Responsible Data Science

Introduction and Overview

January 23, 2023

Prof. Julia Stoyanovich

Center for Data Science &
Computer Science and Engineering
New York University

NYU TANDON SCHOOL OF ENGINEERING
NYU Center for Data Science
r/ai

course logistics

Instructor: Julia Stoyanovich

Assoc. Prof. of Data Science, Computer Science & Engineering
Director of the Center for Responsible AI (R/AI)
New York University

Ph.D. in CS from Columbia University
B.S. in CS & Math from UMass Amherst

Research: data and knowledge management ("databases")

- Responsible Data Science (RDS)
- Preferences and Voting (DB + COMSOC)
- Evolving graphs (Big Data / Systems)

And also:

- Tech policy: NYC ADS task force, NYC algorithmic hiring bill, IEEE AI policy committee, LCO of Ontario, SmartCityPhl task force
- Education & public engagement: Responsible Data Science, We are AI

Office hours: Tuesdays 11am-noon EST and by appointment

Teaching Assistants

Lucius Bynum
Office hours: Wednesdays 2–3pm

Lucas Rosenblatt
Office hours: Thursdays 3-4pm

Assignments and grading

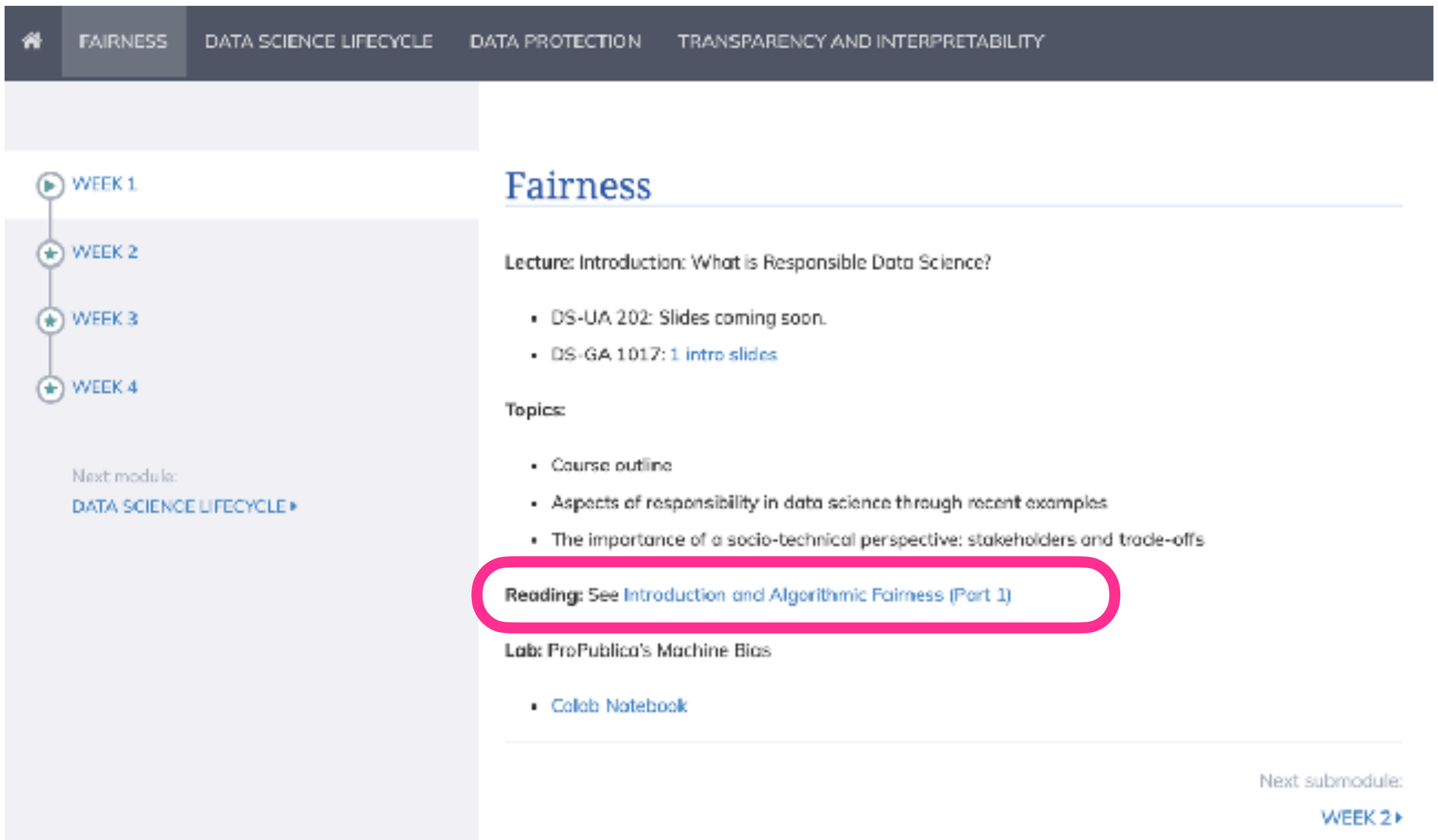
Grading: homeworks - 10% x 3 = 30%
project - 30%
final exam - 20%
labs - 10%
quizzes - 10%

No credit for late homeworks. 2 late days over the term, no questions asked. If a homework is submitted late — a day is used in full.

Assignment schedule posted to Bright Space (under Course information), subject to change.

Where to find information

Website: https://dataresponsibly.github.io/rds/ slides, reading, labs



Bright Space: everything assignment-related, Zoom links for lectures and labs, announcements. **Piazza:** discussion board.

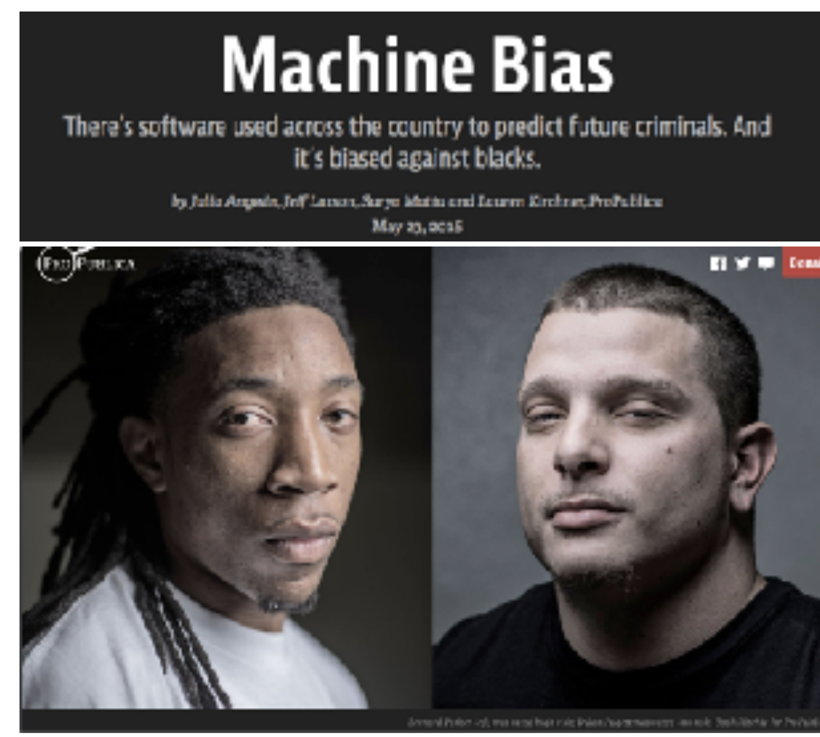
This week's reading

DOI:10.1145/3376898

A group of industry, academic, and government experts convene in Philadelphia to explore the roots of algorithmic bias.

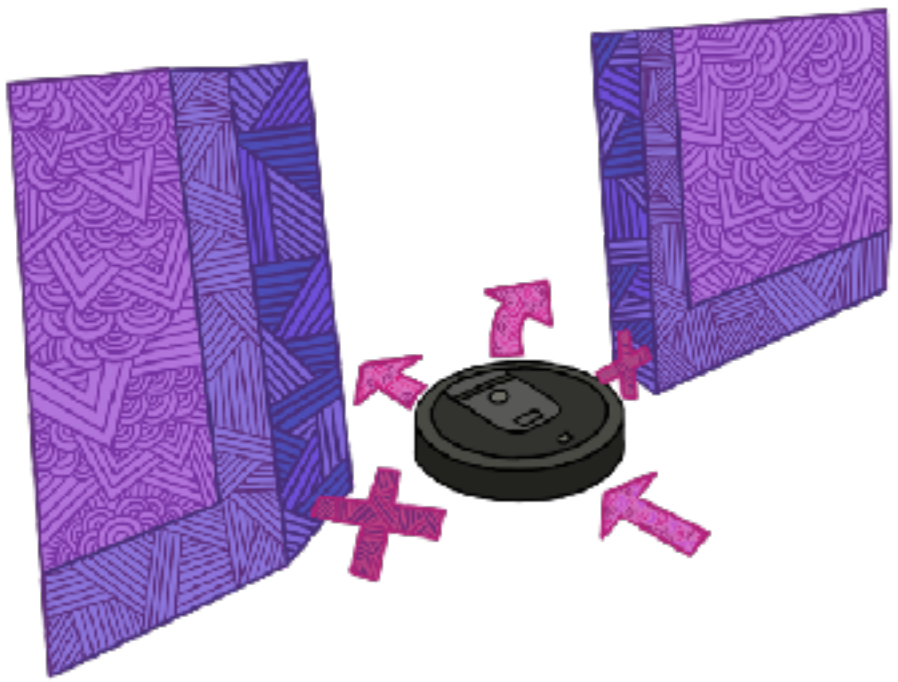
BY ALEXANDRA CHOULDECHOVA AND AARON ROTH

A Snapshot of the Frontiers of Fairness in Machine Learning



what is responsible AI?

AI: algorithms, data, decisions



Artificial Intelligence (AI)

a **system** in which **algorithms** use **data** and make **decisions** on our behalf, or help us make decisions

The promise of AI

Opportunity

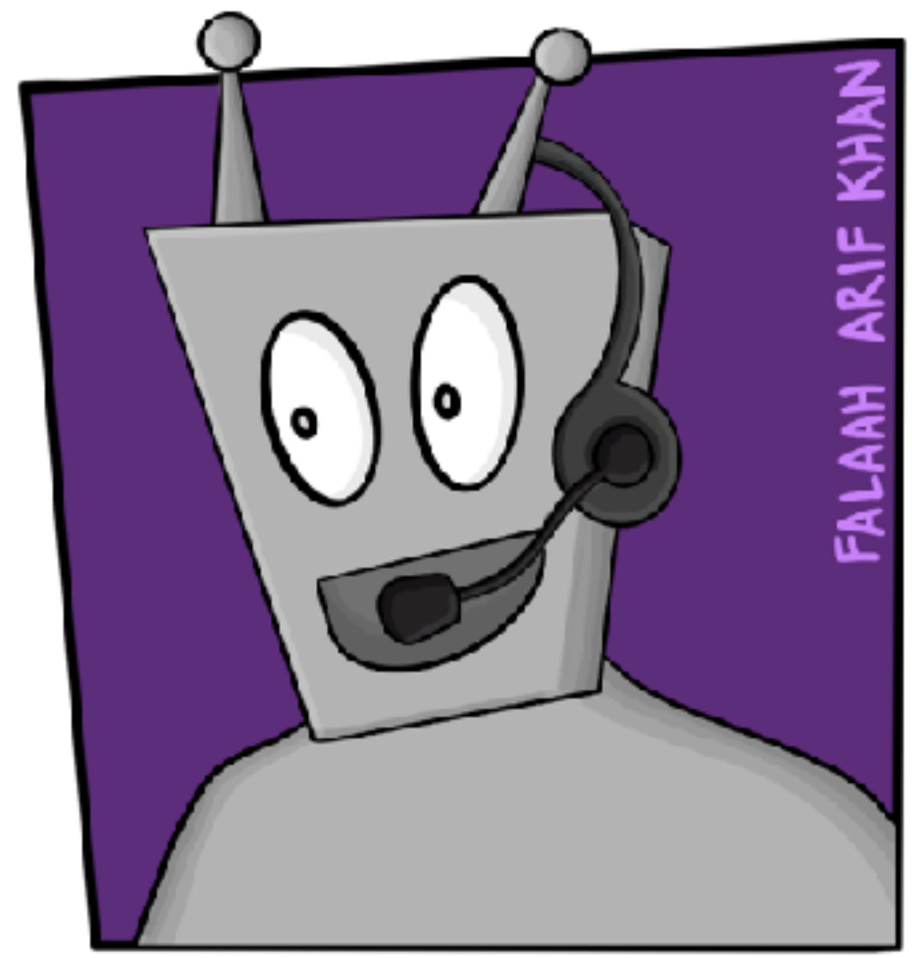
make our lives convenient

accelerate science

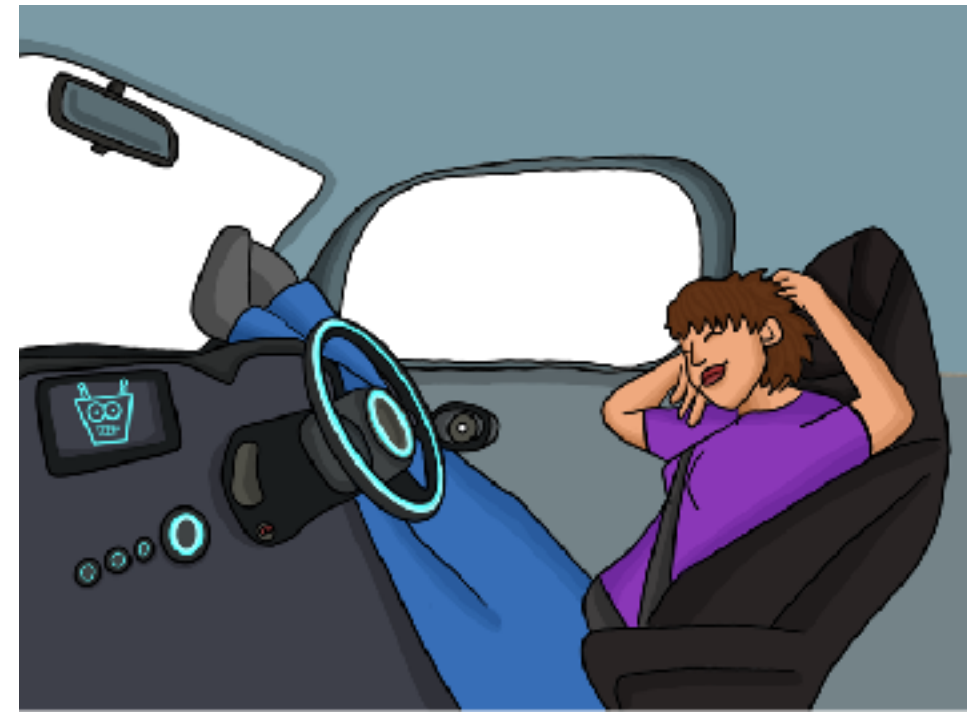
boost innovation

transform government

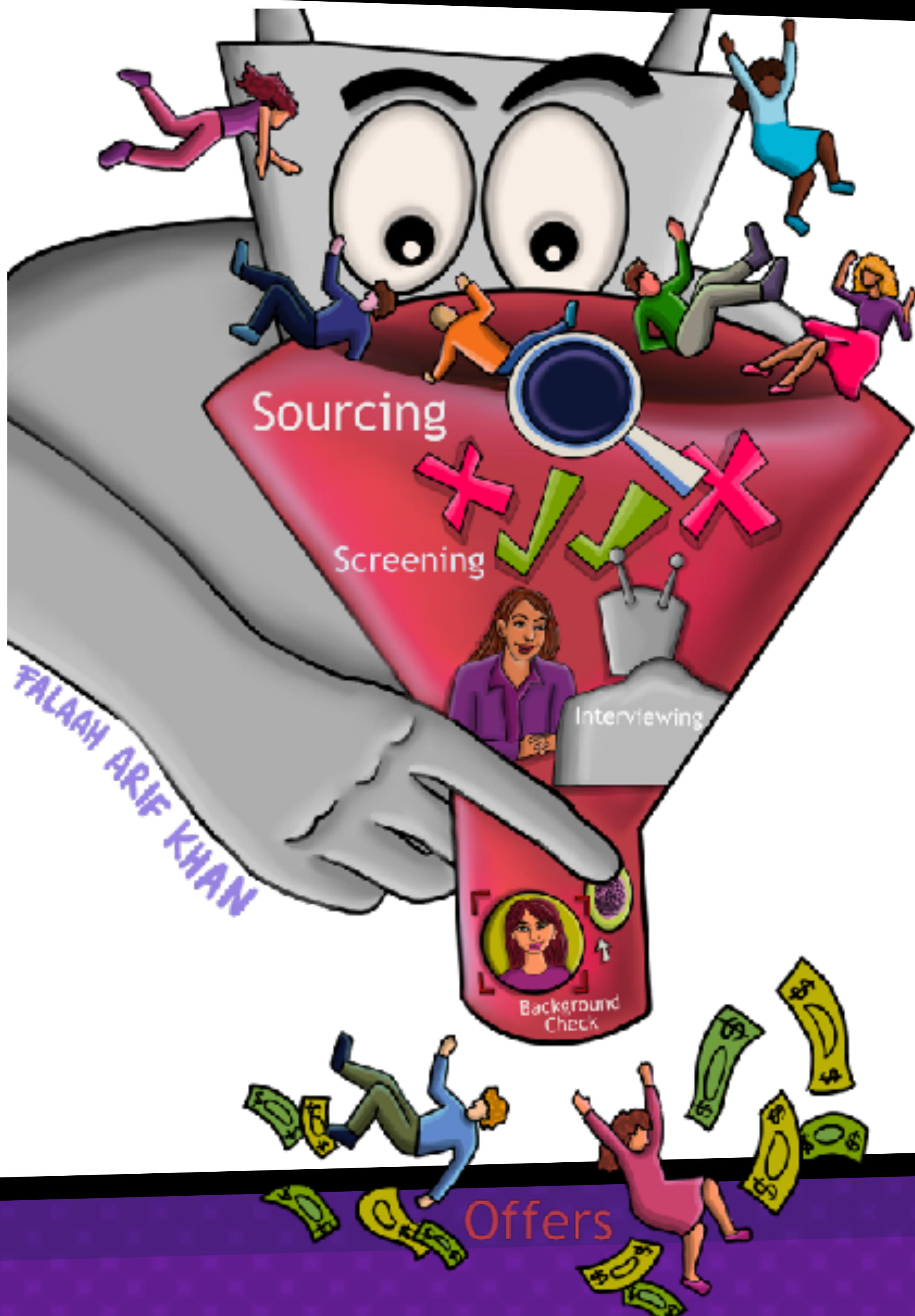
Machines make mistakes

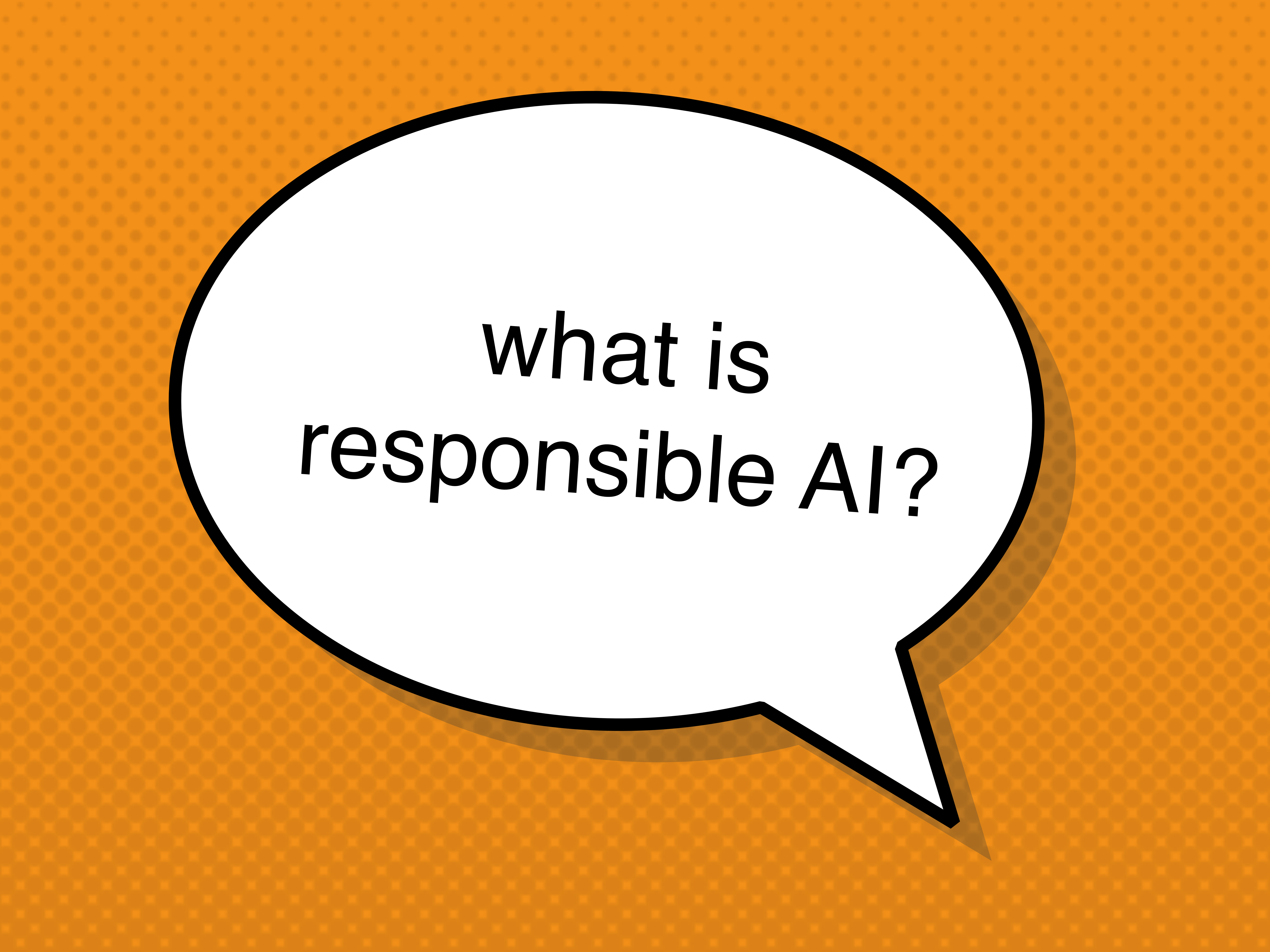


Mistakes lead to harms



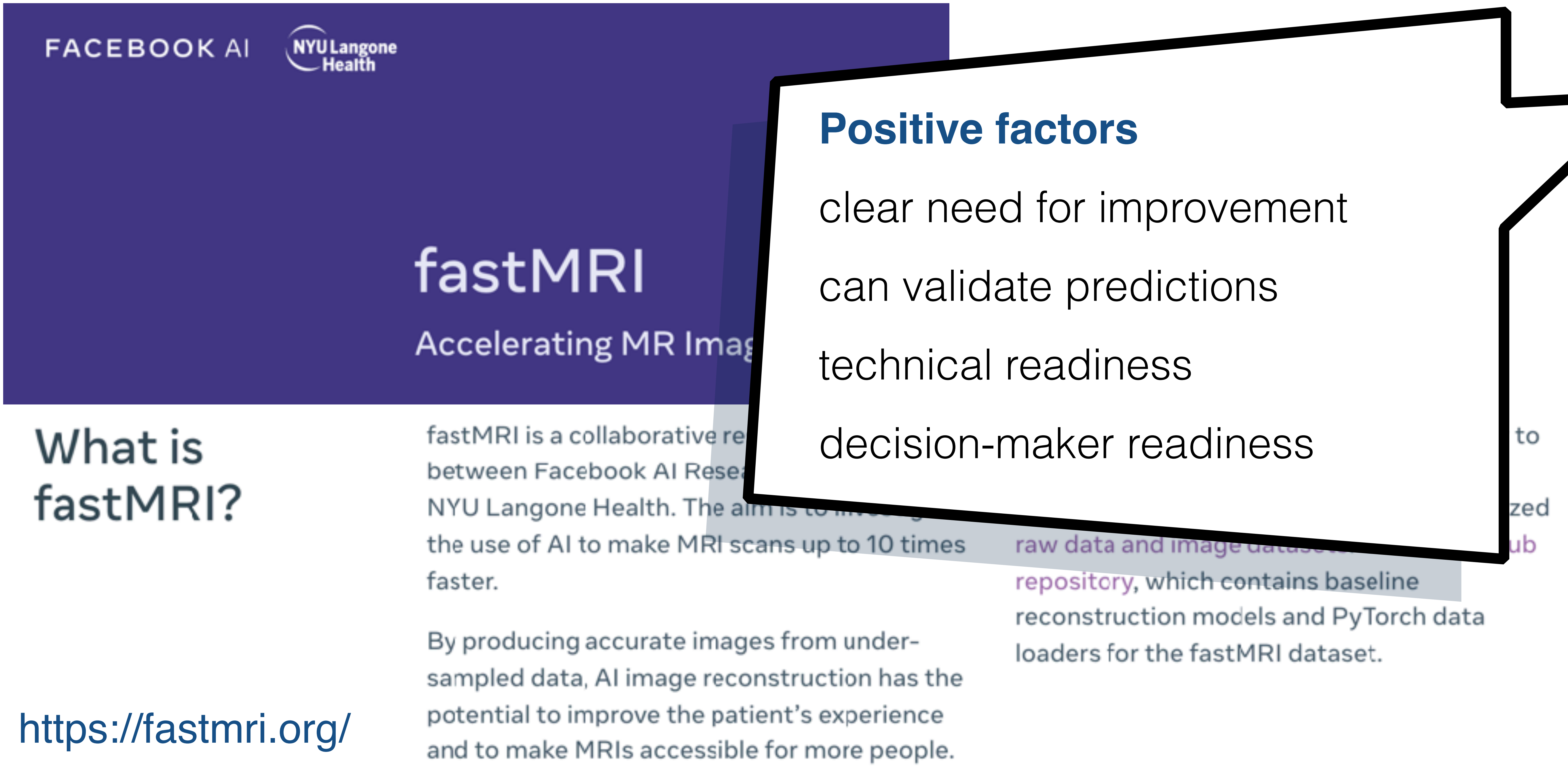
Harms can be cumulative




what is responsible AI?

more examples

Medical imaging



https://fastmri.org/

Positive factors

clear need for improvement

can validate predictions

technical readiness

decision-maker readiness

Automated hiring systems

MIT Technology Review February 2013

Racism is Poisoning Online Ad Delivery, Says Harvard Professor

The New York Times March 2021

We Need Laws to Take On Racism and Sexism in Hiring Technology

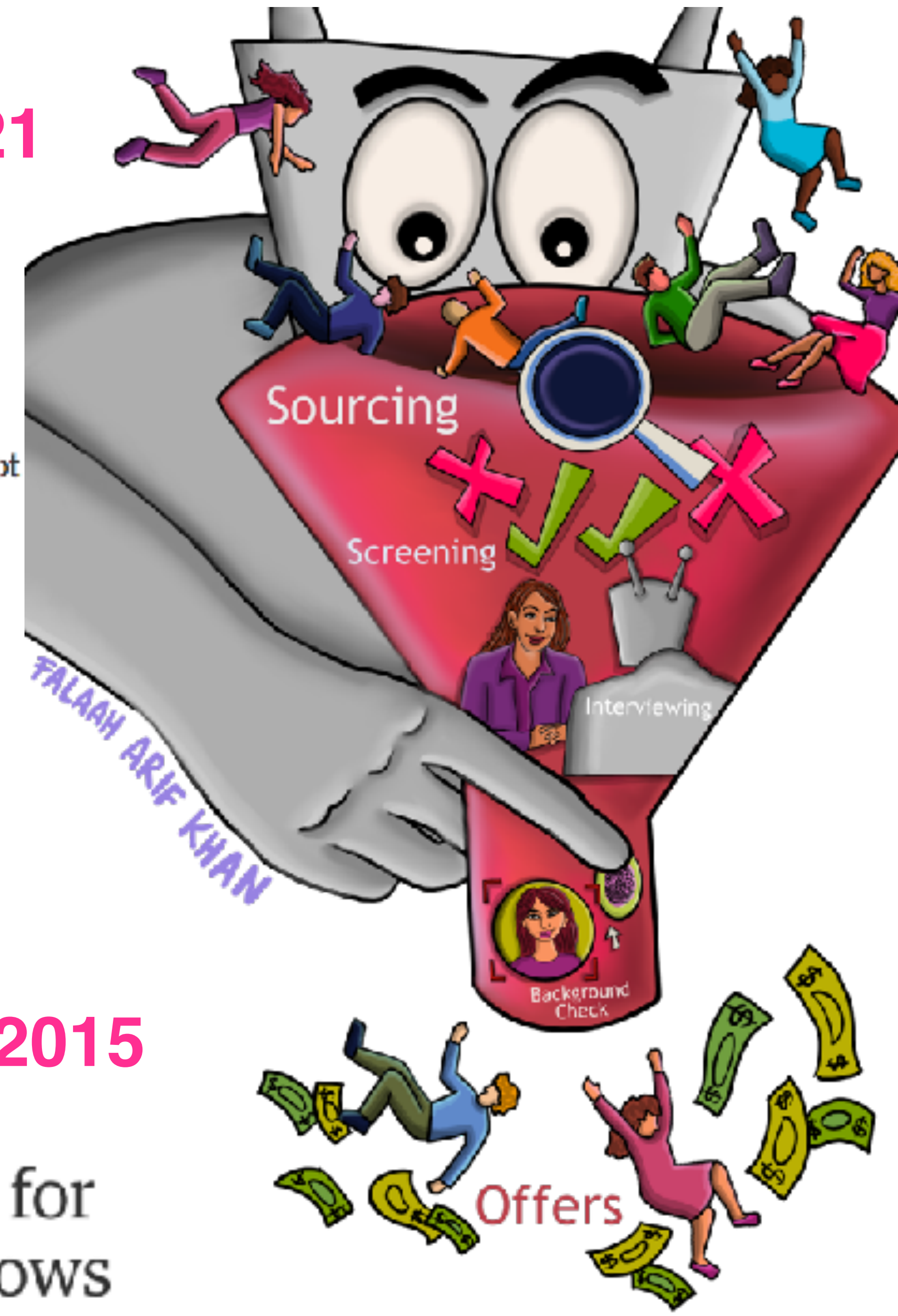
Artificial intelligence used to evaluate job candidates must not become a tool that exacerbates discrimination.

REUTERS October 2018

Amazon scraps secret AI recruiting tool that showed bias against women

theguardian July 2015

Women less likely to be shown ads for high-paid jobs on Google, study shows



Hiring before automation

Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination

September 2004

Marianne Bertrand
Sendhil Mullainathan

AMERICAN ECONOMIC REVIEW
VOL. 94, NO. 4, SEPTEMBER 2004
(pp. 991-1013)

We study race in the labor market by sending fictitious resumes to help-wanted ads in Boston and Chicago newspapers. To manipulate perceived race, resumes are randomly assigned African-American- or White-sounding names. **White names receive 50 percent more callbacks for interviews.** Callbacks are also more responsive to resume quality for White names than for African-American ones. The racial gap is uniform across occupation, industry, and employer size. We also find little evidence that employers are inferring social class from the names. Differential treatment by race still appears to still be prominent in the U. S. labor market.

discussion

Describe a use case

what are the **goals** of the AI system?

what are the **benefits** and to **whom**?

what are the **harms** and to **whom**?

Use case: Staples discounts

THE WALL STREET JOURNAL.

December 2012

WHAT THEY KNOW

Websites Vary Prices, Deals Based on Users' Information

By Jennifer Valentino-DeVries, Jeremy Singer-Vine and Ashkan Soltani
December 24, 2012

WHAT PRICE WOULD YOU SEE?

It was the same Swingline stapler, on the same Staples.com website. But for Kim Wamble, the price was $15.79, while the price on Trude Frizzell's screen, just a few miles away, was $14.29.

A key difference: where Staples seemed to think they were located.

A Wall Street Journal investigation found that the Staples Inc. website displays different prices to people after estimating their locations. More than that, **Staples appeared to consider the person's distance from a rival brick-and-mortar store**, either OfficeMax Inc. or Office Depot Inc. If rival stores were within 20 miles or so, Staples.com usually showed a discounted price.

https://www.wsj.com/articles/SB10001424127887323777204578189391813881534

r/ai

Use case: AdFisher

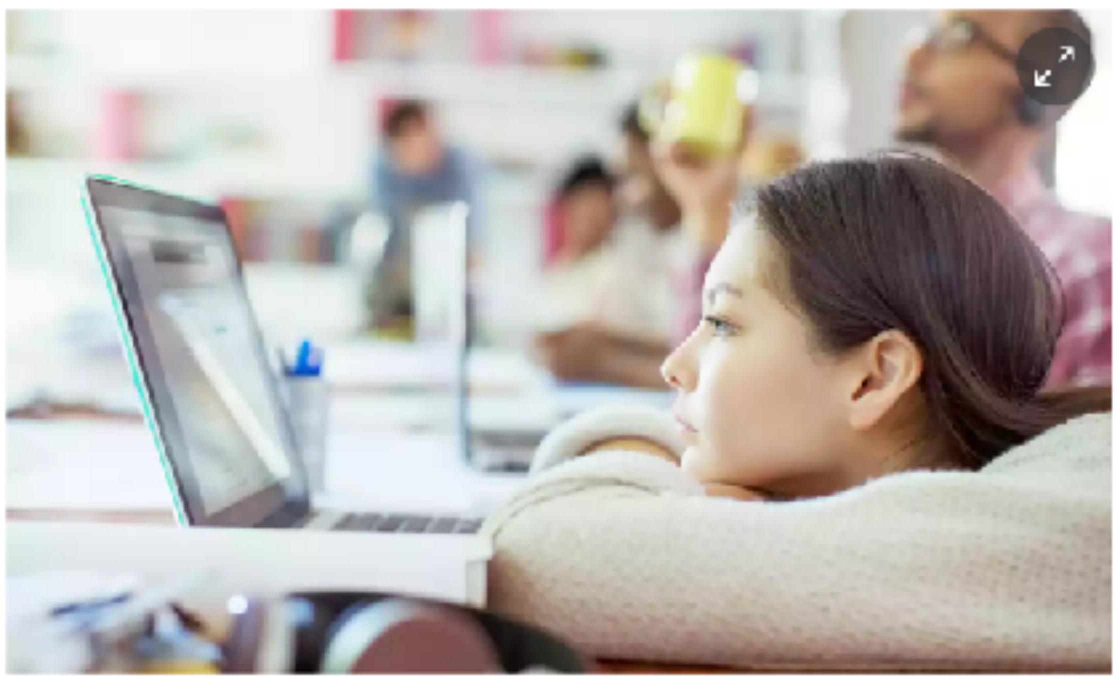
theguardian

July 2015

Samuel Gibbs

Wednesday 8 July 2015 11.29 BST

Automated testing and analysis of company's advertising system reveals male job seekers are shown far more adverts for high-paying executive jobs



One experiment showed that Google displayed adverts for a career coaching service for executive jobs 1,852 times to the male group and only 318 times to the female group. Photograph: Alamy

Women less likely to be shown ads for high-paid jobs on Google, study shows

The AdFisher tool simulated job seekers that did not differ in browsing behavior, preferences or demographic characteristics, except in gender.

One experiment showed that Google displayed ads for a career coaching service for "$200k+" executive jobs **1,852 times to the male group and only 318 times to the female group**. Another experiment, in July 2014, showed a similar trend but was not statistically significant.

https://www.theguardian.com/technology/2015/jul/08/women-less-likely-ads-high-paid-jobs-google-study

r/ai

Use case: Resume screening

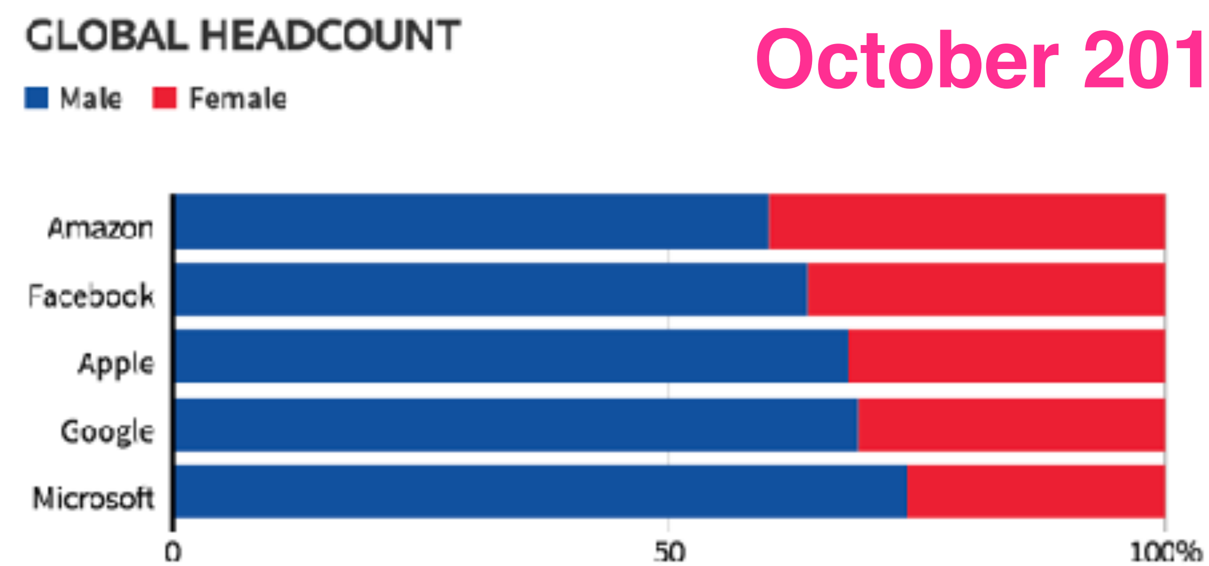
REUTERS Jeffrey Dastin

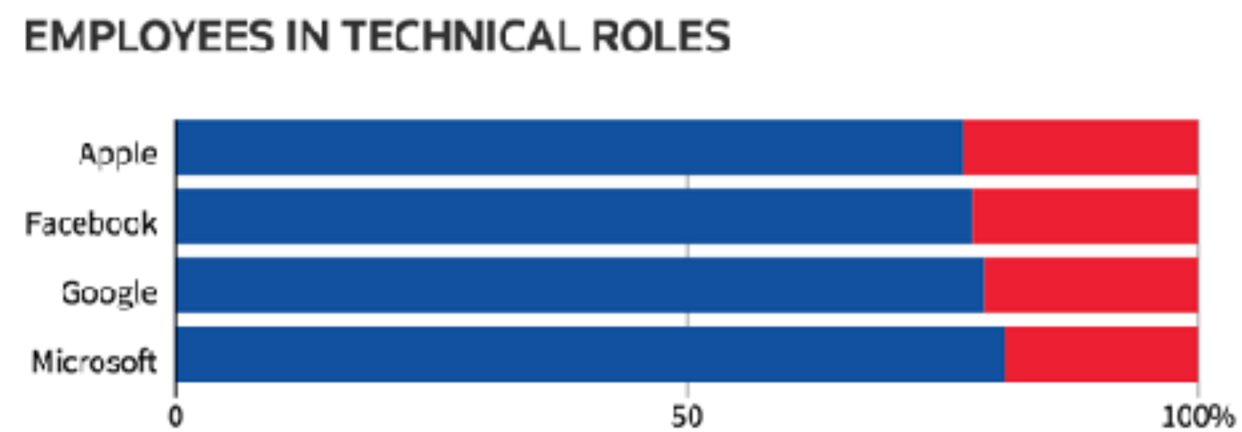
BUSINESS NEWS OCTOBER 9, 2018 / 11:12 PM / 6 MONTHS AGO

Amazon scraps secret AI recruiting tool that showed bias against women

"In effect, **Amazon's system taught itself that male candidates were preferable**. It penalized resumes that included the word "women's," as in "women's chess club captain." And it **downgraded graduates of two all-women's colleges**, according to people familiar with the matter. They did not specify the names of the schools."

October 2018



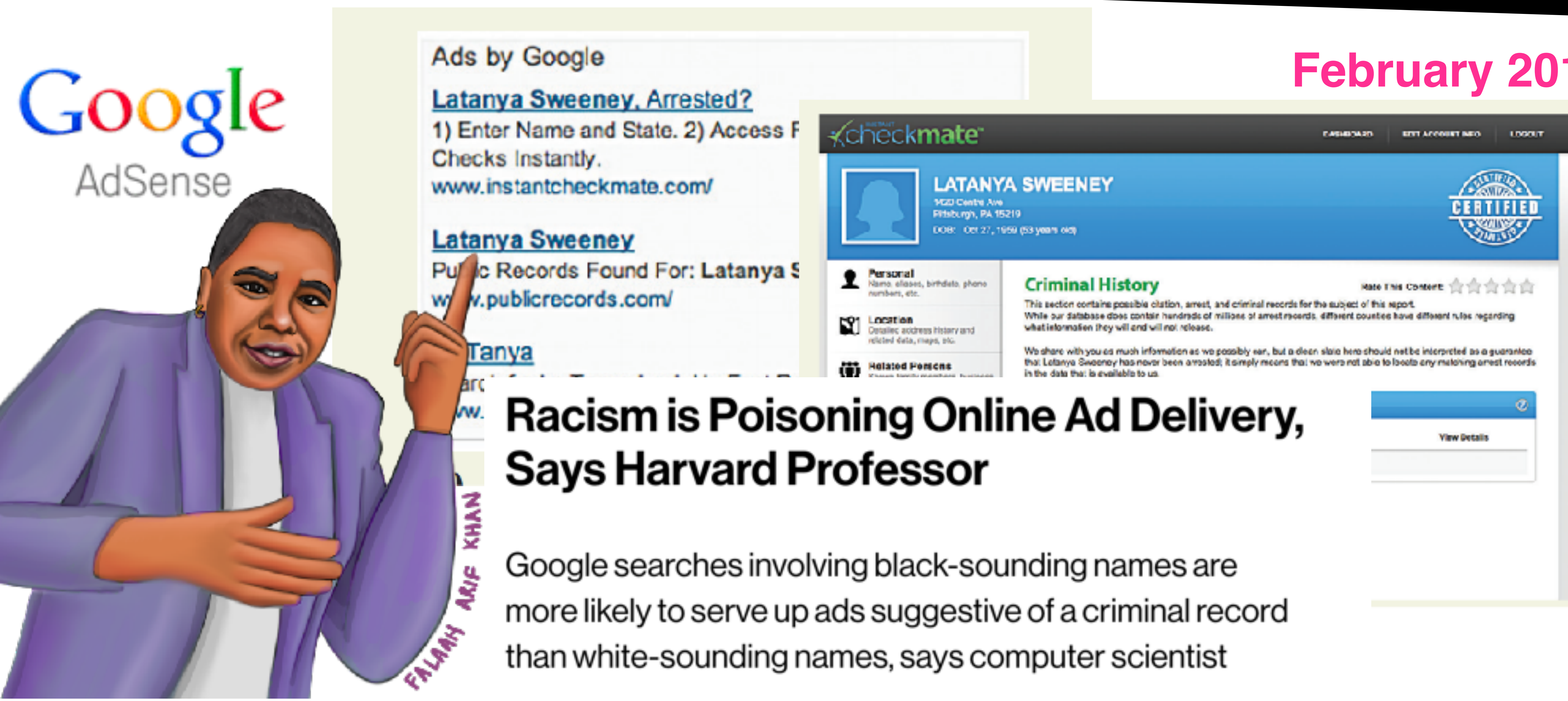


"Note: Amazon does not disclose the gender breakdown of its technical workforce."

https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

Use case: Instant Checkmate

February 2013



racially identifying names trigger ads suggestive of a criminal record

https://www.technologyreview.com/s/510646/racism-is-poisoning-online-ad-delivery-says-harvard-professor/

Use case: Amazon same-day delivery

Bloomberg

Amazon Doesn't Consider the Race of Its Customers. Should It?

"… In six major same-day delivery cities, however, **the service area excludes predominantly black ZIP codes** to varying degrees, according to a Bloomberg analysis that compared Amazon same-day delivery areas with U.S. Census Bureau data."

https://www.bloomberg.com/graphics/2016-amazon-same-day/

Use case: Amazon same-day delivery

Bloomberg

Amazon Doesn't Consider the Race of Its Customers. Should It?

"The most striking gap in Amazon's same-day service is in Boston, where **three ZIP codes encompassing the primarily black neighborhood of Roxbury are excluded** from same-day service, while the neighborhoods that surround it on all sides are eligible."

https://www.bloomberg.com/graphics/2016-amazon-same-day/

examples: racial bias in risk assessment

Racial bias in criminal sentencing

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

Bernard Parker, left, was rated high risk; Dylan Fugett was rated low risk. (Josh Ritchie for ProPublica)

May 2016

A commercial tool COMPAS automatically predicts some categories of future crime to assist in bail and sentencing decisions. It is used in courts in the US.

The tool correctly predicts recidivism **61% of the time.**

Blacks are almost twice as likely as whites to be labeled a higher risk but not actually re-offend.

The tool makes **the opposite mistake among whites**: They are much more likely than blacks to be labeled lower risk but go on to commit other crimes.

https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Racial bias in criminal sentencing

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

May 2016

A commercial tool COMPAS automatically predicts some categories of future crime to assist in bail and sentencing decisions. It is used in courts in the US.

Prediction Fails Differently for Black Defendants

	WHITE	AFRICAN AMERICAN
Labeled Higher Risk, But Didn't Re-Offend	23.5%	44.9%
Labeled Lower Risk, Yet Did Re-Offend	47.7%	28.0%

Overall, Northpointe's assessment tool correctly predicts recidivism 61 percent of the time. But blacks are almost twice as likely as whites to be labeled a higher risk but not actually re-offend. It makes the opposite mistake among whites: They are much more likely than blacks to be labeled lower risk but go on to commit other crimes. (Source: ProPublica analysis of data from Broward County, Fla.)

https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Racial bias in healthcare

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

October 2019

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

Science 25 Oct 2019:
Vol. 366, Issue 6464, pp. 447-453
DOI: 10.1126/science.aax2342

Science

Health systems rely on commercial prediction algorithms to identify and help patients with complex health needs. We show that a widely used algorithm, typical of this industry-wide approach and **affecting millions of patients**, exhibits significant **racial bias**: **At a given risk score, Black patients are considerably sicker than White patients, as evidenced by signs of uncontrolled illnesses**. Remedying this disparity would increase the percentage of Black patients receiving additional help from 17.7 to 46.5%. The bias arises because the algorithm **predicts health care costs rather than illness**, but unequal access to care means that we spend less money caring for Black patients than for White patients. Thus, **despite health care cost appearing to be an effective proxy for health by some measures of predictive accuracy, large racial biases arise**. We suggest that the choice of convenient, seemingly effective proxies for ground truth can be an important source of algorithmic bias in many contexts.

r/ai

Fixing bias in algorithms?

The New York Times

By Sendhil Mullainathan

Dec. 6, 2019

ECONOMIC VIEW

Biased Algorithms Are Easier to Fix Than Biased People

Racial discrimination by algorithms or by people is harmful — but that's where the similarities end.

Tim Cook

December 2019

In one study published 15 years ago, **two people applied for a job**. Their résumés were about as similar as two résumés can be. One person was named Jamal, the other Brendan.

In a study published this year, **two patients sought medical care**. Both were grappling with diabetes and high blood pressure. One patient was black, the other was white.

Both studies documented **racial injustice**: In the first, the applicant with a black-sounding name got fewer job interviews. In the second, the black patient received worse care.

But they differed in one crucial respect. In the first, hiring managers made biased decisions. In the second, the culprit was a computer program.

https://www.nytimes.com/2019/12/06/business/algorithm-bias-fix.html

Fixing bias in algorithms?

The New York Times

By Sendhil Mullainathan

Dec. 6, 2019

ECONOMIC VIEW

Biased Algorithms Are Easier to Fix Than Biased People

Racial discrimination by algorithms or by people is harmful — but that's where the similarities end.

Tim Cook

December 2019

Changing algorithms is easier than changing people: software on computers can be updated; the "wetware" in our brains has so far proven much less pliable.

[…] In a 2018 paper […], I took a cautiously optimistic perspective and argued that **with proper regulation, algorithms can help to reduce discrimination**.

But the key phrase here is "proper regulation," which we do not currently have.

We must ensure all the necessary inputs to the algorithm, including the data used to test and create it, are carefully stored. * […] **We will need a well-funded regulatory agency with highly trained auditors to process this data.**

https://www.nytimes.com/2019/12/06/business/algorithm-bias-fix.html a push for regulation

Automated Decision Systems (ADS)

Automated Decision Systems (ADS)

process data about people

help make consequential decisions

combine human & automated decision making

aim to improve **efficiency** and promote **equity**

are subject to **auditing** and **public disclosure**

Regulating ADS?

New York City Local Law 49 of 2018

January 11, 2018

An **Automated Decision System (ADS)** is a “computerized implementation of algorithms, including those derived from machine learning or other data processing or artificial intelligence techniques, which are used to make or assist in making decisions.”

Form task force that surveys the current use of ADS in City agencies and develops procedures for:

- requesting and receiving an **explanation** of an algorithmic decision affecting an individual (3(b))
- interrogating ADS for **bias and discrimination** against members of legally-protected groups (3(c) and 3(d))
- allowing the **public** to **assess** how ADS function and are used (3(e)), and archiving ADS together with the data they use (3(f))

ADS regulation in NYC: take 1

Principles

- using ADS **where** they promote innovation and efficiency in service delivery
- promoting **fairness**, **equity**, **accountability**, and **transparency** in the use of ADS
- reducing potential harm **across the lifespan** of ADS

New York City Local Law 144 of 2021

THE NEW YORK CITY COUNCIL

Corey Johnson, Speaker

December 11, 2021

This bill would require that a **bias audit** be conducted on an automated employment decision tool prior to the use of said tool. The bill would also require that candidates or employees that reside in the city **be notified about the use of such tools** in the assessment or evaluation for hire or promotion, as well as, **be notified about the job qualifications and characteristics that will be used** by the automated employment decision tool. Violations of the provisions of the bill would be subject to a civil penalty.

course overview

module 1: algorithmic fairness

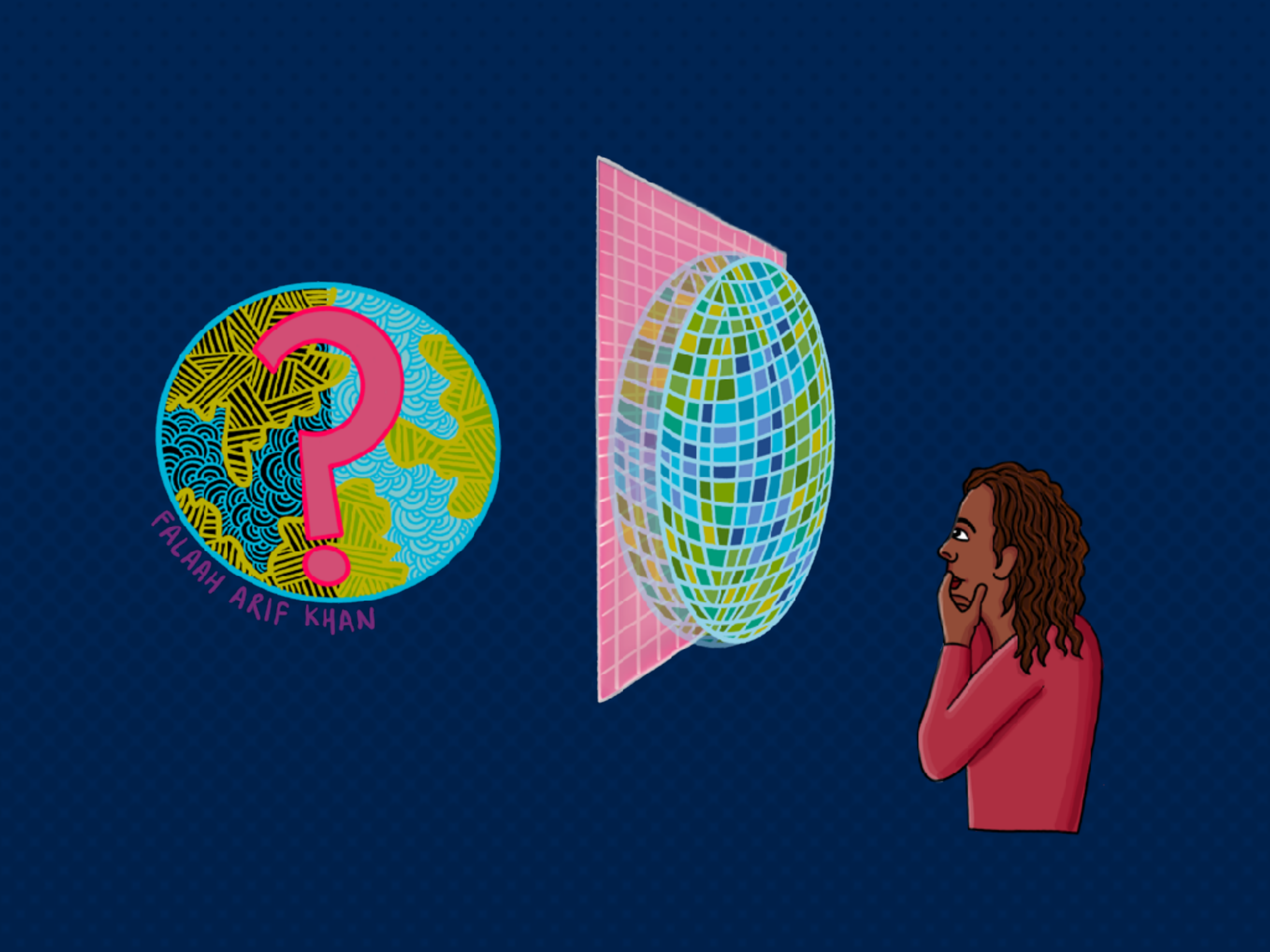
Bias in computer systems

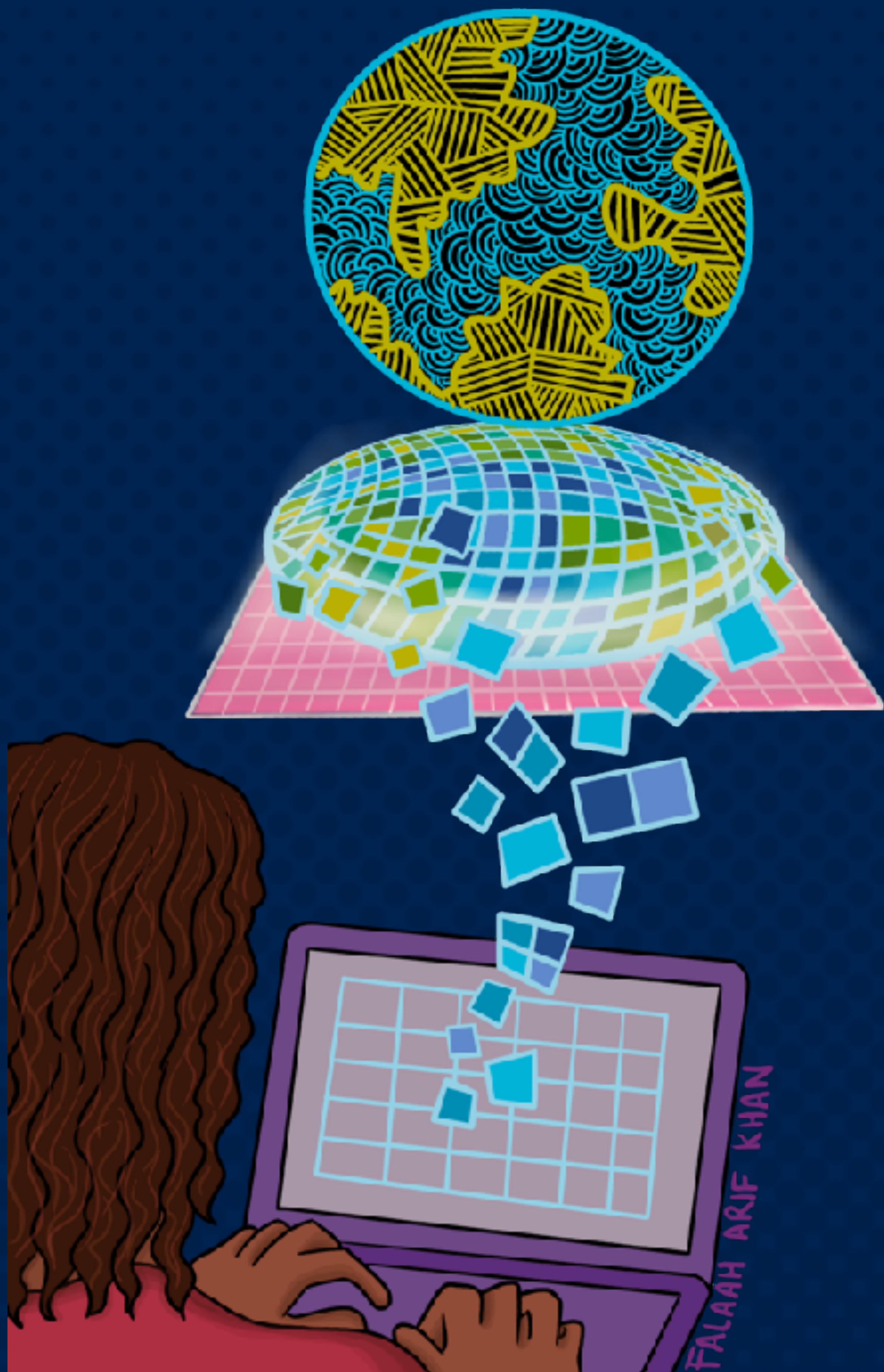
Pre-existing: exists independently of algorithm, has origins in society

Technical: introduced or exacerbated by the technical properties of an ADS

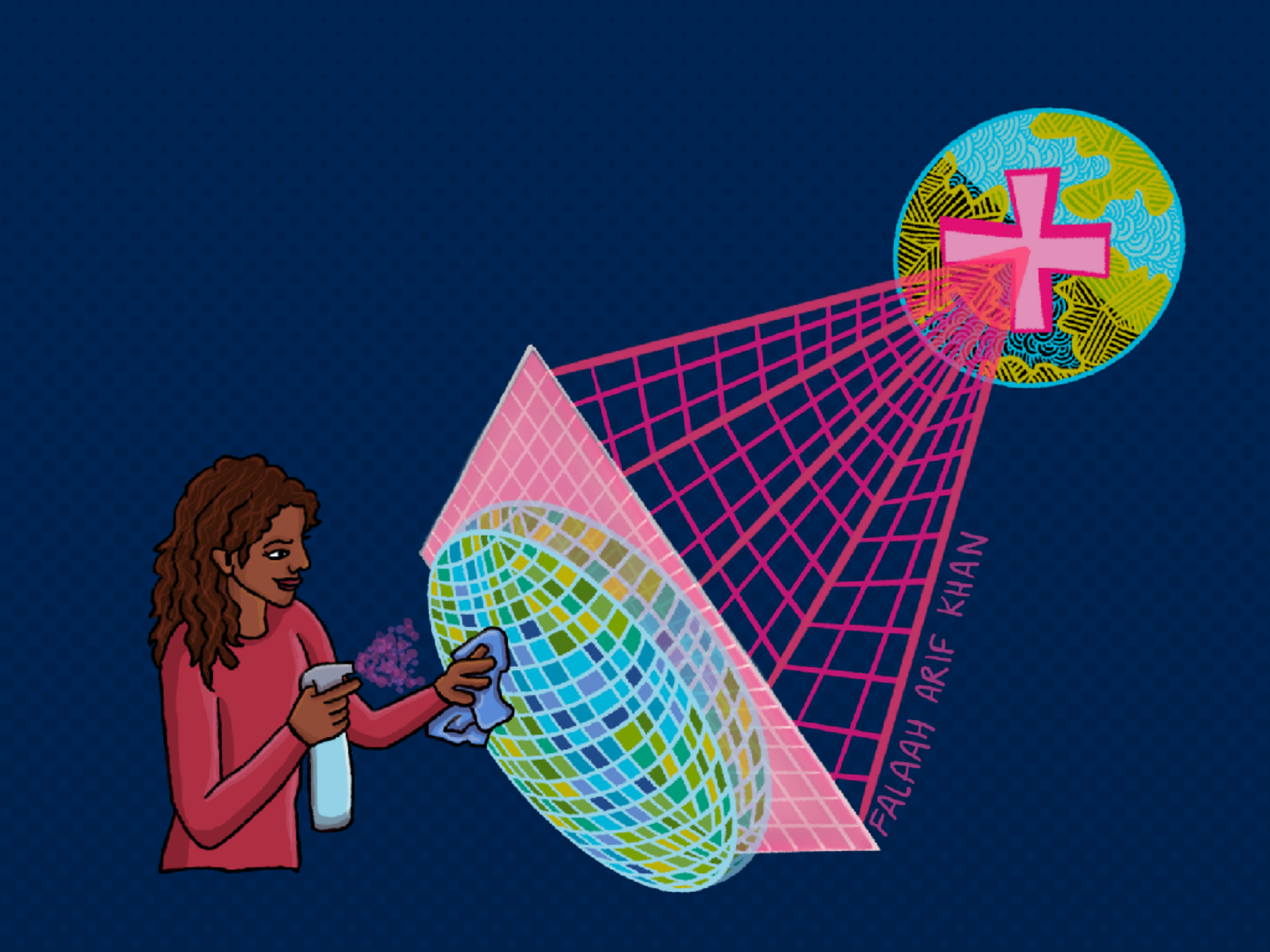
Emergent: arises due to context of use

[Friedman & Nissenbaum (1996)]


FALAAH ARIF KHAN


FALAAH ARIF KHAN


FALAAH ARIF KHAN


FALAAH ARIF KHAN

module 2:
the data science
lifecycle

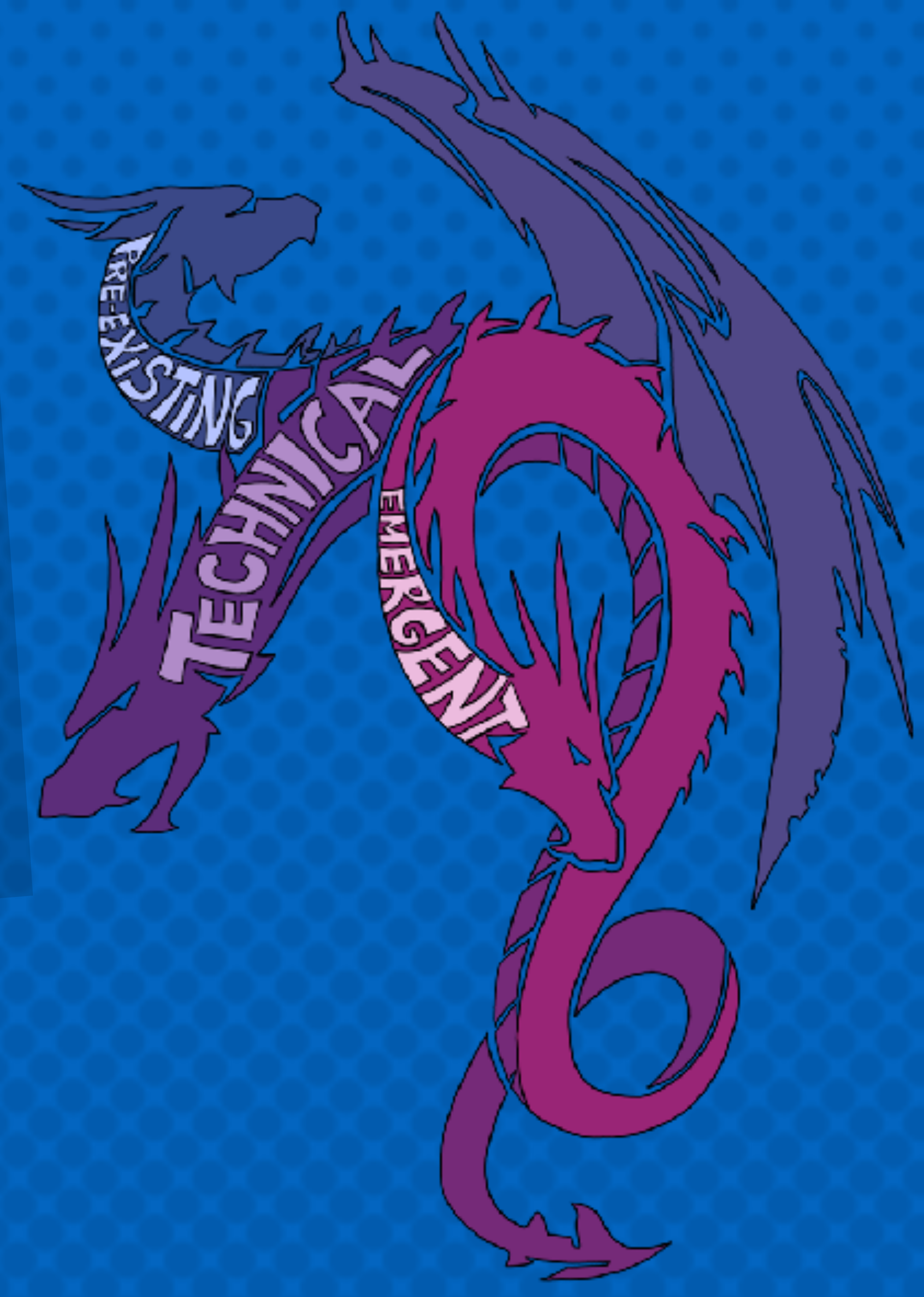
Bias in computer systems

Pre-existing: exists independently of algorithm, has origins in society

Technical: introduced or exacerbated by the technical properties of an ADS

Emergent: arises due to context of use

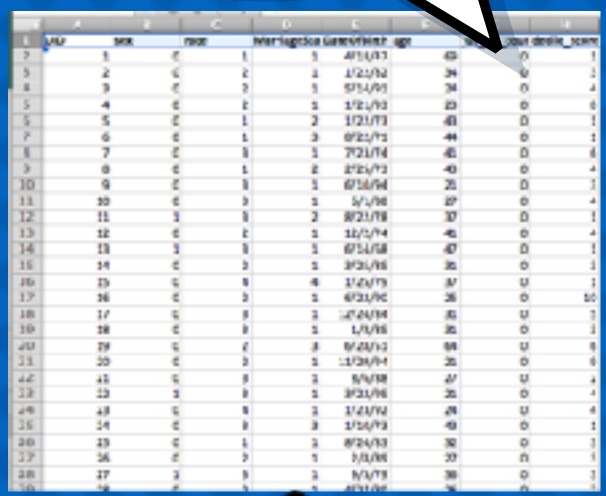
to fight bias, state beliefs and assumptions explicitly



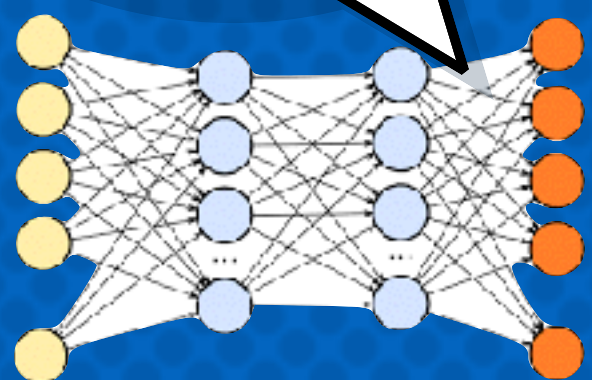
[Friedman & Nissenbaum (1996)]

Fair-ML view

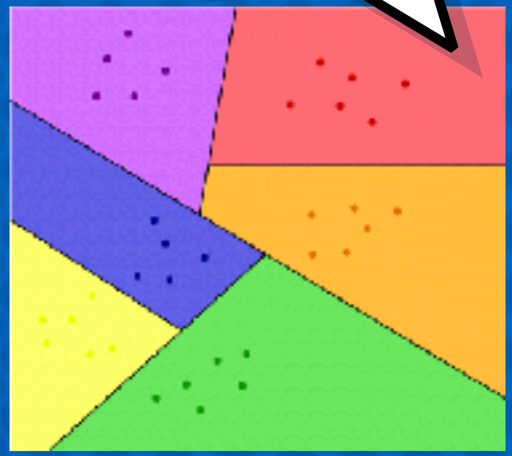
where did the data come from?



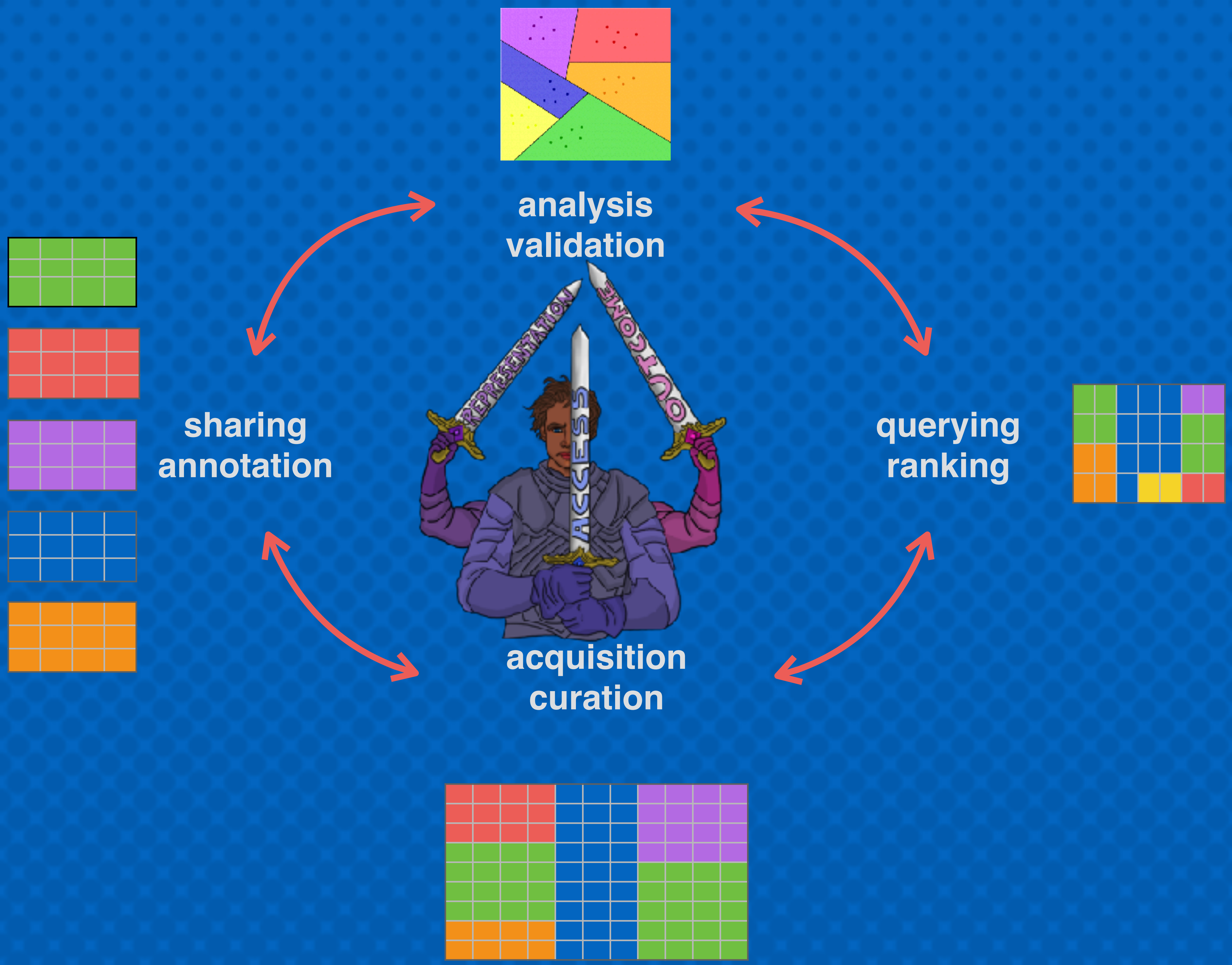
what happens inside the box?



how are results used?



Lifecycle view



[Stoyanovich, Howe, Jagadish (2020)]

Models and assumptions

[Stoyanovich, Howe, Jagadish (2020)]

module 3:
data protection & privacy

Privacy: two sides of the same coin

Did you go out drinking over the weekend?

$\dfrac{\text{protecting an individual}}{\text{plausible deniability}}$	$\dfrac{\text{learning about the population}}{\text{noisy estimates}}$

Truth or dare

Did you go out drinking over the weekend?

let's call this property **P** (Truth=Yes) and estimate **p**, the fraction of the group for whom **P** holds

thus, we estimate **p** as:

$$\tilde{p} = 2A - \frac{1}{2}$$

1. flip a coin **C1**
    1. if **C1** is tails, then **respond truthfully**
    2. if **C1** is heads, then flip another coin **C2**
        1. if **C2** is heads then **Yes**
        2. else **C2** is tails then respond **No**

randomization - adding noise - is what gives plausible deniability

a process privacy method

the expected number of **Yes** answers is:

$$A = \frac{3}{4}p + \frac{1}{4}(1-p) = \frac{1}{4} + \frac{p}{2}$$

privacy comes from plausible deniability

Differential privacy

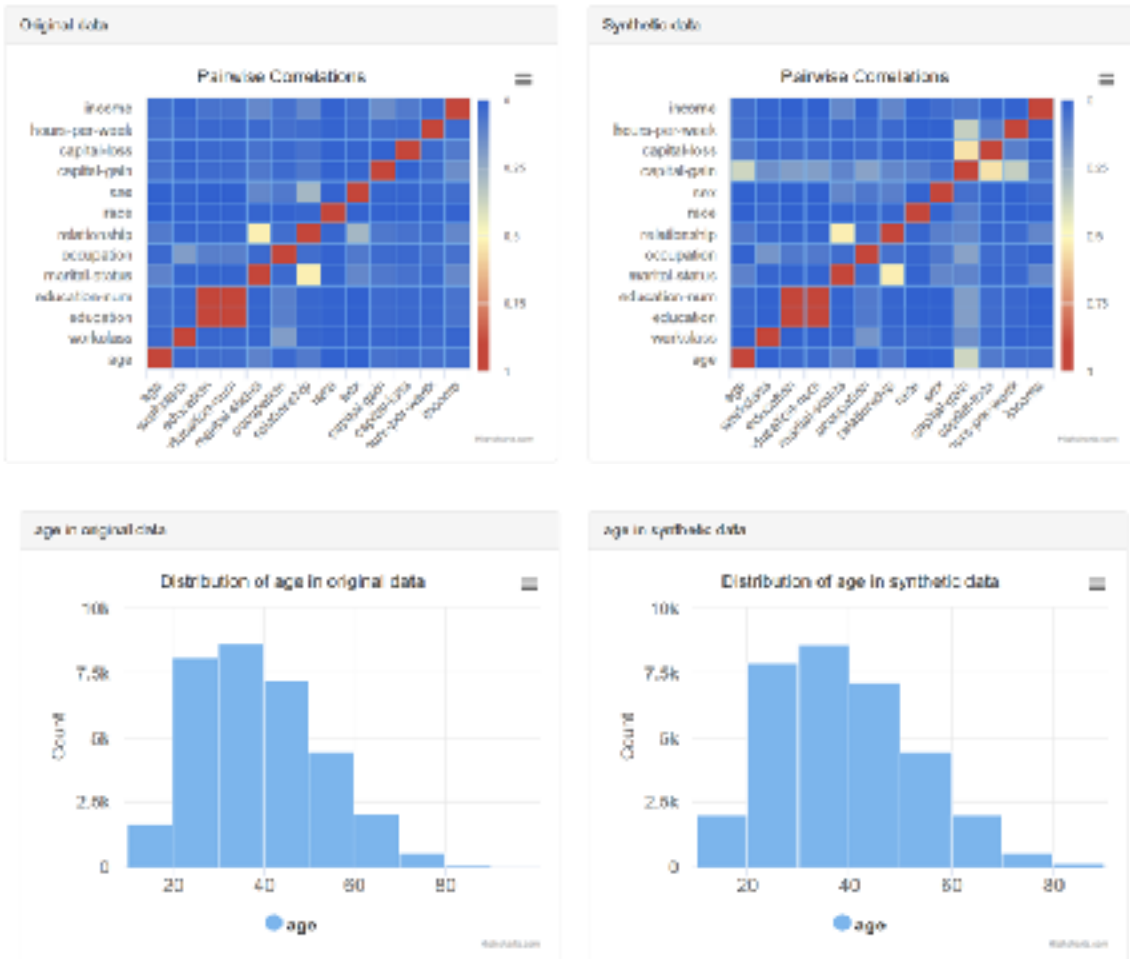
review articles

DOI:10.1145/1866739.1866758

What does it mean to preserve privacy?

BY CYNTHIA DWORK

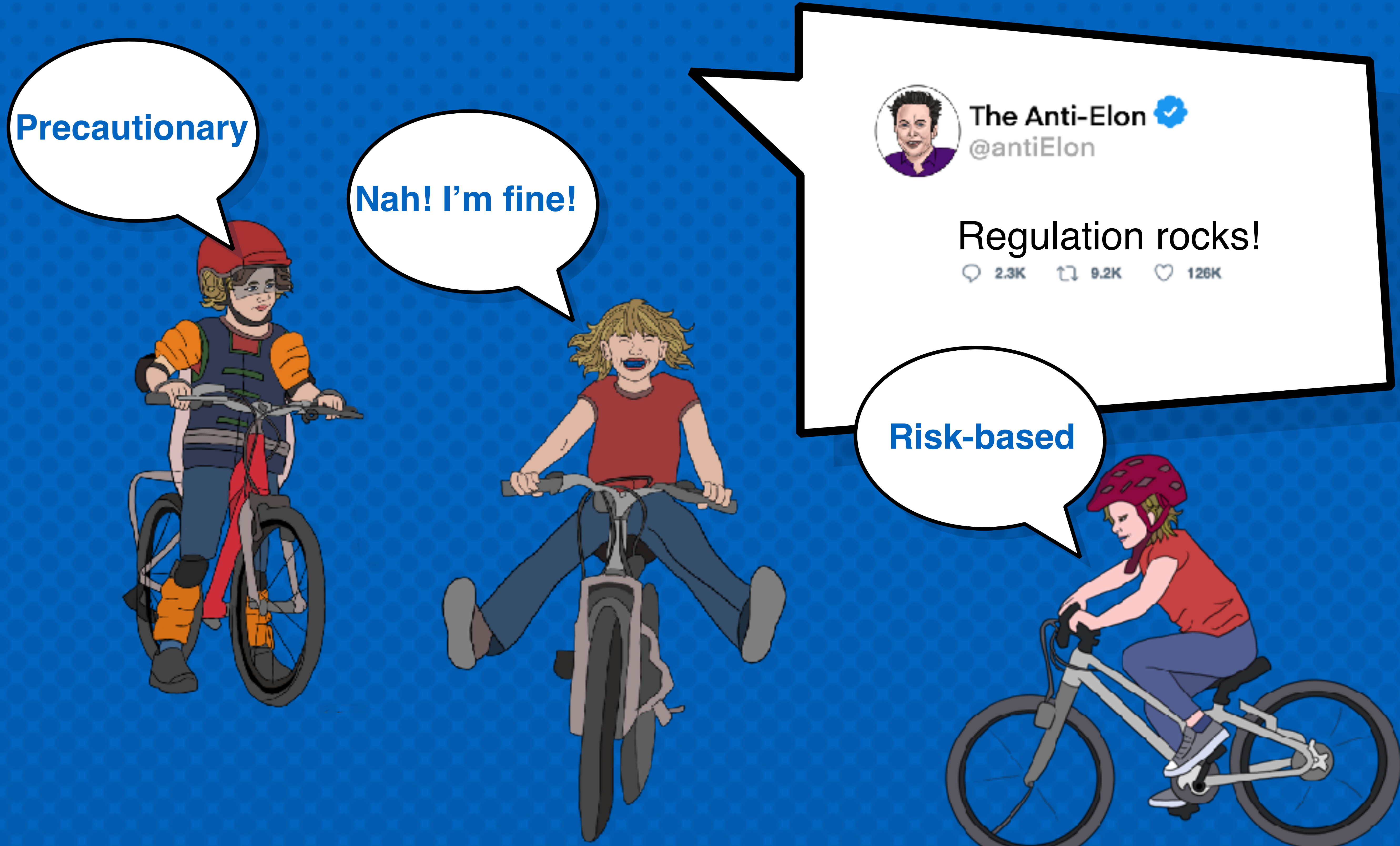
A Firm Foundation for Private Data Analysis



Communications of the ACM CACM Homepage archive

Volume 54 Issue 1, January 2011
Pages 86-95

Regulating ADS?



Legal frameworks


GENERAL DATA PROTECTION REGULATION (GDPR) | RECITALS | KEY ISSUES | Deutsch

GDPR

- Chapter 1 (Art. 1 – 4) General provisions
- Chapter 2 (Art. 5 – 11) Principles
- Chapter 3 (Art. 12 – 23) Rights of the data subject
- Chapter 4 (Art. 24 – 43) Controller and processor
- Chapter 5 (Art. 44 – 50) Transfers of personal data to third countries or international organisations
- Chapter 6 (Art. 51 – 59) Independent supervisory authorities
- Chapter 7 (Art. 60 – 76) Cooperation and consistency
- Chapter 8 (Art. 77 – 84) Remedies, liability and penalties
- Chapter 9 (Art. 85 – 91) Provisions relating to specific processing situations
- Chapter 10 (Art. 92 – 93) Delegated acts and implementing acts
- Chapter 11 (Art. 94 – 99) Final provisions

General Data Protection Regulation
GDPR

Welcome to gdpr-info.eu. Here you can find the official PDF of the Regulation (EU) 2016/679 (General Data Protection Regulation) in the current version of the OJ L 119, 04.05.2016; cor. OJ L 127, 23.5.2018 as a neatly arranged website. All Articles of the GDPR are linked with suitable recitals. The European Data Protection Regulation is applicable as of May 25th, 2018 in all member states to harmonize data privacy laws across Europe. If you find the page useful, feel free to support us by sharing the project.

Quick Access

Chapter 1 – 1 2 3 4
Chapter 2 – 5 6 7 8 9 10 11
Chapter 3 – 12 13 1
Chapter 4 – 24 25 2
Chapter 5 – 44 45 4
Chapter 6 – 51 52 5
Chapter 7 – 60 61 6
Chapter 8 – 77 78 7
Chapter 9 – 85 86 8


Government of Canada | Gouvernement du Canada

Home → How government works → Policies, directives, standards and guidelines

Directive on Automated Decision-Making

The Government of Canada is increasingly looking to utilize artificial intelligence to make, or assist in making, administrative decisions to improve service delivery. The Government is committed to doing so in a manner that is compatible with core administrative law principles such as transparency, accountability, legality, and procedural fairness. Understanding that this technology is changing rapidly, this Directive will continue to evolve to ensure that it remains relevant.

Date modified: 2019-02-05

module 4:
transparency & interpretability

The evils of discrimination

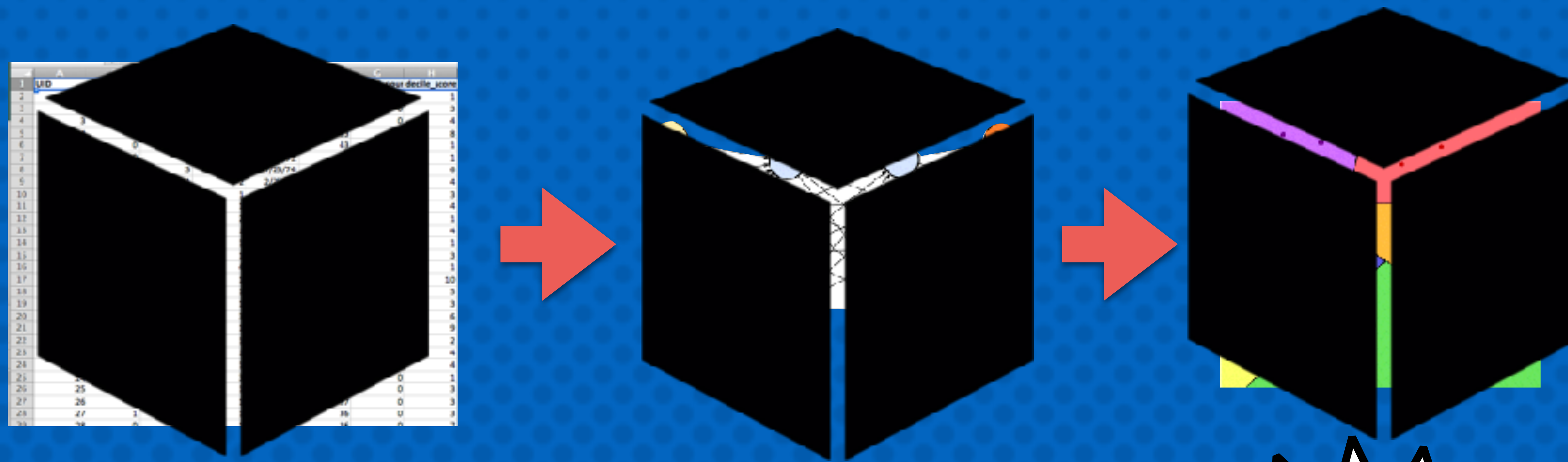
Disparate treatment

is the illegal practice of treating an entity, such as a job applicant or an employee, differently based on a **protected characteristic** such as race, gender, age, religion, sexual orientation, or national origin.

Disparate impact

is the result of systematic disparate treatment, where disproportionate **adverse impact** is observed on members of a **protected class**.

Regulating automated decisions

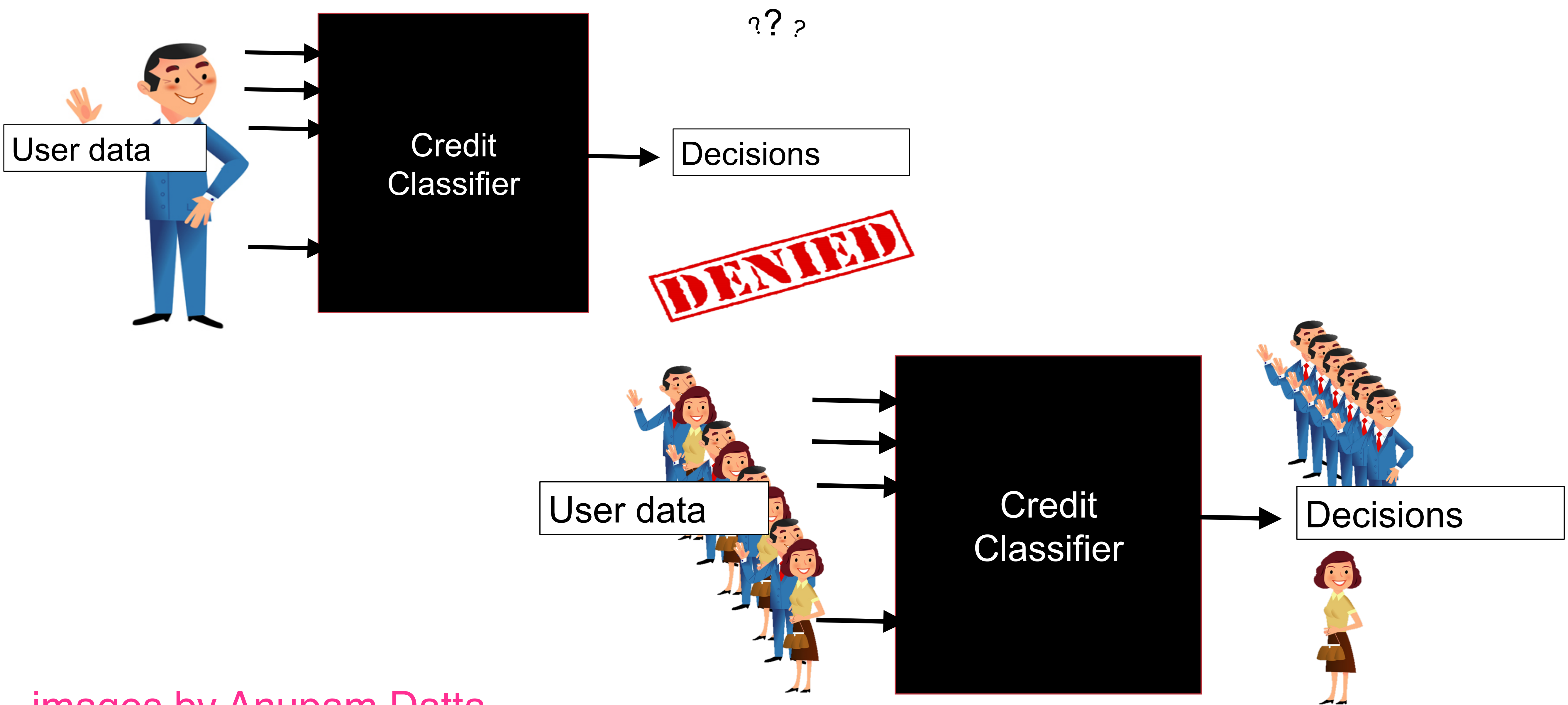


Fair Housing Act

Equal Credit Opportunity Act, 1964

Civil Rights Act, 1964

Auditing black-box models



images by Anupam Datta

Nutritional labels

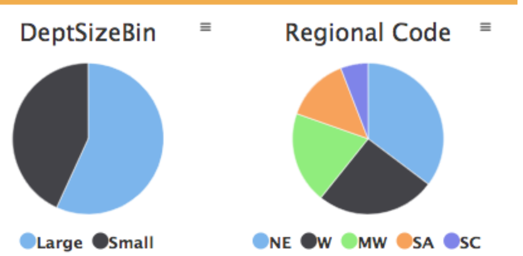
Ranking Facts

Ingredients

Attribute	Importance
PubCount	1.0
CSRankingAllArea	0.24
Faculty	0.12

Importance of an attribute in a ranking is quantified by the correlation coefficient between attribute values and items scores, computed by a linear regression model. Importance is high if the absolute value of the correlation coefficient is over 0.75, medium if this value falls between 0.25 and 0.75, and low otherwise.

Diversity overall



Fairness

DeptSizeBin	FA*IR	Pairwise	Proportion
Large	Fair	Fair	Fair
Small	Unfair	Unfair	Unfair

A ranking is considered unfair when the p-value of the corresponding statistical test falls below 0.05.

Stability

Top-K	Stability
Top-10	Stable
Overall	Stable

comprehensible: short, simple, clear

consultative: provide actionable info

comparable: implying a standard


in summary

So what is RDS?

As advertised: ethics, legal compliance, personal responsibility.
But also: **data quality**!

A technical course, with content drawn from:

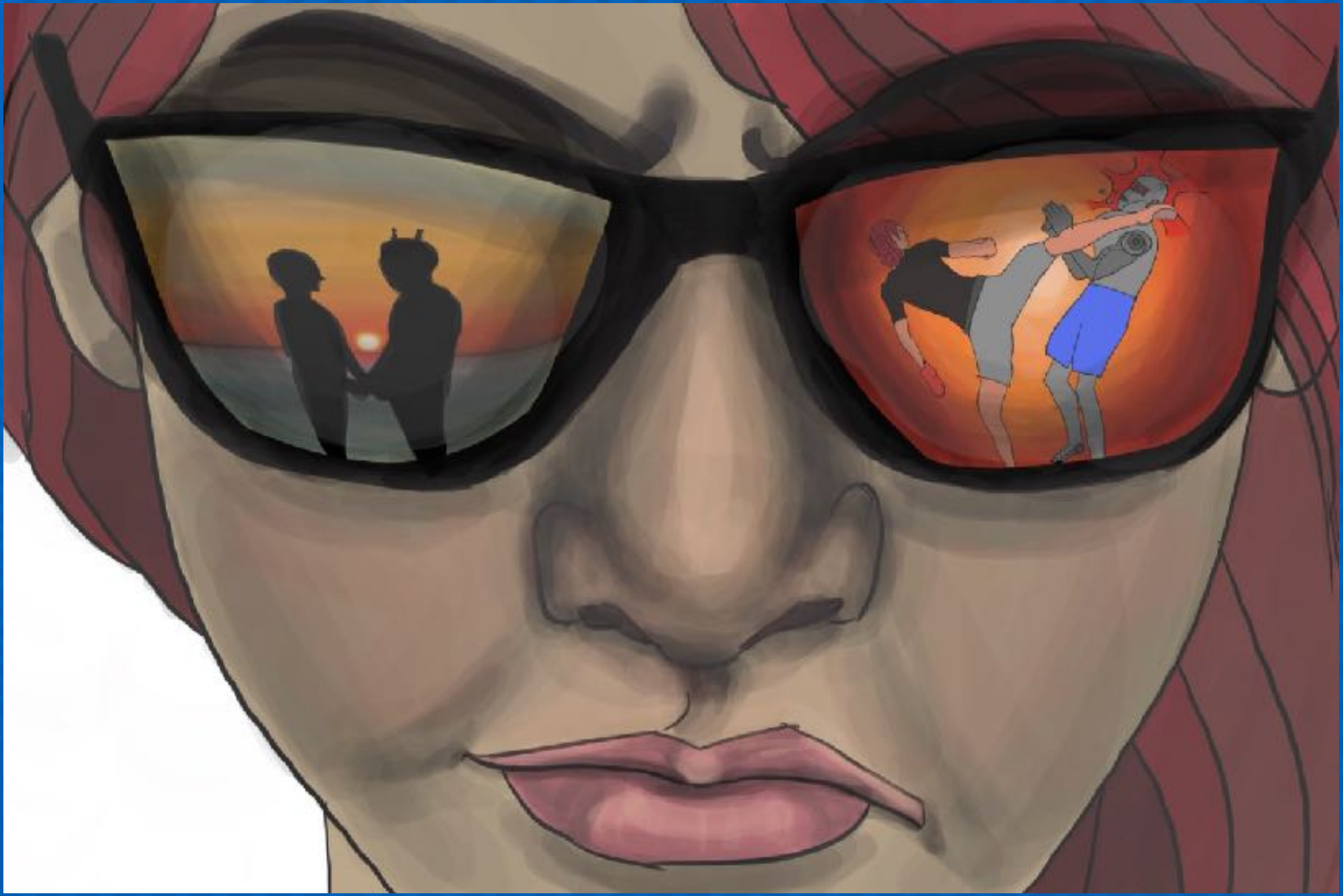
1. fairness, accountability and transparency
2. data engineering
3. privacy & data protection

We will learn **algorithmic techniques** for data analysis.
We will also learn about recent **laws** / **regulatory frameworks**.

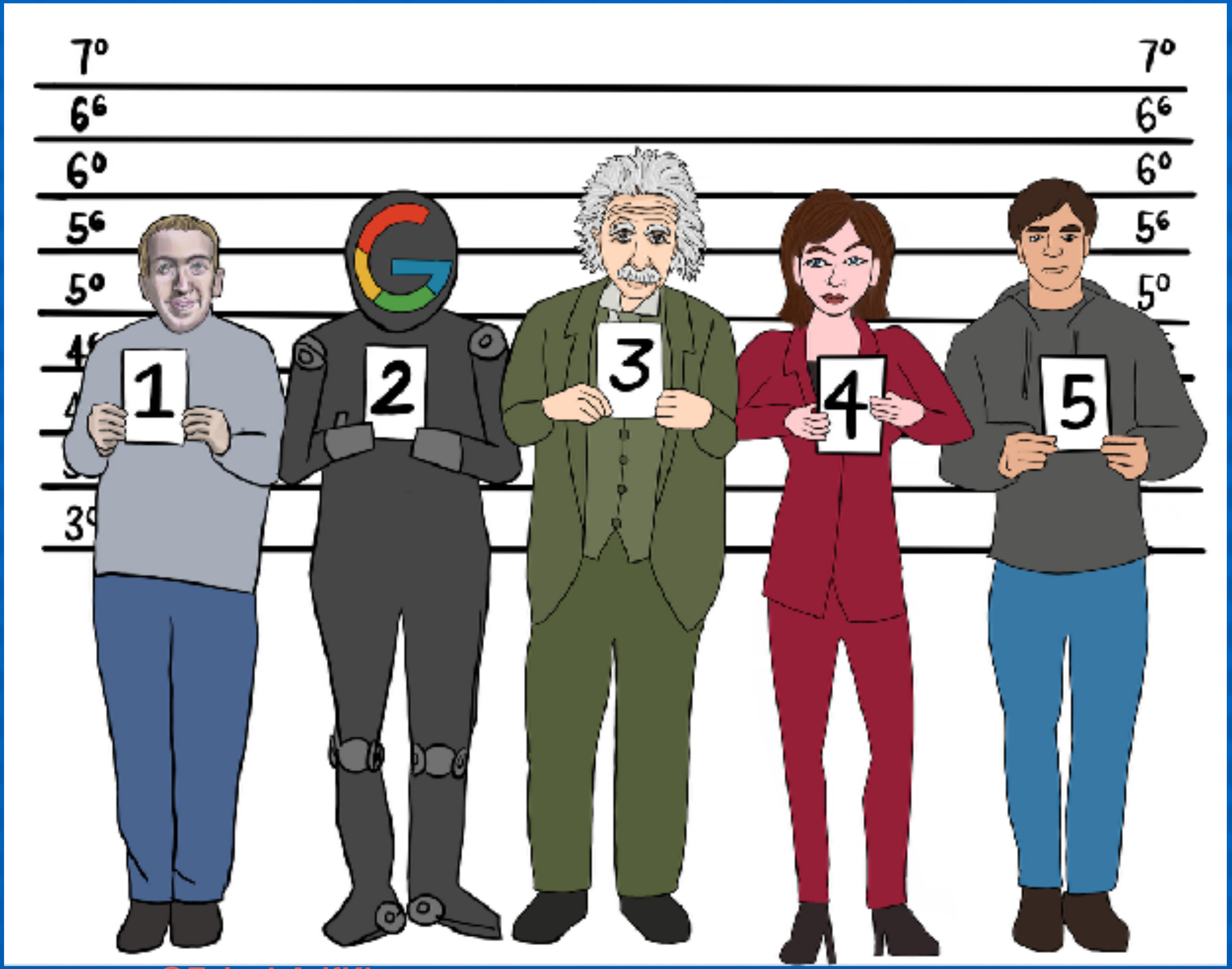
Bottom line: we will learn that many of the problems are **socio-technical**, and so cannot be "solved" with technology alone.

My perspective: a pragmatic engineer, **not** a technology skeptic.

Nuance, please!



We all are responsible



@FalaahArifKhan

Responsible Data Science

Introduction and Overview

Thank you!