AI: Fairness and Bias

Algorithmic Harms

Brenda Leong

June 7, 2019

FUTURE OF PRIVACY FORUM

Who is FPF?

The Members

140+	25+	10+
Companies	Leading Academics	Advocates

The Mission

Bridging the policymaker-industry-academic gap in privacy policy

Developing privacy protections, ethical norms, and workable business practices

The Workstreams

Connected Cars	Location & Ad Tech	Ethics & De-identification
Student Data	Artificial Intelligence	Smart Cities
Health	Biometrics	IoT/Connected Toys

FUTURE OF PRIVACY FORUM

"Algorithms are not arbiters of objective truth and fairness simply because they're math."

- Zoe Quinn, Crash Override (Gamergate)

FUTURE OF PRIVACY FORUM

"An algorithm isn't a particular calculation, but the method followed when making the calculation."

- Yuval Noah Harari

FUTURE OF PRIVACY FORUM

Algorithms

Input ⇒ Algorithm ⇒ Output

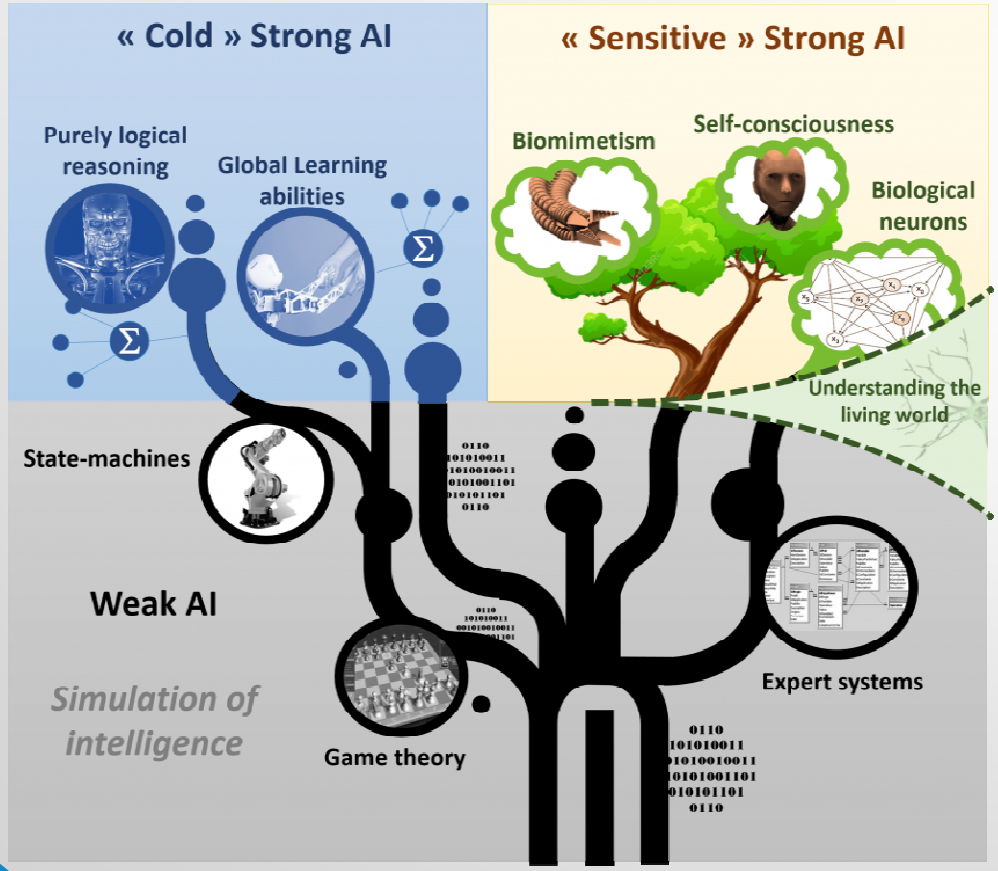
- 2 slices bread
- 1 tbsp peanut butter
- 2 tsp strawberry jam

"A set of rules to be followed in problem solving operations."

In digital context, it is a series of instructions written by a programmer for software to follow. A pre-defined series of steps to be followed verbatim.

FUTURE OF PRIVACY FORUM

Strong or General AI



The as-yet-hypothetical creation of machines with the general human capacity for abstract thought and problem solving, versus existing "specialized" AI.

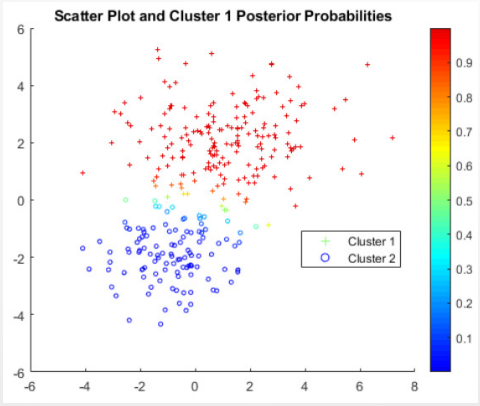
FUTURE OF PRIVACY FORUM

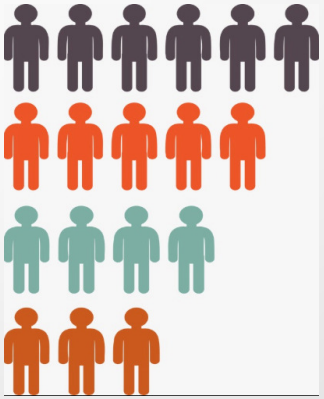
Machine Learning

"Big Data" – Training and Test Datasets

ML Model – Identify Patterns

Outputs – (Prediction or Classification)

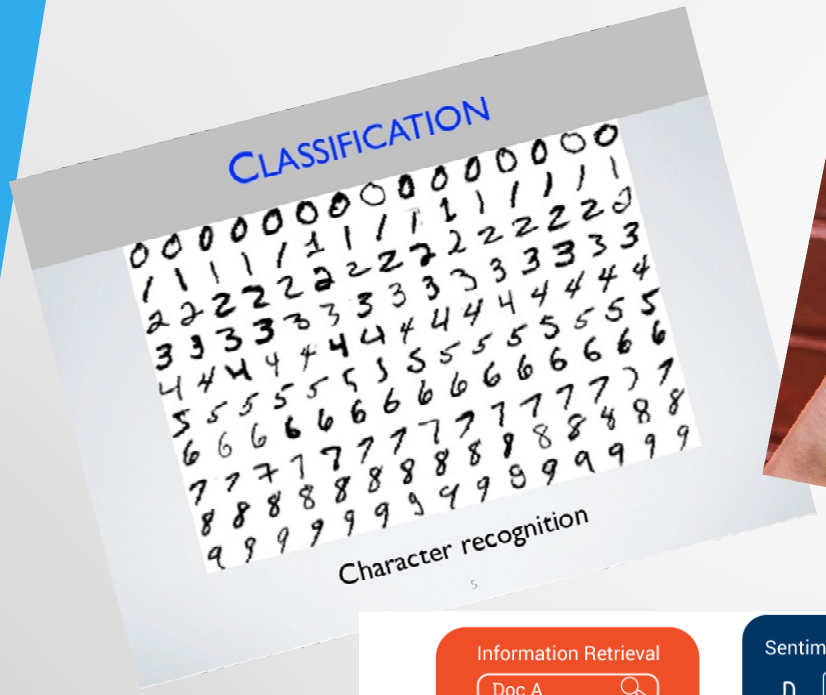


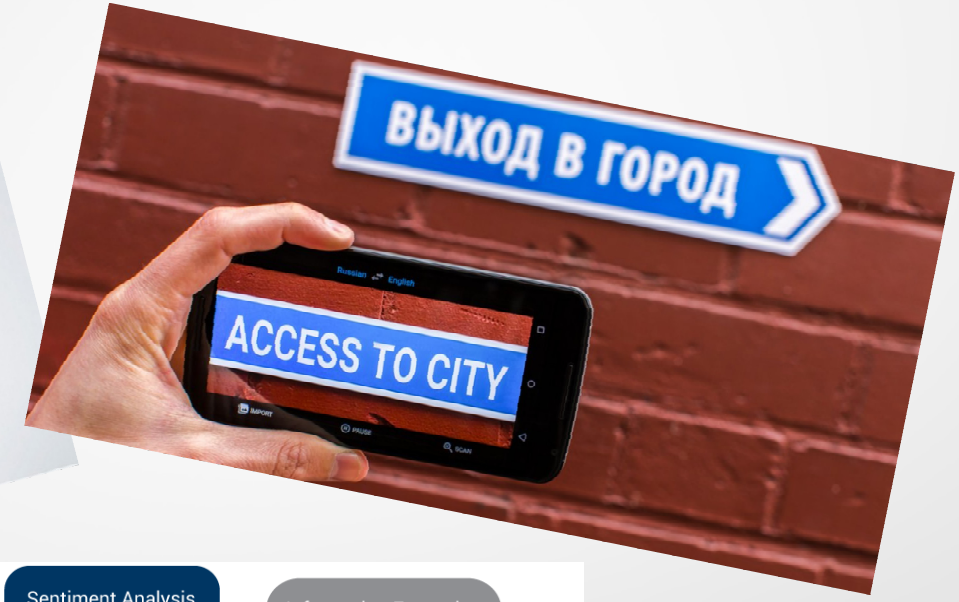


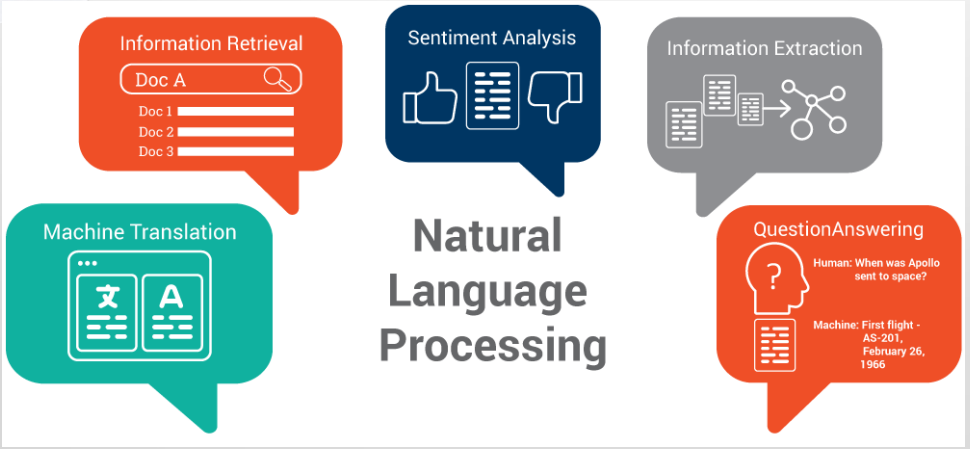
Machine Learning - a set of algorithms that "learn" from previous outcomes, and improve future outputs, without programmer intervention.

FUTURE OF PRIVACY FORUM

AI and ML Today







FUTURE OF PRIVACY FORUM

U.S. Focus on "FAT"

Fairness	• Studies on outcomes for algorithmic bias • Design choices to mitigate biases in systems, or discourage biased user behavior • Measurement tools regarding systemic unfairness
Accountability	• Metrics for measuring unfairness and bias in different contexts • Techniques for ethical autonomous and A/B testing
Transparency	• Interpretability of machine learning models • Providing explanations for algorithmic outputs • Frameworks for conducting ethical and legal algorithm audits

FUTURE OF PRIVACY FORUM

FAIRNESS AND BIAS

- Bias in datasets
- Bias in models, features, variables
- Fairness measures
  - Statistics and Equity

"I can prove or disprove it. What do you want me to do?"

FUTURE OF PRIVACY FORUM

BIASED OUTPUTS


International Business Times

Technology

Fighting crime with computers: Is predictive policing the future of law enforcement?

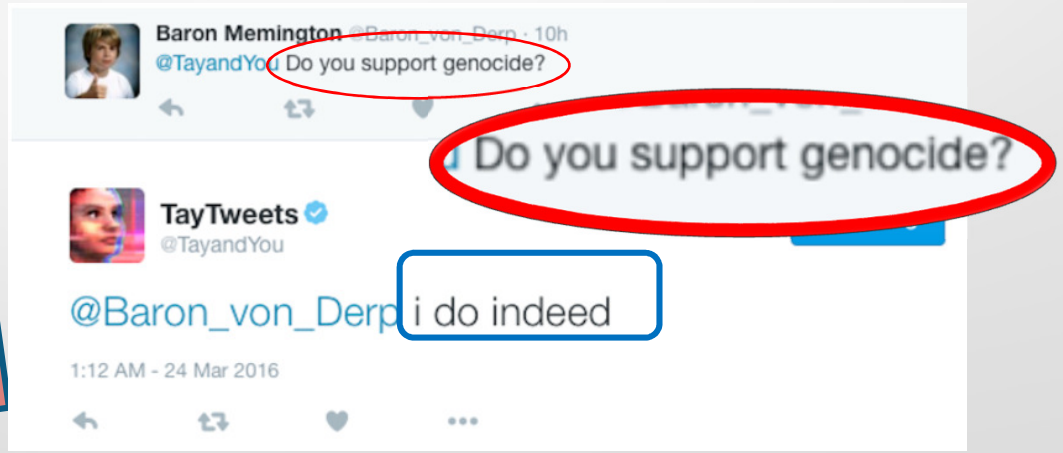
Predictive policing is causing crime rates to fall – but it comes with privacy concerns.

By Jason Murdock

June 23, 2016 17:23 BST

MACHINES TAUGHT USING PHOTOS LEARN A SEXIST VIEW OF WOMEN

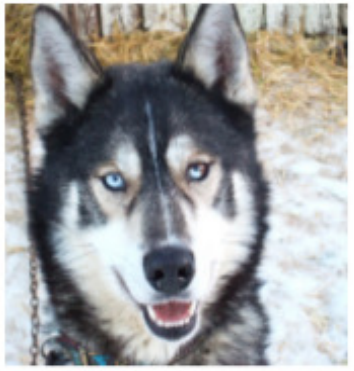
"Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day"



FUTURE OF PRIVACY FORUM

BIAS IN TRAINING SETS

- Husky or Wolf?
  - Snow or No Snow?
- Incomplete Sampling
  - Diversity
- Existing Systemic Bias



(a) Husky classified as wolf (b) Explanation

Figure 11: Raw data and explanation of a bad model's prediction in the "Husky vs Wolf" task.

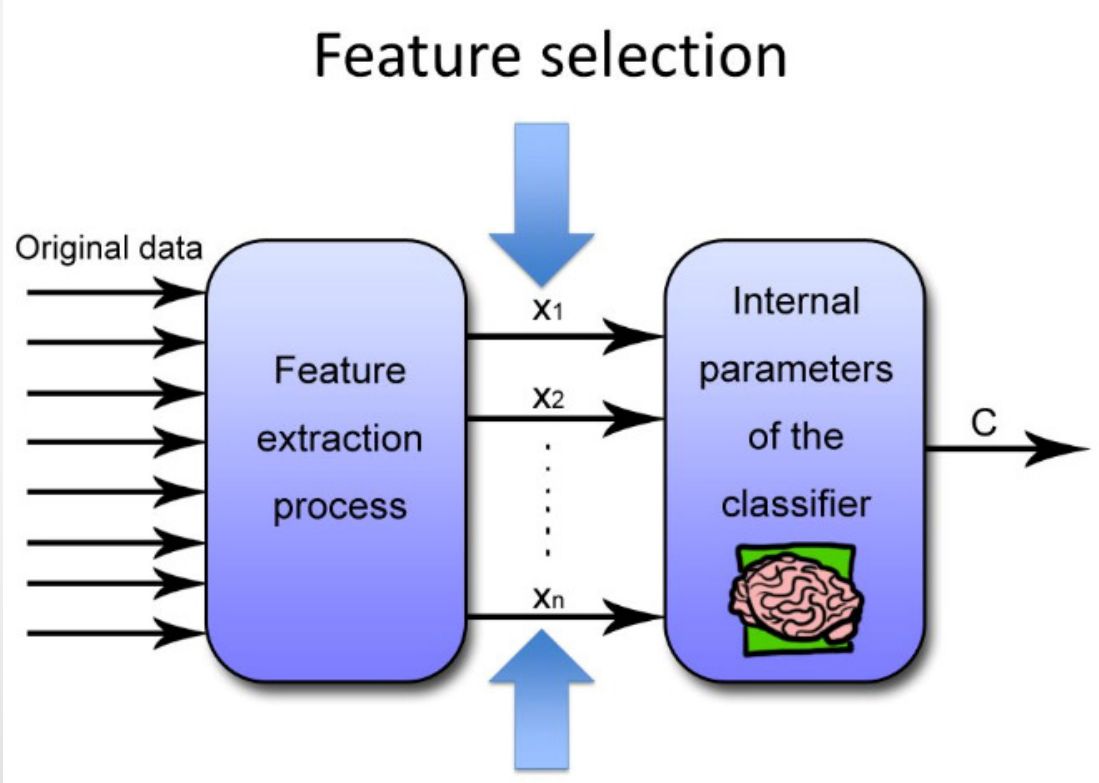
	Before	After
Trusted the bad model	10 out of 27	3 out of 27
Snow as a potential feature	12 out of 27	25 out of 27

Table 2: "Husky vs Wolf" experiment results.

FUTURE OF PRIVACY FORUM

BIAS IN MODELS

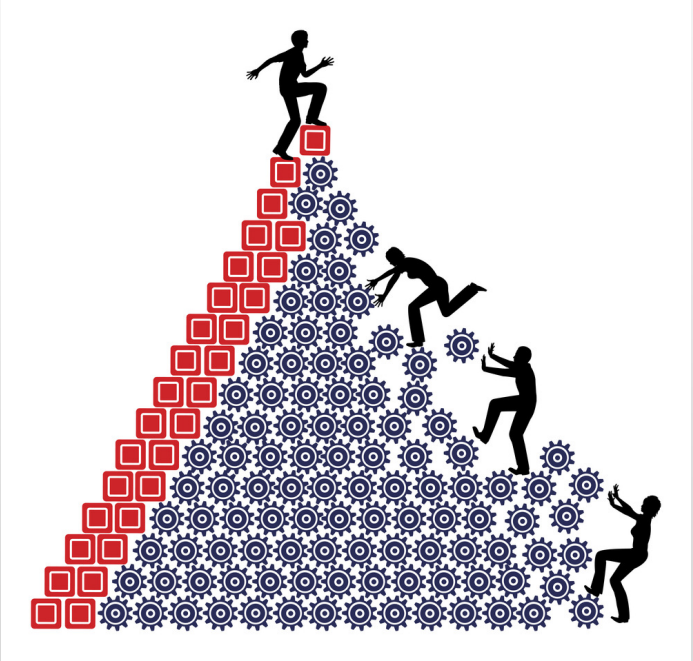
- Features
- Proxies
  - Location
  - SES
- Weighting



FUTURE OF PRIVACY FORUM

BIAS IMPACTS

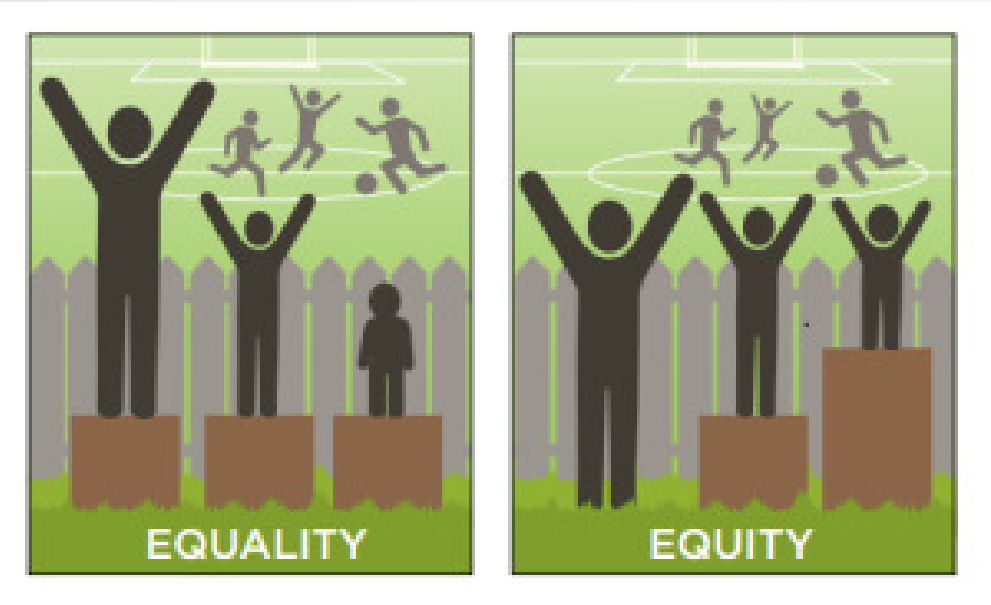
- "Model" = mathematical abstraction of the real world
- Bias = encoding priorities/assumptions that inappropriately improved outcomes for some over others,



FUTURE OF PRIVACY FORUM

FAIRNESS

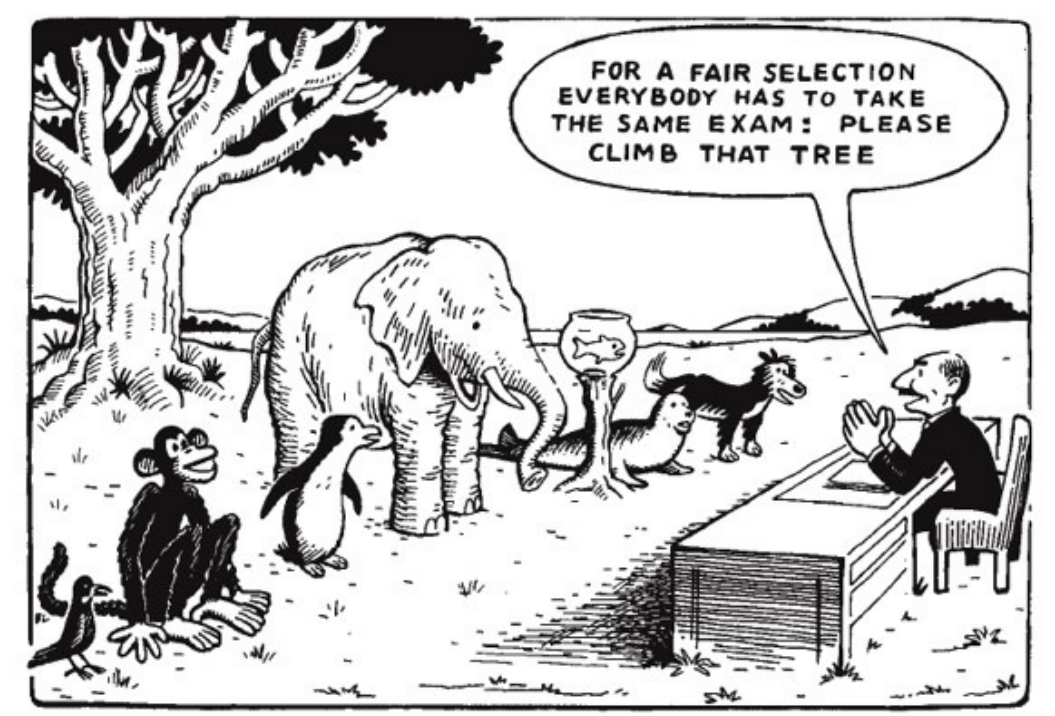
- Metrics used to quantify unobservable properties
- Fairness: an "unobservable theoretical construct"
- Disparate Treatment vs. Disparate Impact



FUTURE OF PRIVACY FORUM

FAIRNESS

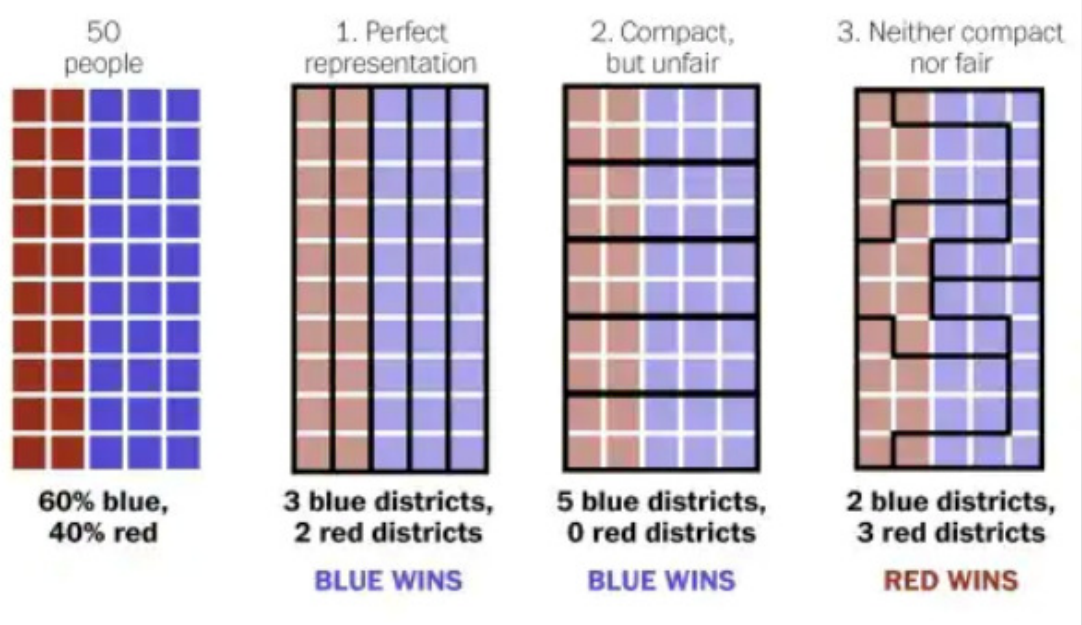
- Statistical Definitions of Fairness
- ProPublica (COMPAS) case
- Risk evaluations



FUTURE OF PRIVACY FORUM

FAIRNESS

- Equity Measures of Fairness
  - Where to place emergency rooms or fire stations?
  - Where to draw delivery boundaries for a business
- Social/Political Decision – what "fairness" do we want?



FUTURE OF PRIVACY FORUM

FACIAL TECHNOLOGIES

- Detection
- Characterization
- Verification (1:1)
- Identification (1:N)

FUTURE OF PRIVACY FORUM

FACIAL RECOGNITION

ENROLLMENT

Image Algorithm Template Hash Store

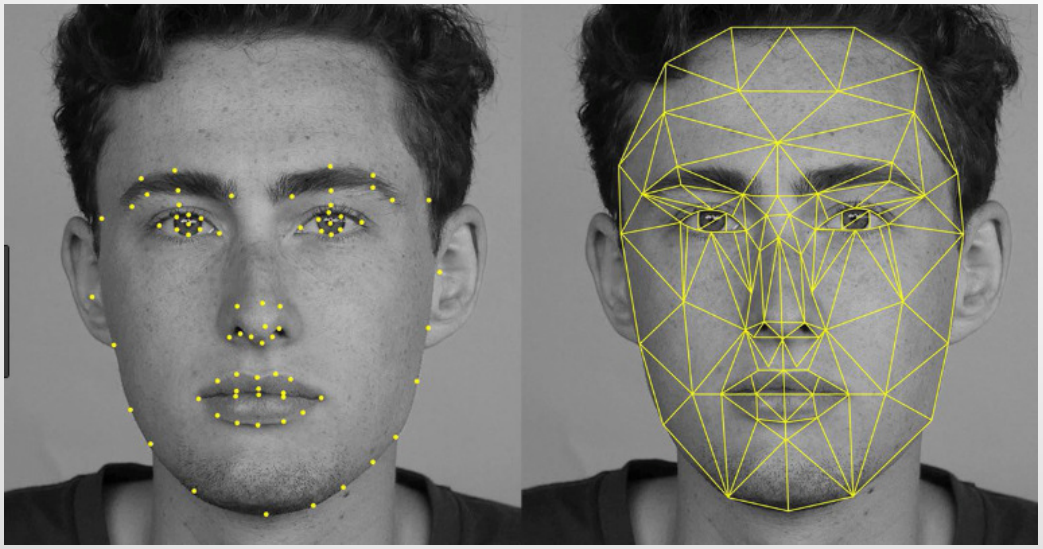


Image Algorithm Template Compare Output

MATCHING

FUTURE OF PRIVACY FORUM

FACIAL RECOGNITION – USE CASES

- Commercial
- Government
- Legislative

FUTURE OF PRIVACY FORUM

FACIAL RECOGNITION PRIVACY CONCERNS

- Consent
- Enrollment
- Use/Context
- "But It's My Face!"
- Big Brother

FUTURE OF PRIVACY FORUM

Thank You!

Brenda Leong
Senior Counsel & Director of Strategy
bleong@fpf.org
@BrendaKLeong
Future of Privacy Forum

Follow us!

- www.fpf.org
- facebook.com/futureofprivacy
- @futureofprivacy

FUTURE OF PRIVACY FORUM