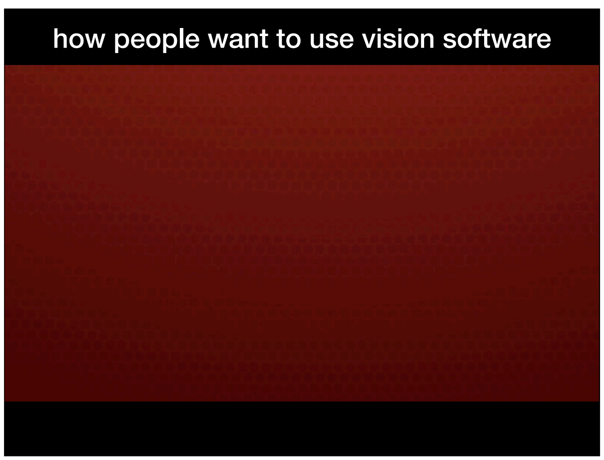

YouTube

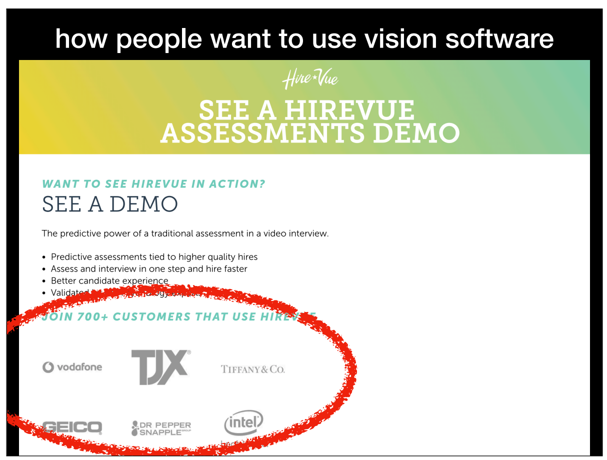

YouTube automatic captions
Oh Jessica I am this stove I play the
heroine me I am









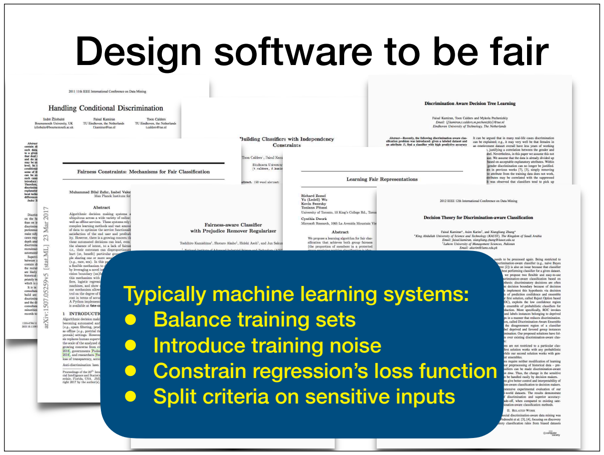


today's goals

Define software discrimination.

Operationalize measuring discrimination through causal software testing.

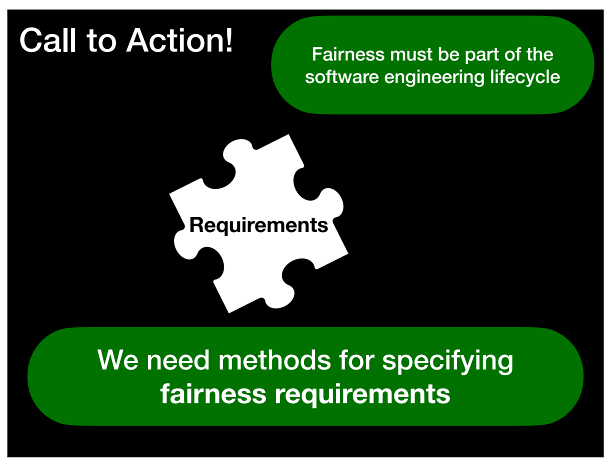
Provide provable fairness guarantees.

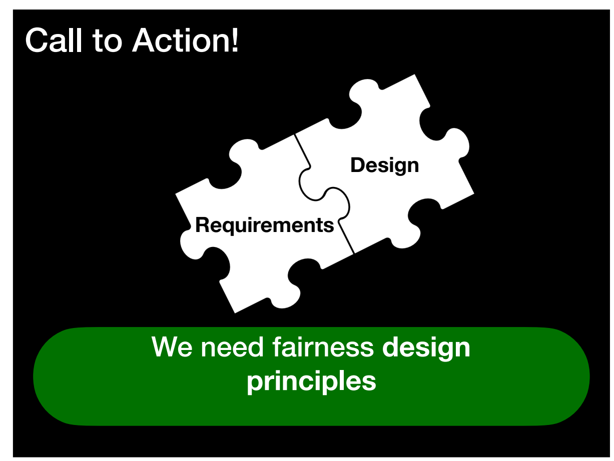

Design software to be fair
Typically machine learning systems:
• Balance training sets
• Introduce training noise
• Constrain regression's loss function
• Split criteria on sensitive inputs

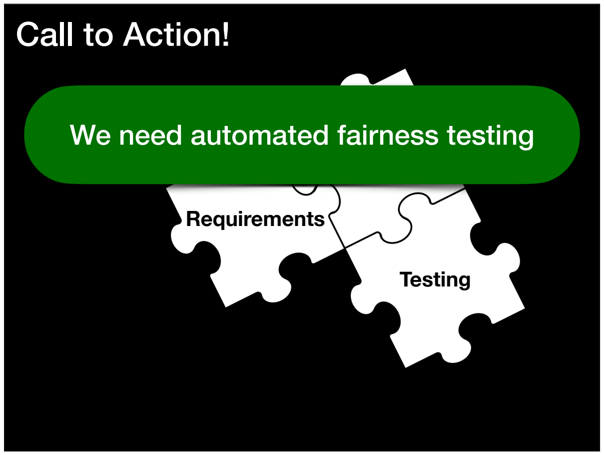

Design alone is not enough

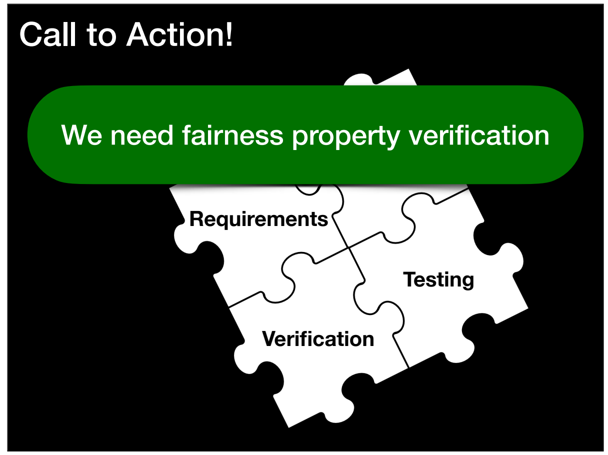

possible causes
biased data
poor design
implementation bugs
unintended interactions and mismatched components

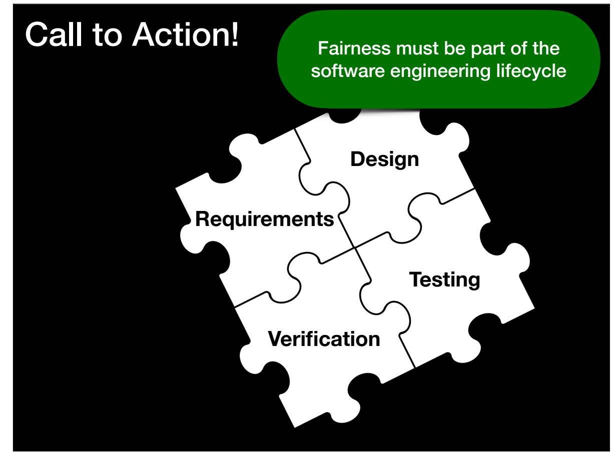

Fairness is just like quality and security
Fairness must be part of the software engineering lifecycle


Call to Action!
Fairness must be part of the software engineering lifecycle
Requirements
We need methods for specifying fairness requirements

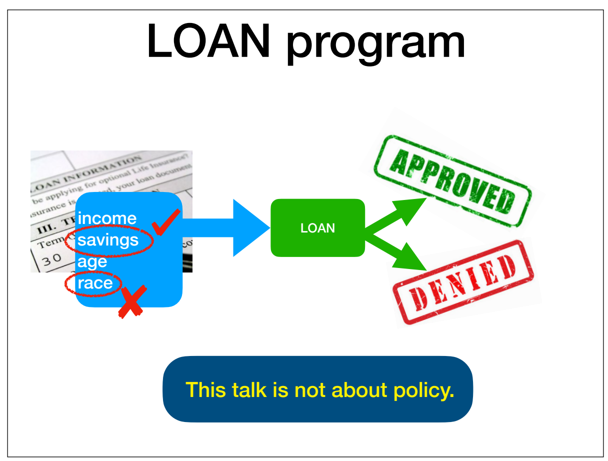

Call to Action!
Design
Requirements
We need fairness design principles

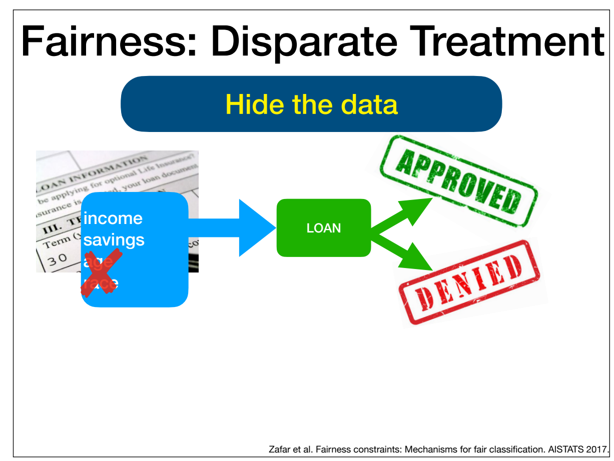

Call to Action!
We need automated fairness testing
Requirements
Testing


Call to Action!
We need fairness property verification
Requirements
Testing
Verification


Call to Action!
Fairness must be part of the software engineering lifecycle
Design
Requirements
Testing
Verification

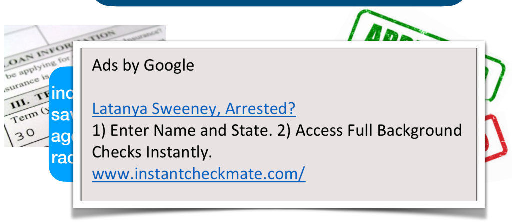

Let's talk about
what it means for
systems to discriminate


LOAN program
income
savings
age
race
LOAN
APPROVED
DENIED
This talk is not about policy.


Fairness: Disparate Treatment
Hide the data
income
savings
LOAN
APPROVED
DENIED
Zafar et al. Fairness constraints: Mechanisms for fair classification. AISTATS 2017.

Fairness: Disparate Treatment

Hide the data



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

BUSINESS INSIDER

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

Isobel Asher Hamilton

- 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.

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

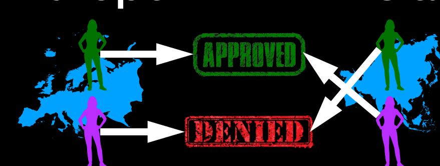
often called group discrimination

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

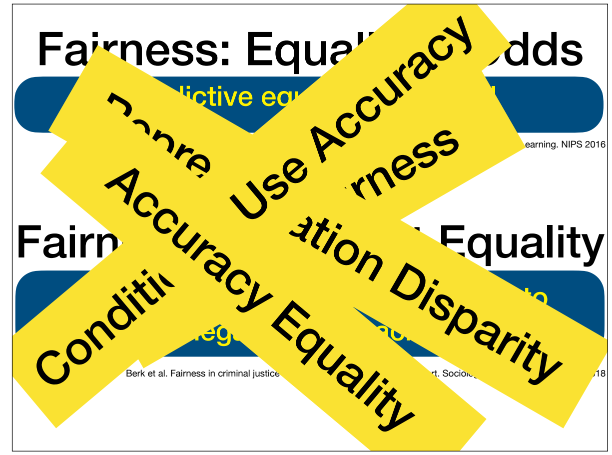
False positive rates should not differ

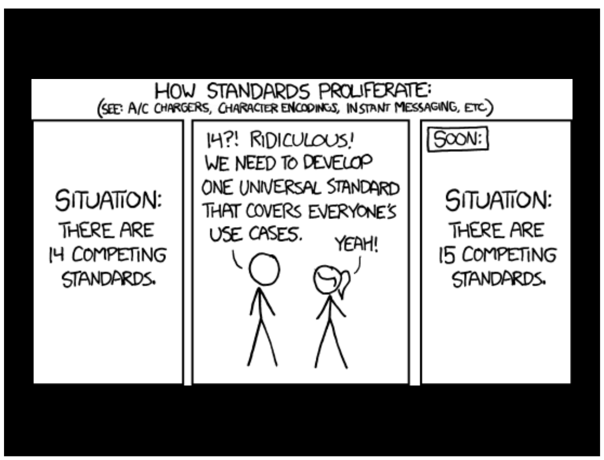
Chouldechova. Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. FATML 2016
Corbett-Davies. Algorithmic decision making and the cost of fairness. KDD 2017

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





Fairness: Correlation

correlation(race, APPROVED) = 0.8

mutual information(race, APPROVED) = 0.6

Correlation does not measure causation

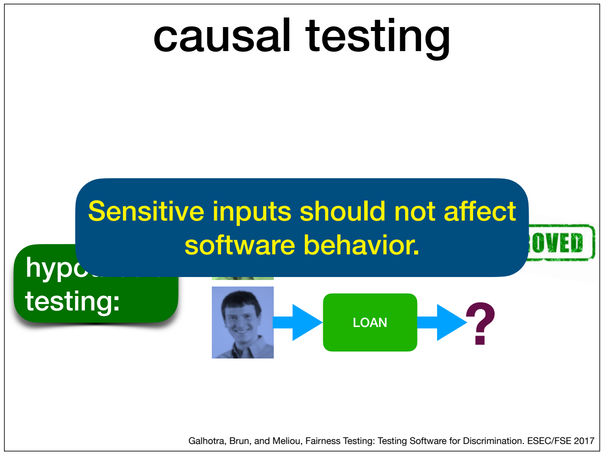
Atlidakis et al. FairTest: Discovering unwarranted associations in data-driven applications. EuroS&P 2017

What is fairness?

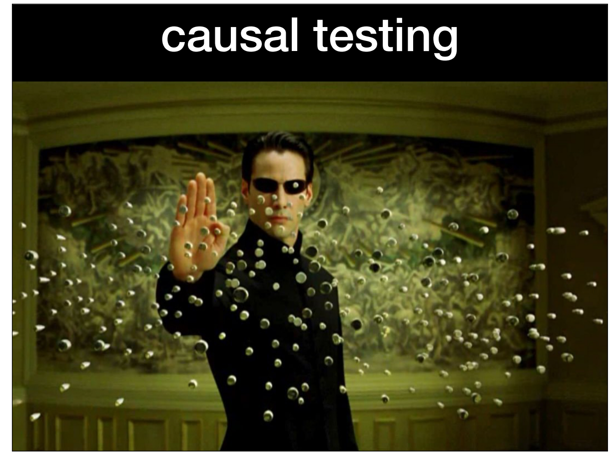
Sensitive inputs should not affect software behavior.

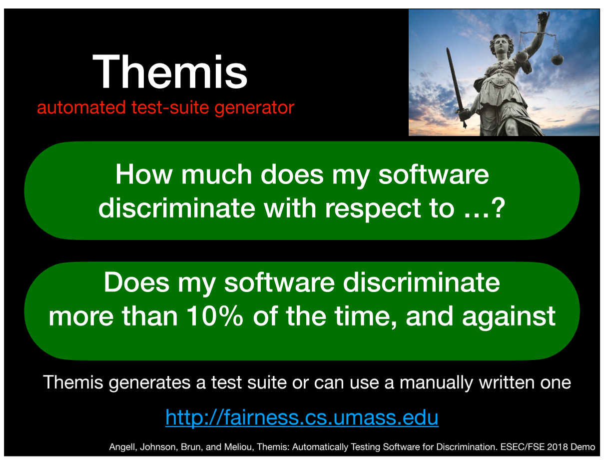
We want to measure causality!

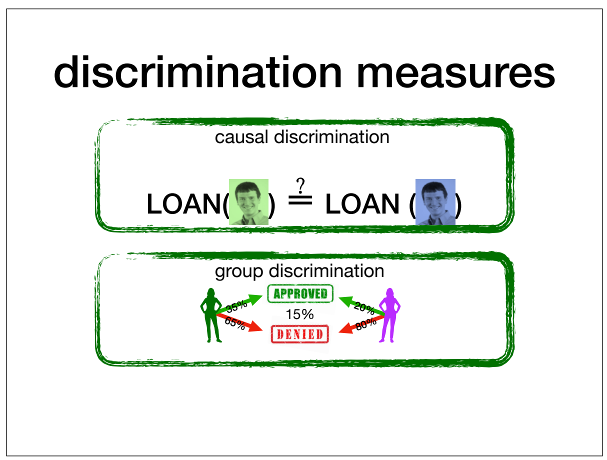
Judea Pearl. Causal inference in statistics: An overview. Statistics Surveys 2009

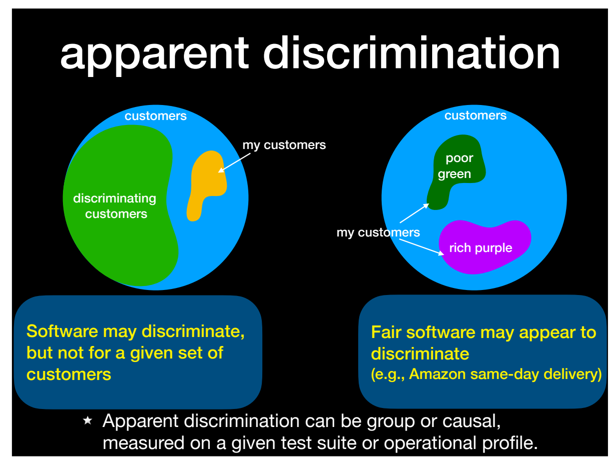

causal testing
Sensitive inputs should not affect software behavior.
hypo
testing:
OVED
LOAN
?
Galhotra, Brun, and Meliou, Fairness Testing: Testing Software for Discrimination. ESEC/FSE 2017


causal testing
No need for an oracle!


causal testing

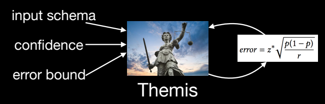

Themis
automated test-suite generator
How much does my software discriminate with respect to …?
Does my software discriminate more than 10% of the time, and against
Themis generates a test suite or can use a manually written one
http://fairness.cs.umass.edu
Angell, Johnson, Brun, and Meliou, Themis: Automatically Testing Software for Discrimination. ESEC/FSE 2018 Demo


discrimination measures
causal discrimination
LOAN( ) ≟ LOAN ( )
group discrimination
APPROVED
15%
DENIED

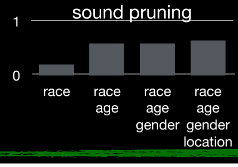

apparent discrimination
customers
my customers
discriminating customers
customers
poor green
my customers
rich purple
Software may discriminate, but not for a given set of customers
Fair software may appear to discriminate
(e.g., Amazon same-day delivery)
★ Apparent discrimination can be group or causal, measured on a given test suite or operational profile.

How does Themis work?

adaptive, confidence-driven sampling



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

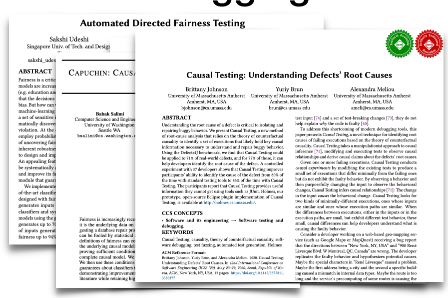
Training a decision tree not to discriminate against gender made it discriminate against race 38.4% of the time.

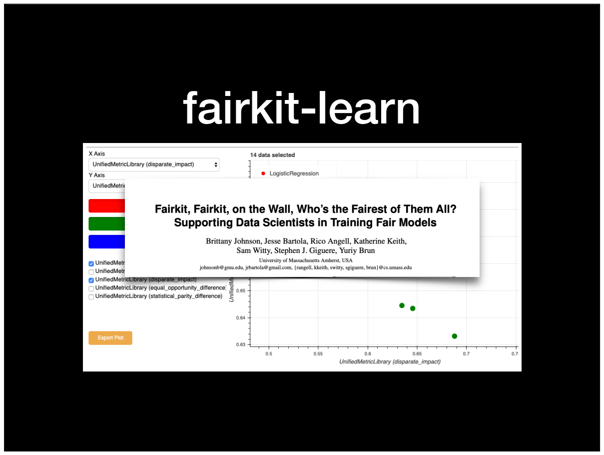
findings

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×.

Debugging




fairkit-learn
X Axis
UnifiedMetricLibrary (disparate_impact)
Y Axis
14 data selected
LogisticRegression
Fairkit, Fairkit, on the Wall, Who's the Fairest of Them All?
Supporting Data Scientists in Training Fair Models
Brittany Johnson, Jesse Bartola, Rico Angell, Katherine Keith,
Sam Witty, Stephen J. Giguere, Yuriy Brun
University of Massachusetts Amherst, USA
johnsonb@gmu.edu, jrbartola@gmail.com, {rangell, kkeith, switty, sgiguere, brun}@cs.umass.edu
UnifiedMetricLibrary (disparate_impact)
UnifiedMetricLibrary (equal_opportunity_difference)
UnifiedMetricLibrary (statistical_parity_difference)
Export Plot
UnifiedMetricLibrary (disparate_impact)