# Responsible Data Science

## Introduction and Overview

*January 23, 2024*

**Prof. Umang Bhatt**

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: Umang Bhatt

Assistant Professor and Faculty Fellow
Center for Data Science
New York University

Ph.D. in Engineering from University of Cambridge
M.S. and B.S. in ECE from Carnegie Mellon University

**Research**: Trustworthy Machine Learning

- Algorithmic Transparency (Explainable AI and Uncertainty Quantification)
- Human-Machine Collaboration
- Decision Support Systems

**And also**:

- NGOs: Center for Democracy and Technology, OECD, Mozilla Foundation, Partnership on AI, Responsible AI Institute
- Outreach: Deep Learning Indaba, The Alan Turing Institute

**Office hours:** Tuesdays 5-6 ET and by appointment

http://r-ai.co

# Teaching Assistants

Andrew Bell

Raphael Meyer

Aradhita Bhandari

Venetia Pliatsika

Marcia Ma

# Assignments and grading

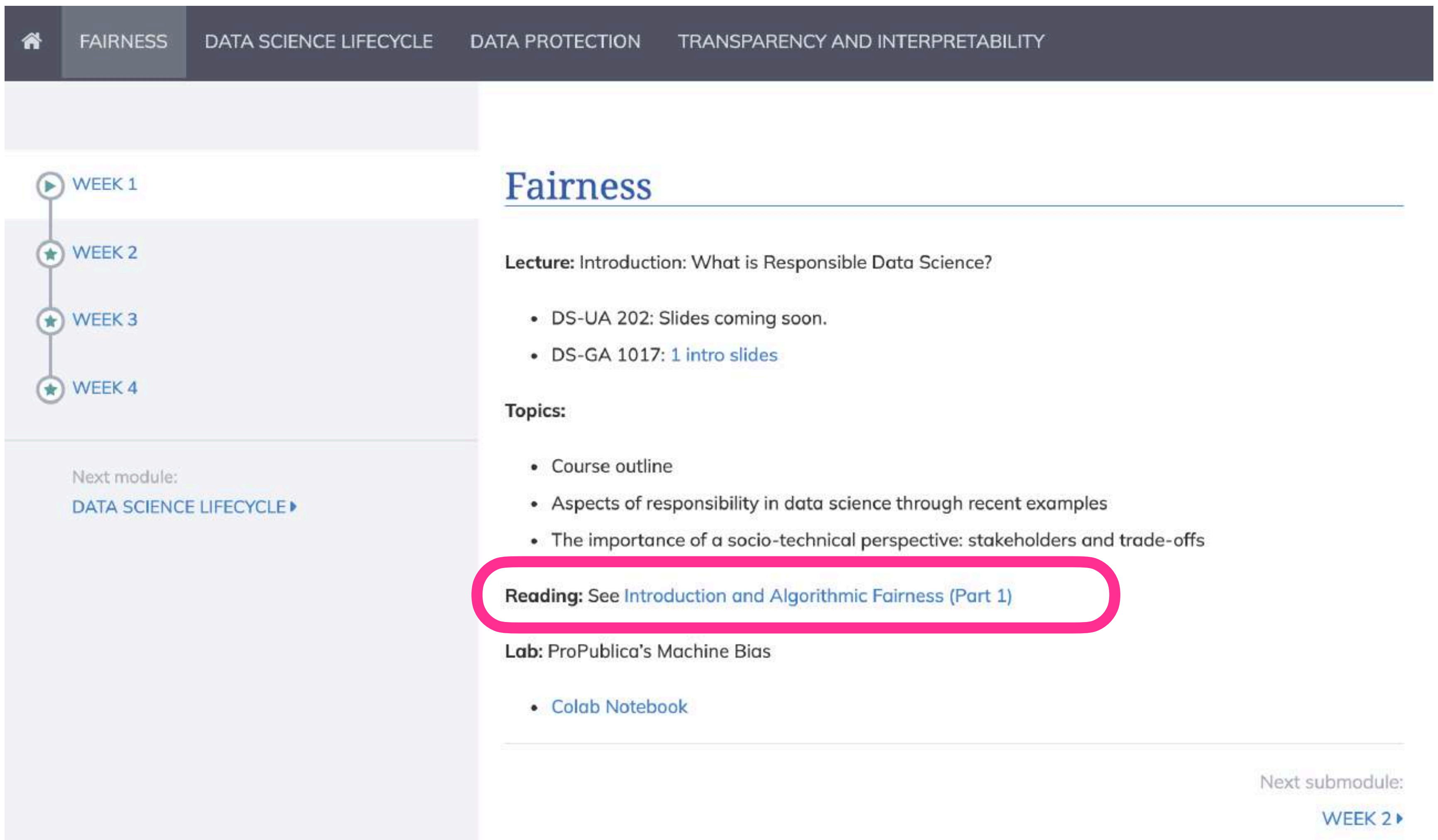
**Grading:** homeworks - 10% x 3 = 30%
project - 30%
final exam - 20%
labs - 10%
quizzes - 10%

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.

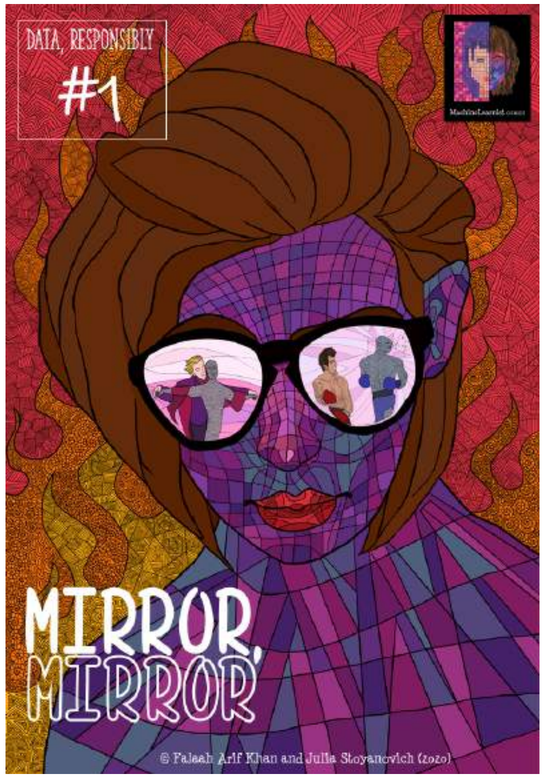
# Where to find information

**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



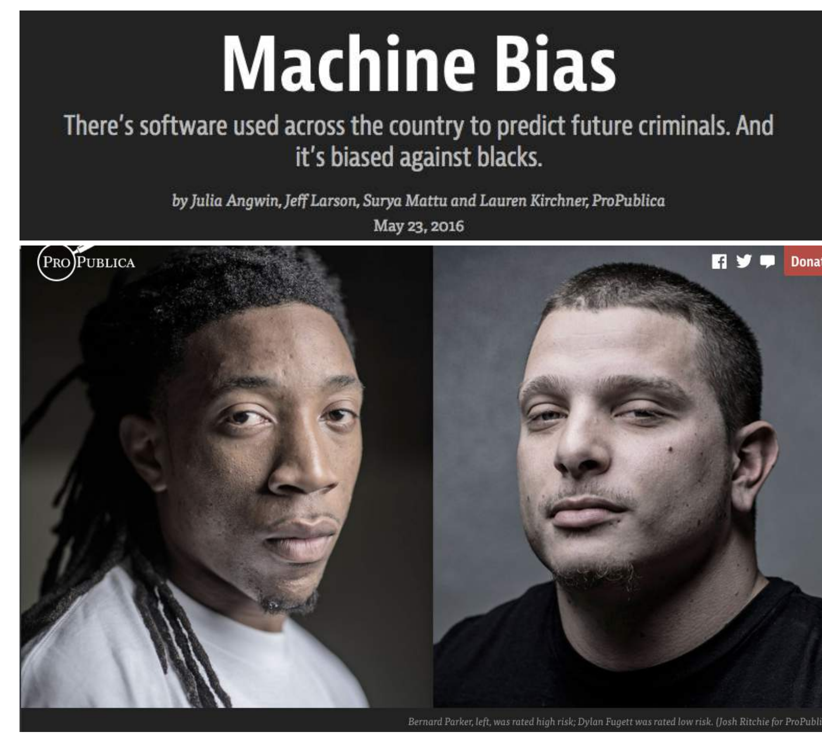


DOI:10.1145/3376898

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



Machine Bias

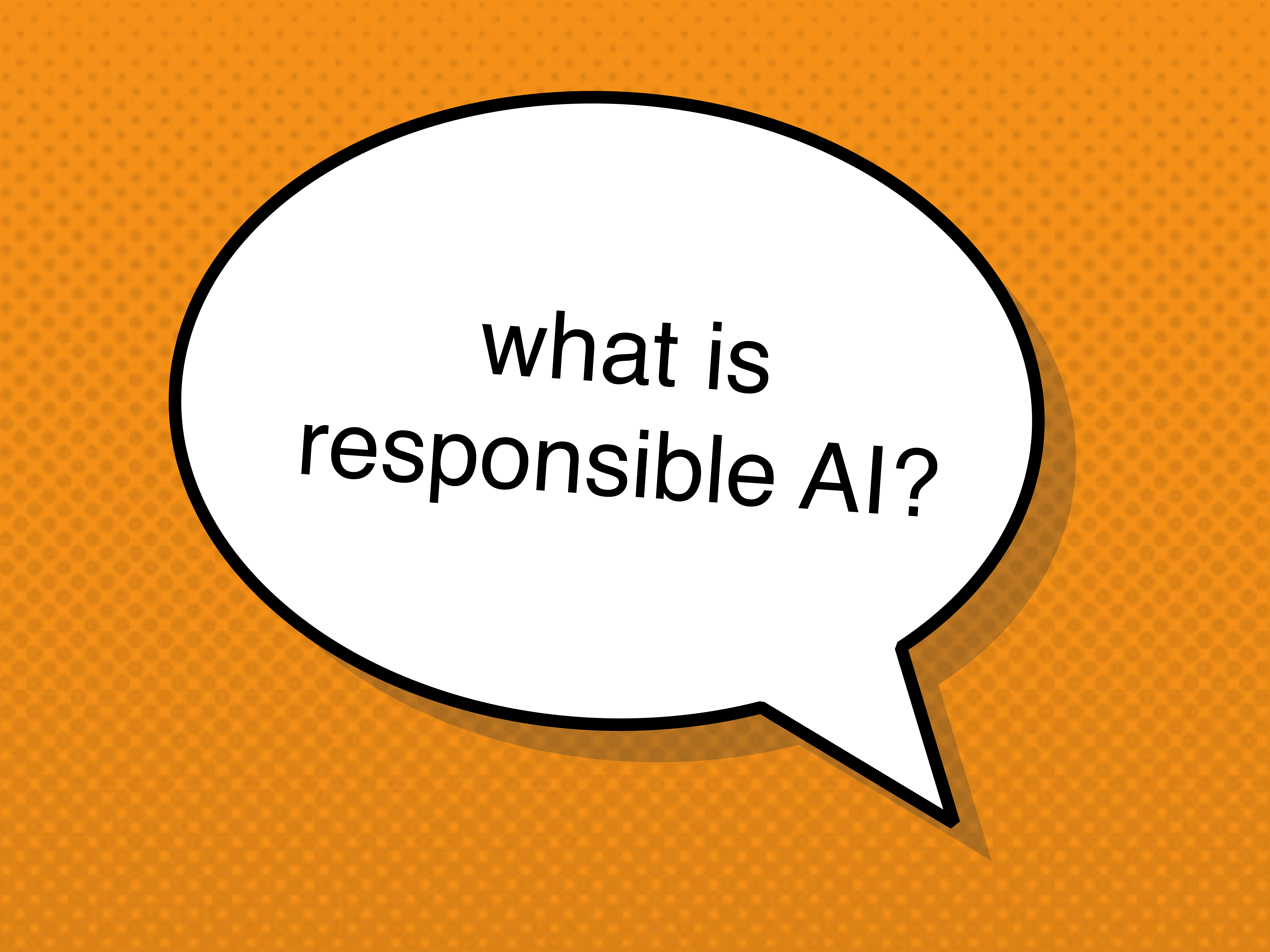
There's software used across the country to predict future criminals. And it's biased against blacks.

by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica

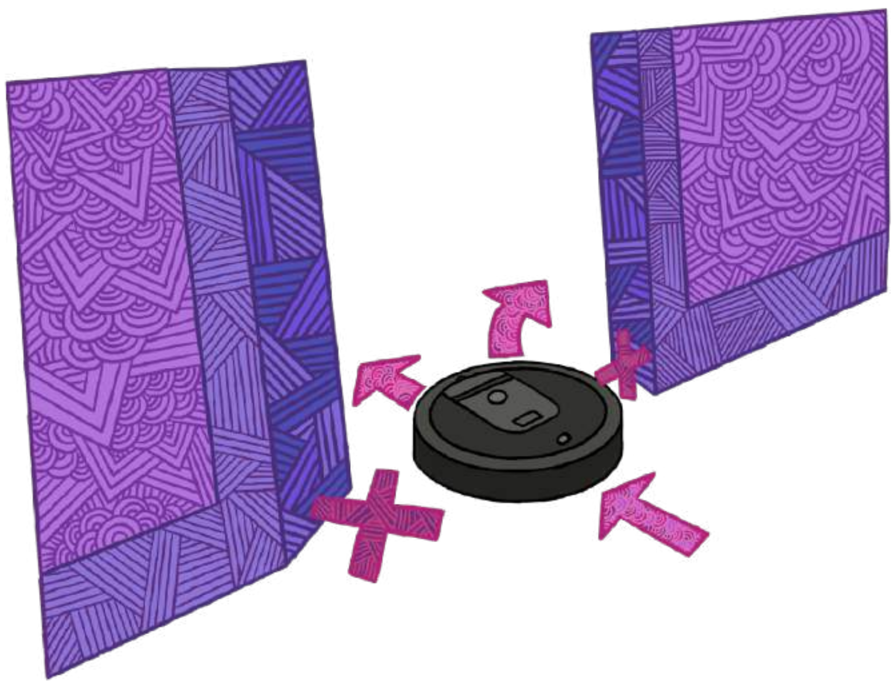
May 23, 2016



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

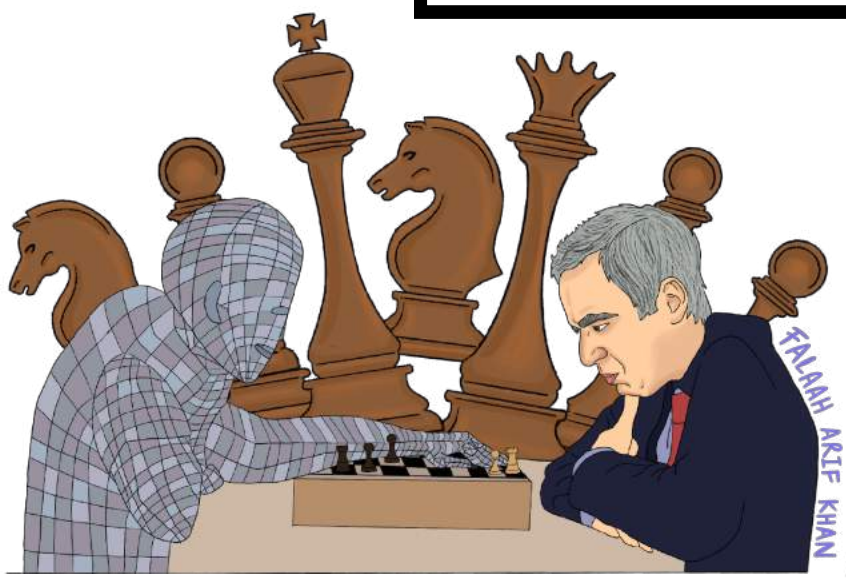

what is
responsible AI?

# 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**

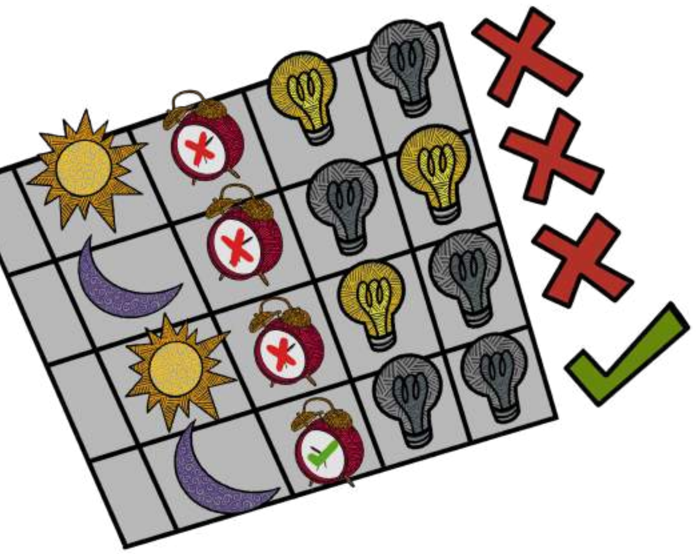
make our lives convenient

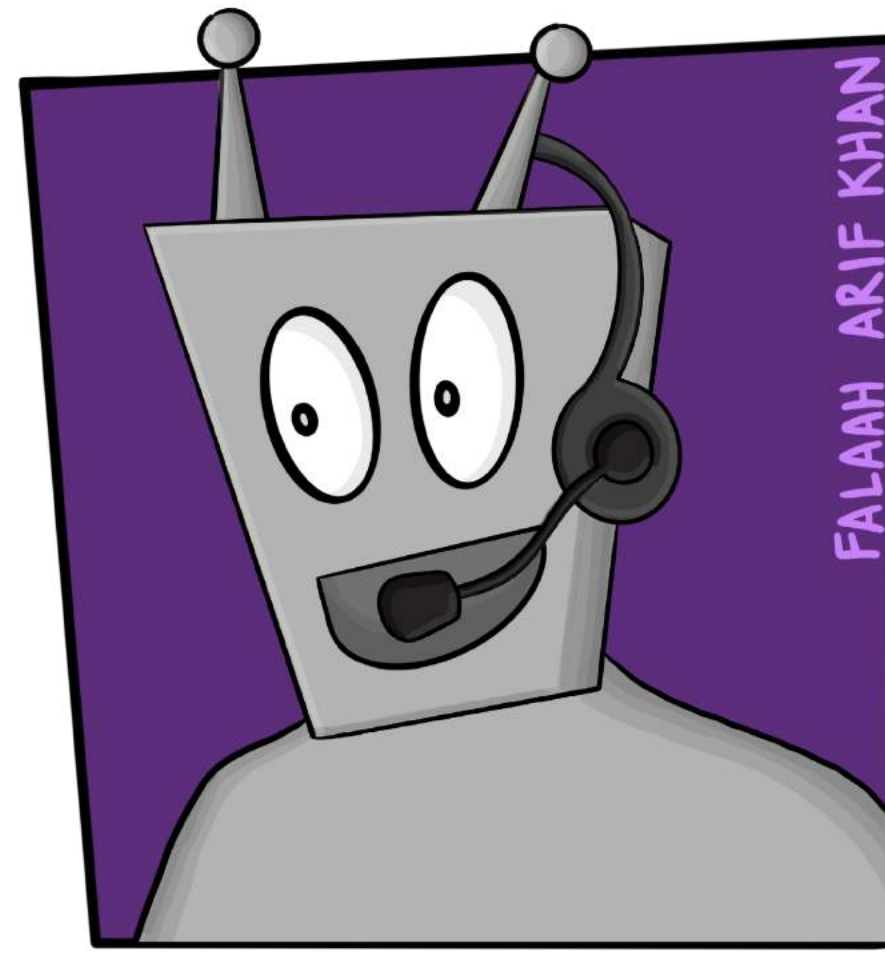
accelerate science

boost innovation

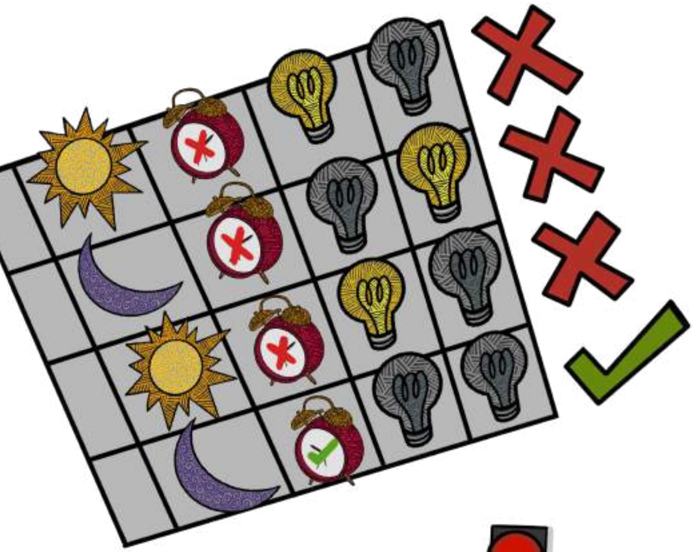
transform government

# Machines make mistakes

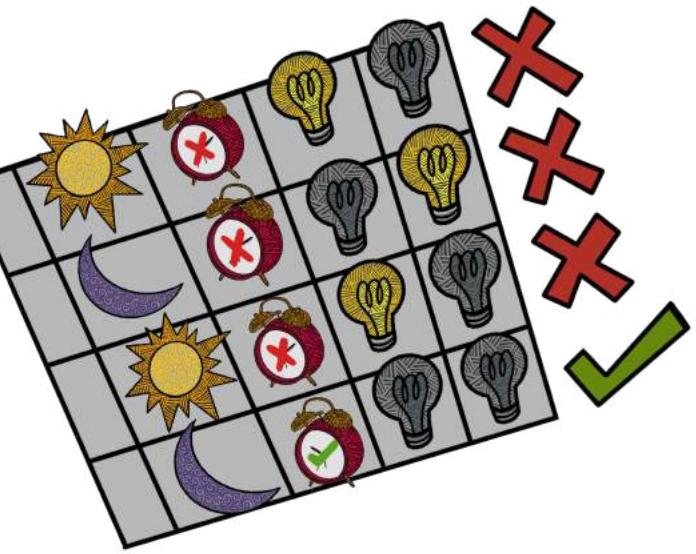


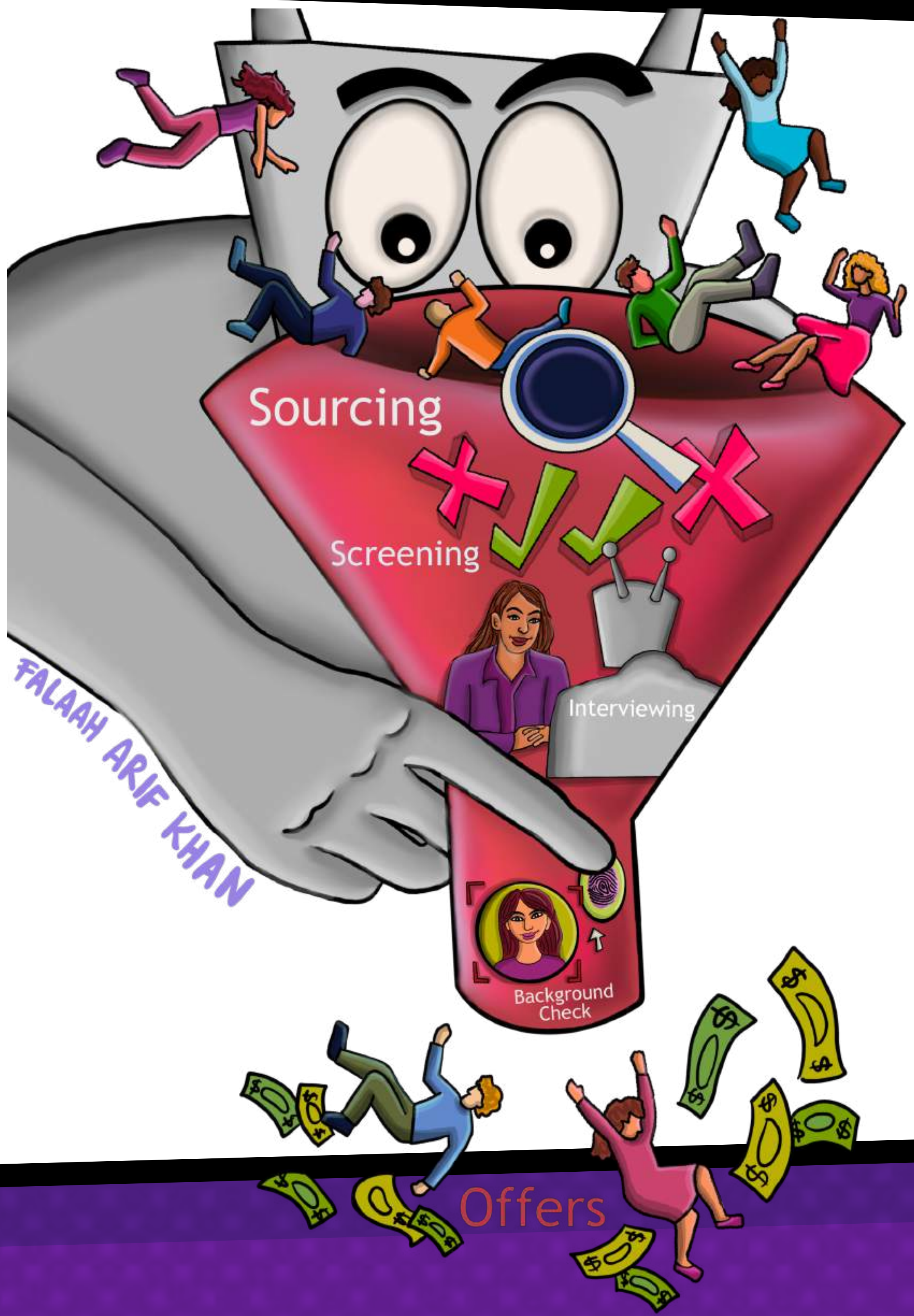



# Mistakes lead to harms



# Harms can be cumulative





what is responsible AI?

# more examples

# Medical imaging

FACEBOOK AI NYU Langone Health

fastMRI

Accelerating MR Imag

## What is fastMRI?

fastMRI is a collaborative re
between Facebook AI Resea
NYU Langone Health. The aim is to
the use of AI to make MRI scans up to 10 times faster.

By producing accurate images from under-sampled data, AI image reconstruction has the potential to improve the patient's experience and to make MRIs accessible for more people.

to

zed

raw data and image data ub
repository, which contains baseline reconstruction models and PyTorch data loaders for the fastMRI dataset.

**Positive factors**

clear need for improvement

can validate predictions

technical readiness

decision-maker readiness

https://fastmri.org/

# 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**

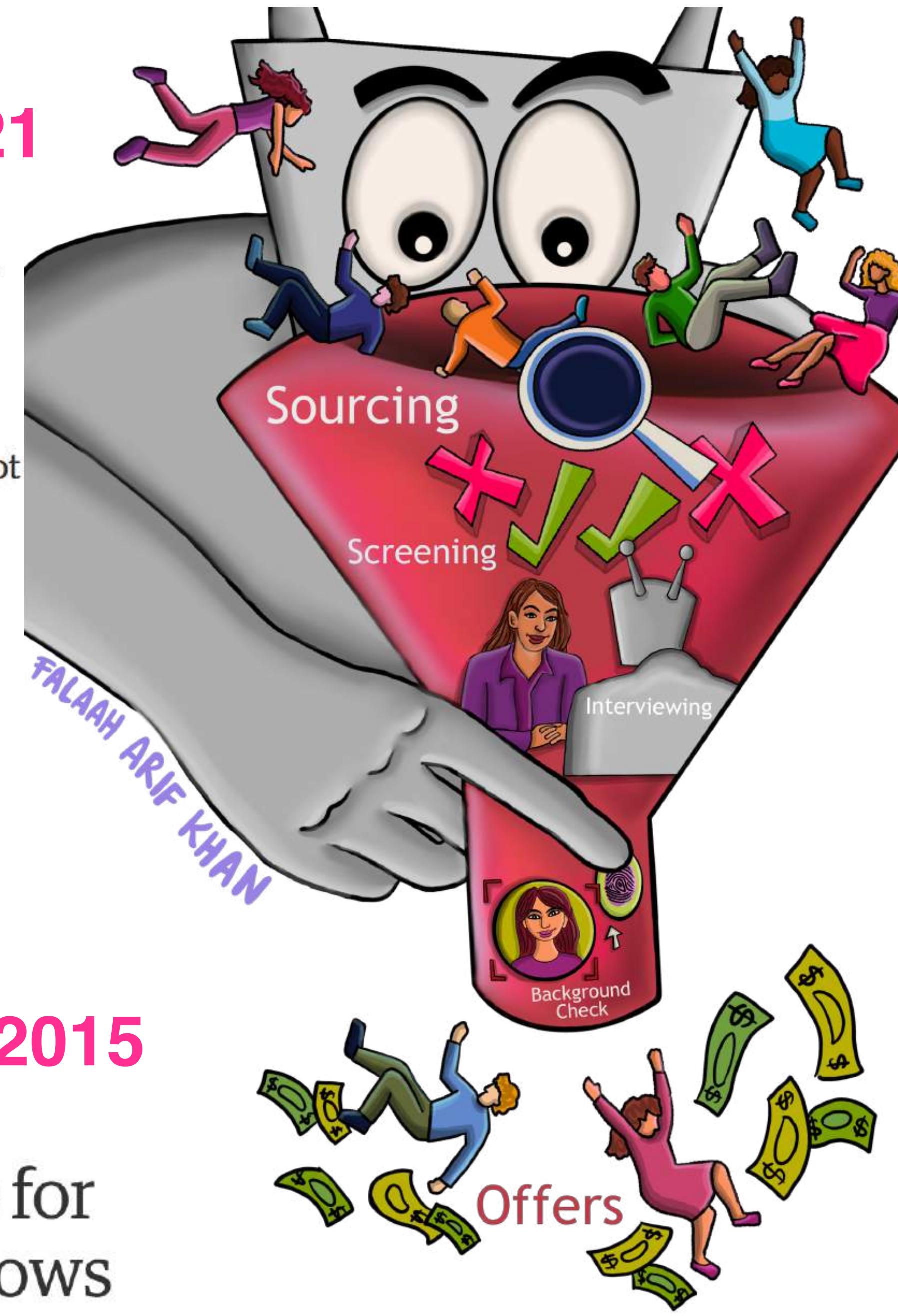
Artificial intelligence used to evaluate job candidates must not become a tool that exacerbates discrimination.

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

# Use case: AdFisher

theguardian

**July 2015**

Samuel Gibbs

Wednesday 8 July 2015 11.29 BST

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

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

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

# Use case: Resume screening

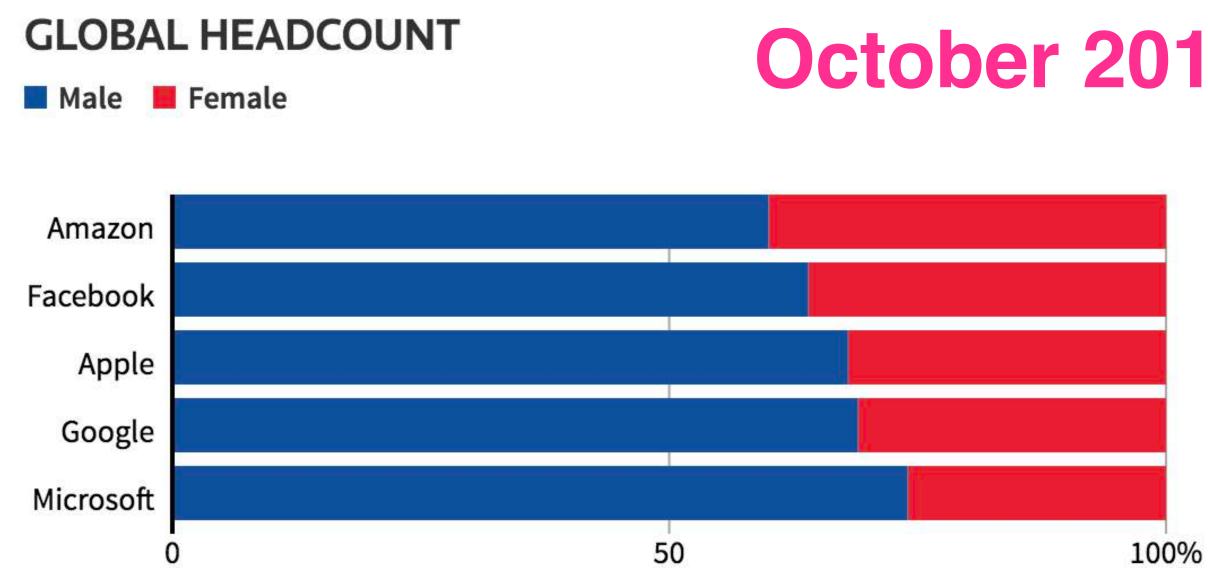
REUTERS Jeffrey Dastin

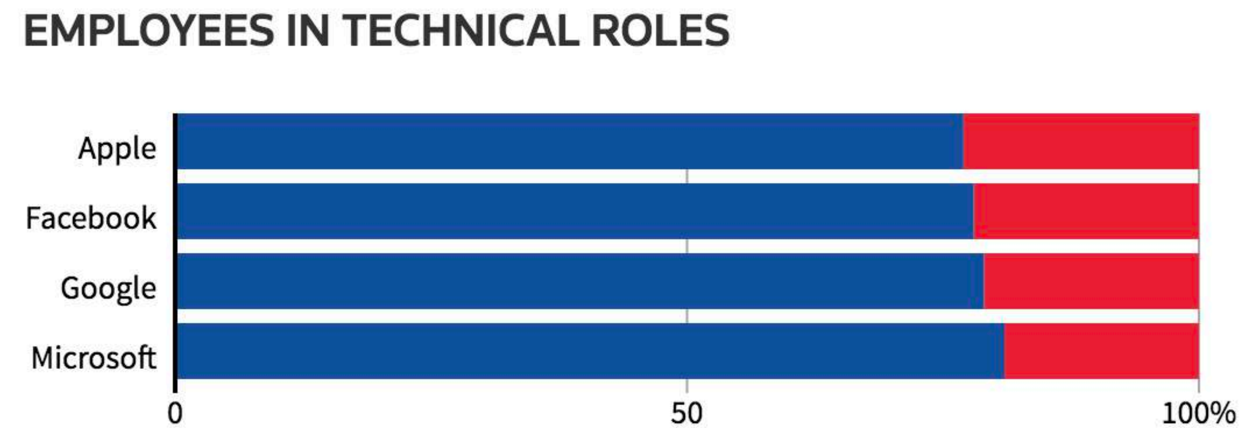
BUSINESS NEWS OCTOBER 9, 2018 / 11:12 PM / 6 MONTHS AGO

## 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**



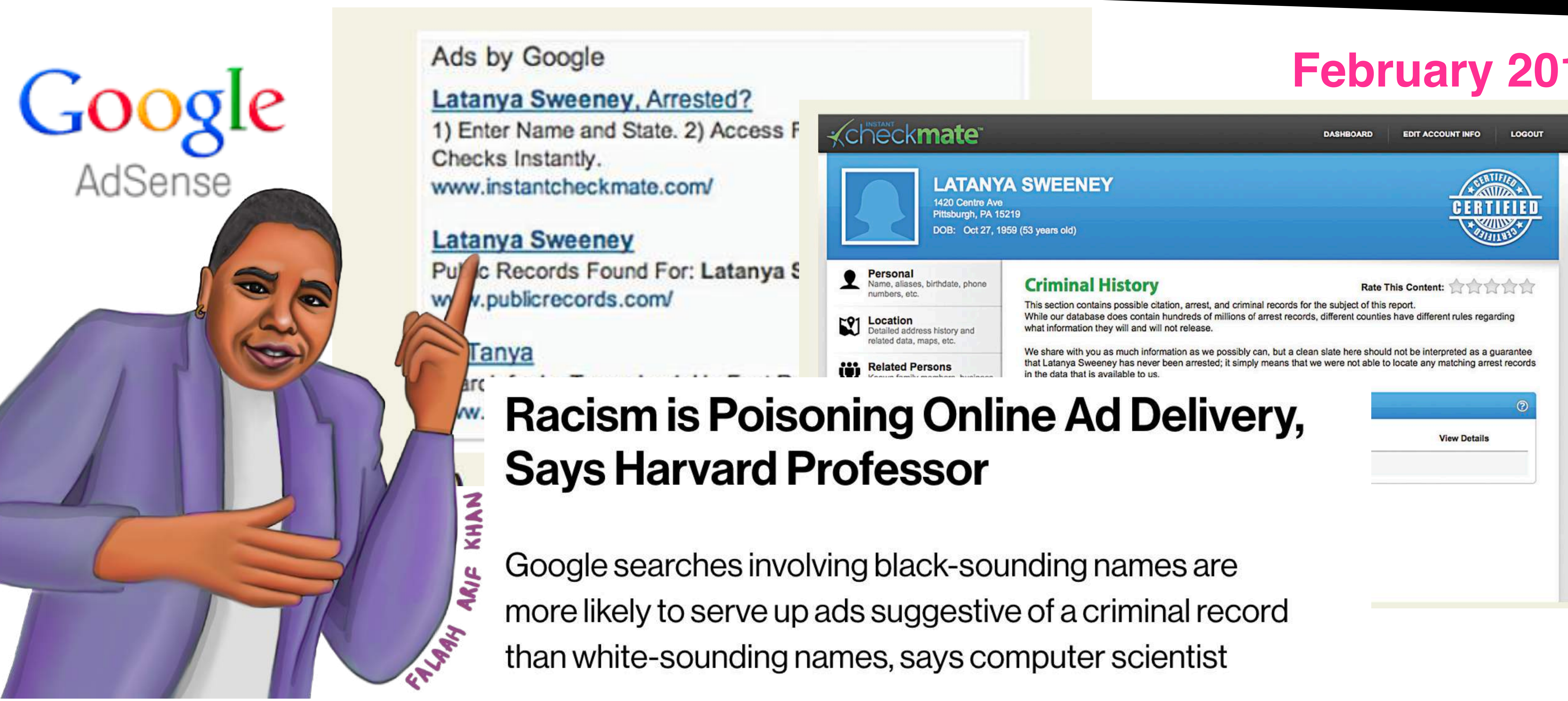




"Note: Amazon does not disclose the gender breakdown of its technical workforce."

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/

r/ai

# 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.

https://www.science.org/doi/10.1126/science.aax2342

r/ai

# Racial bias in healthcare

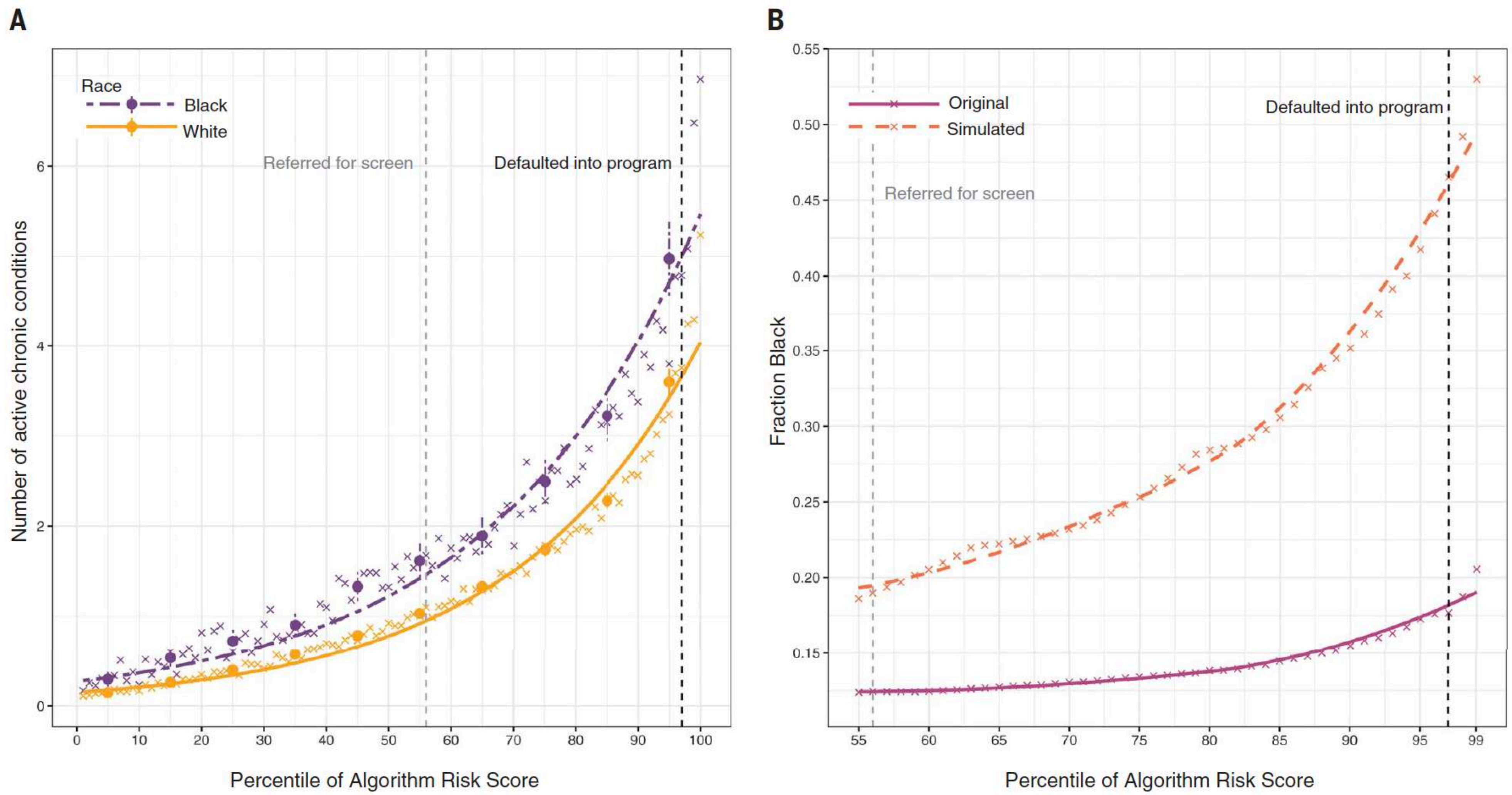



**Fig. 1. Number of chronic illnesses versus algorithm-predicted risk, by race.** (**A**) Mean number of chronic conditions by race, plotted against algorithm risk score. (**B**) Fraction of Black patients at or above a given risk score for the original algorithm ("original") and for a simulated scenario that removes algorithmic bias ("simulated": at each threshold of risk, defined at a given percentile on the *x* axis, healthier Whites above the threshold are replaced with less healthy Blacks below the threshold, until the marginal patient is equally healthy). The × symbols show risk percentiles by race; circles show risk deciles with 95% confidence intervals clustered by patient. The dashed vertical lines show the auto-identification threshold (the black line, which denotes the 97th percentile) and the screening threshold (the gray line, which denotes the 55th percentile).

https://www.science.org/doi/10.1126/science.aax2342

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

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

**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.**

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))

r/ai

# 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

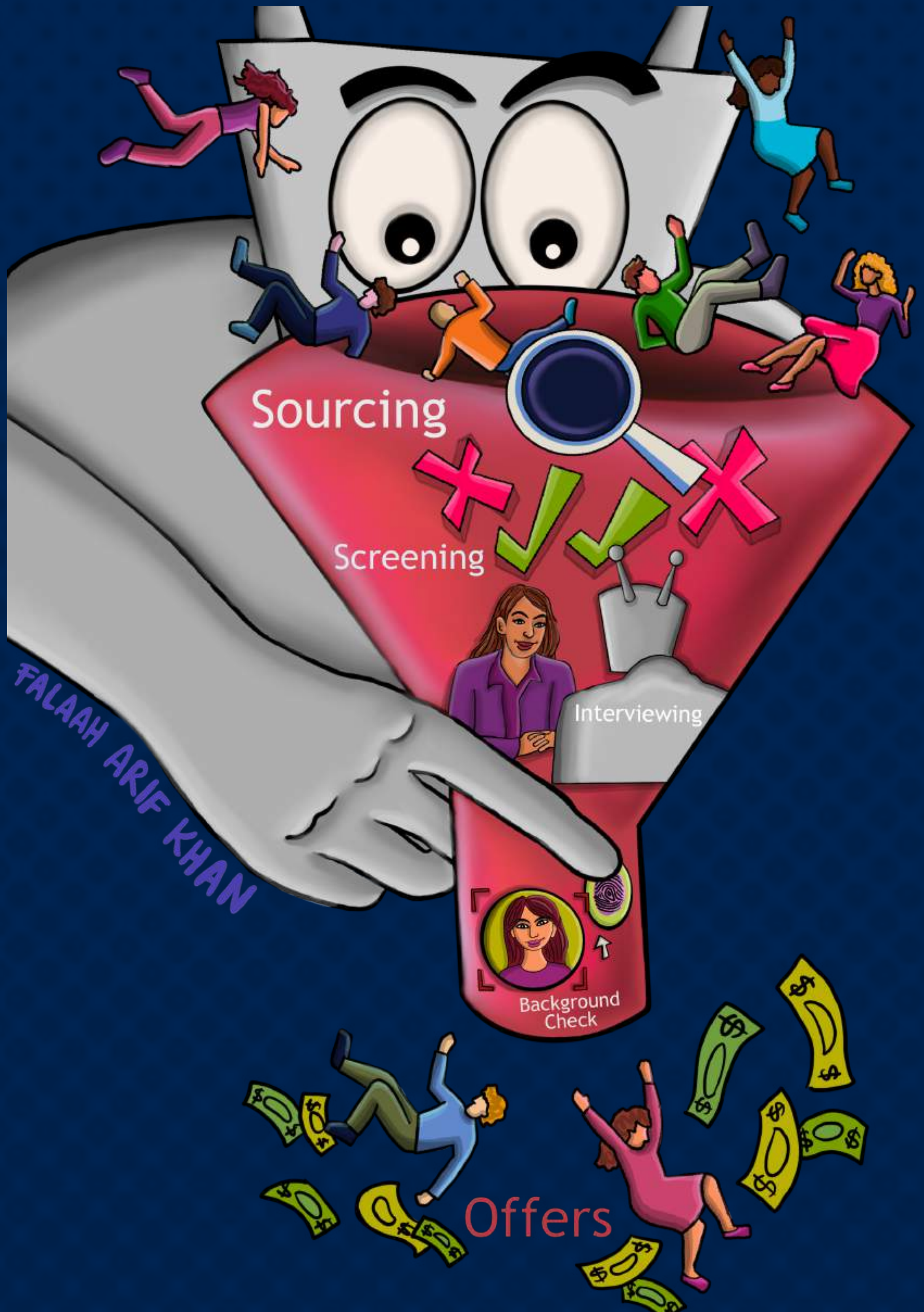
**December 11, 2021**

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.

# A related domain: AI in hiring

"Automated hiring systems act as modern gatekeepers to economic opportunity."

*Jenny Yang*





center for responsible ai

# Algorithmic discrimination

**the guardian** July 2015

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

**MIT Technology Review** February 2013

**Racism is Poisoning Online Ad Delivery, Says Harvard Professor**

**THE WALL STREET JOURNAL.** September 2014

**Are Workplace Personality Tests Fair?**

Growing Use of Tests Sparks Scrutiny Amid Questions of Effectiveness and Workplace Discrimination

**REUTERS** October 2018

**Amazon scraps secret AI recruiting tool that showed bias against women**


center for responsible ai

# The need for regulation

Opinion

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

The New York Times

**March 17, 2021**

**By Alexandra Reeve Givens, Hilke Schellmann and Julia Stoyanovich**

Ms. Givens is the chief executive of the Center for Democracy & Technology. Ms. Schellman and Dr. Stoyanovich are professors at New York University focusing on artificial intelligence.

The bill should also require validity testing, to ensure that the tools actually measure what they claim to, and it must make certain that they measure characteristics that are relevant for the job. **Such testing would interrogate whether, for example, candidates' efforts to blow up a balloon in an online game really indicate their appetite for risk in the real world — and whether risk-taking is necessary for the job.**

# The need for regulation

Opinion

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

The New York Times

**March 17, 2021**

**By Alexandra Reeve Givens, Hilke Schellmann and Julia Stoyanovich**

Ms. Givens is the chief executive of the Center for Democracy & Technology. Ms. Schellman and Dr. Stoyanovich are professors at New York University focusing on artificial intelligence.

In addition, the City Council must require vendors to tell candidates how they will be screened by an automated tool **before** the screening, so candidates know what to expect. **People who are blind, for example, may not suspect that their video interview could score poorly if they fail to make eye contact with the camera.** If they know what is being tested, they can engage with the employer to seek a fairer test.


THESE GHOSTS ARE MAKING THEIR WAY INTO DATA-DRIVEN PRODUCTS AS WELL.
TAKE THE INFAMOUS FACIAL RECOGNITION SOFTWARE THAT HAS BEEN ALL OVER THE NEWS RECENTLY. RACIAL INJUSTICES ARE PROBLEMATIC ENOUGH, BUT HAVE YOU CONSIDERED HOW THESE MODELS DISCRIMINATE AGAINST BLACK DISABLED PEOPLED?
AS DISABILITY RIGHTS ADVOCATE **HABEN GIRMA** EXPLAINS (7),
"MY EYES MOVE INVOLUNTARILY, EACH ONE SWINGING TO ITS OWN MUSIC. THEY'VE DANCED THIS WAY FOR AS LONG AS I CAN REMEMBER."
HOW WELL DO YOU THINK **FACIAL RECOGNITION** WOULD PERFORM ON **BLIND BLACK PEOPLE**?
HAVING BEEN TRAINED ON THE FACIAL DYNAMICS OF SIGHTED WHITE PEOPLE, FACIAL RECOGNITION TECHNOLOGY PEDDLES AN ABLEIST AND RACIST NARRATIVE.

# Nutritional labels for job seekers

THE WALL STREET JOURNAL.

September 22, 2021

## Hiring and AI: Let Job Candidates Know Why They Were Rejected

Labels that explain a hiring process that uses AI could allow job seekers to opt out if they object to the employer's data practices.
PHOTO: ISTOCKPHOTO/GETTY IMAGES

*By Julia Stoyanovich*
Updated Sept. 22, 2021 11:00 am ET

Artificial-intelligence tools are seeing ever broader use in hiring. But this practice is also hotly criticized because we rarely understand how these tools select candidates, and whether the candidates they select are, in fact, better qualified than those who are rejected.

To help answer these crucial questions, **we should give job seekers more information about the hiring process and the decisions**. The solution I propose is a twist on something we see every day: **nutritional labels**. Specifically, job candidates would see simple, standardized labels that show the factors that go into the AI's decision.

https://www.wsj.com/articles/hiring-job-candidates-ai-11632244313

center for responsible ai

# Nutritional labels for job seekers

THE WALL STREET JOURNAL.

September 22, 2021

## Hiring and AI: Let Job Candidates Know Why They Were Rejected

Labels that explain a hiring process that uses AI could allow job seekers to opt out if they object to the employer's data practices.

PHOTO: ISTOCKPHOTO/GETTY IMAGES

*By Julia Stoyanovich*

Updated Sept. 22, 2021 11:00 am ET

**ACCOUNTANT**

**Acme Partners**

| | |
|---|---|
| **Qualifications:** | BS in accounting, GPA >3.0, Knowledge of financial and accounting systems and applications |
| **Personal data to be analyzed:** | An AI program could be used to review and analyze the applicant's personal data online, including LinkedIn profile, social media accounts and credit score. |
| **Additional assessment:** | AI-assisted personality scoring |

**ALERT: Applicants for this position DO NOT have the option to selectively decline use of AI analysis for any of their personal data or to review and challenge the results of such analysis.**

https://www.wsj.com/articles/hiring-job-candidates-ai-11632244313

# Nutritional labels for public disclosure

## ACCOUNTANT
**Acme Partners**

| | |
|---|---|
| **Qualifications:** | BS in accounting, GPA >3.0, Knowledge of fi accounting systems and applications |
| **Personal data to be analyzed:** | An AI program could be used to review and analyze the applicant's personal data online, including LinkedIn profile, social media accounts and credit score. |
| **Additional assessment:** | AI-assisted personality scoring |

**ALERT: Applicants for this position DO NOT have the option to selectively decline use of AI analysis for any of their personal data or to review and challenge the results of such analysis.**

https://www.wsj.com/articles/hiring-job-candidates-ai-11632244313

**comprehensible**: short, simple, clear

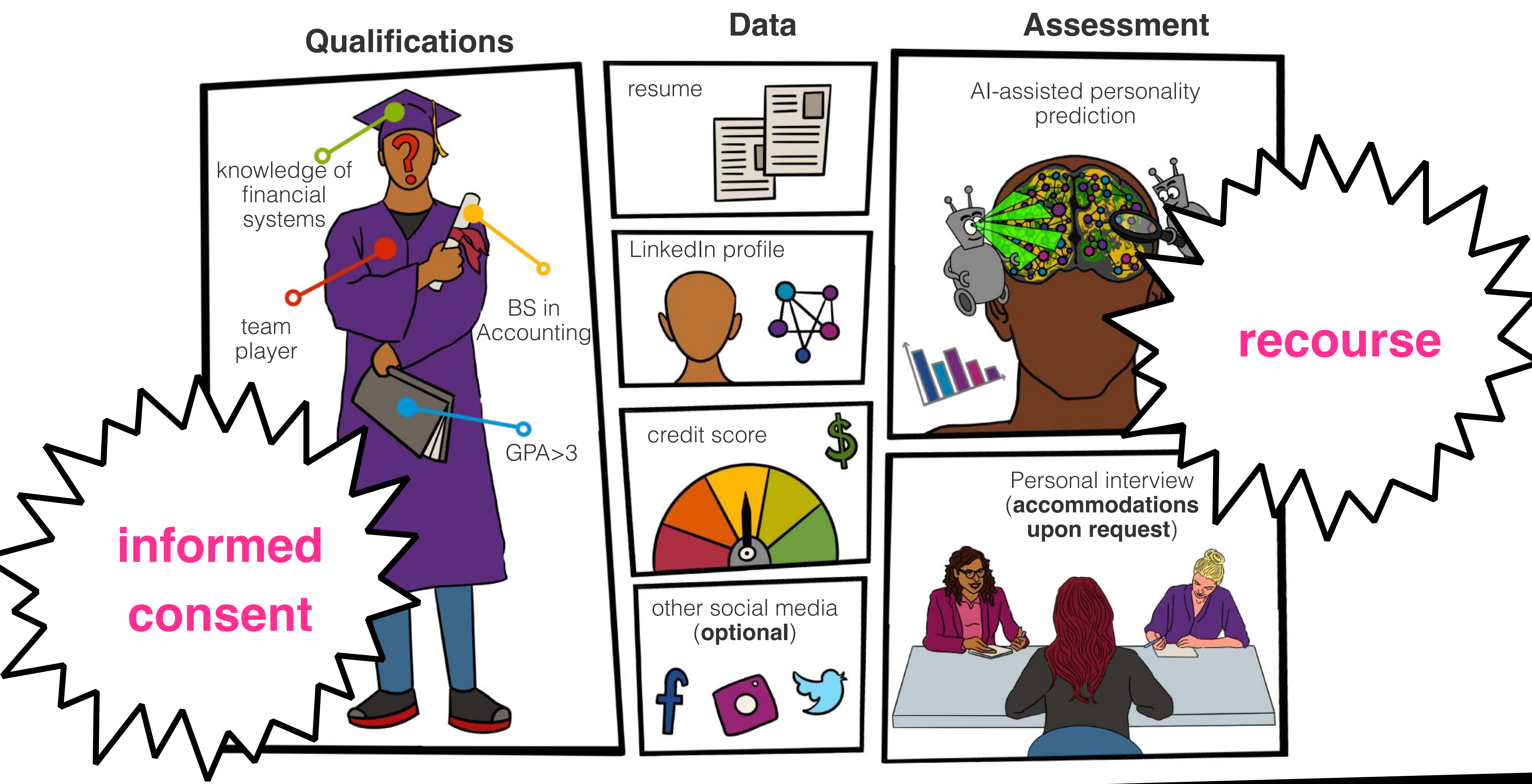
**consultative**: provide actionable info

**comparable:** implying a standard

[Stoyanovich & Howe, 2019]

r/ai center for responsible ai

# Anatomy of a job posting label




https://www.wsj.com/articles/hiring-job-candidates-ai-11632244313

r/ai center for responsible ai

# course overview

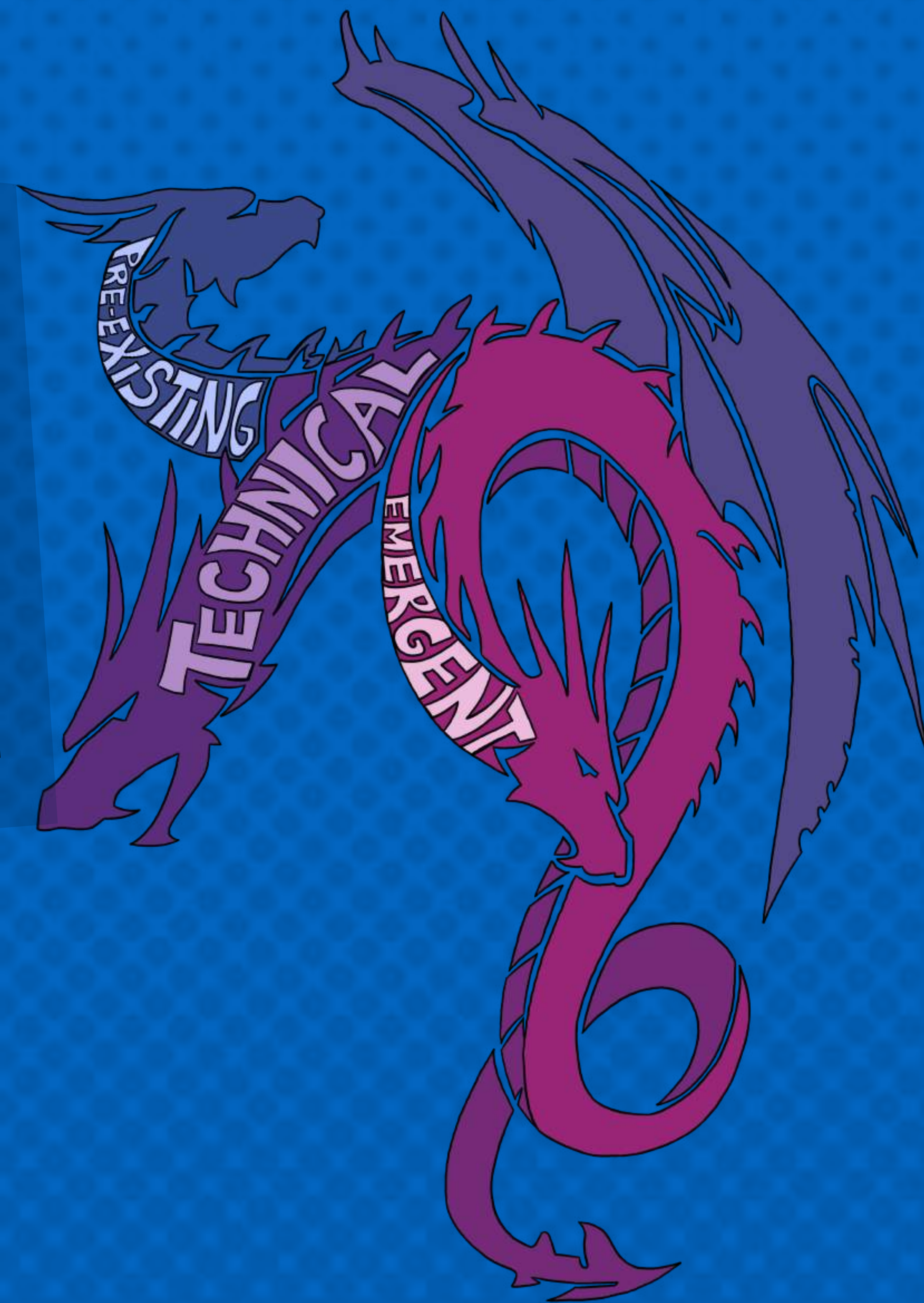
# **module 1:** algorithmic fairness

# Bias in computer systems

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

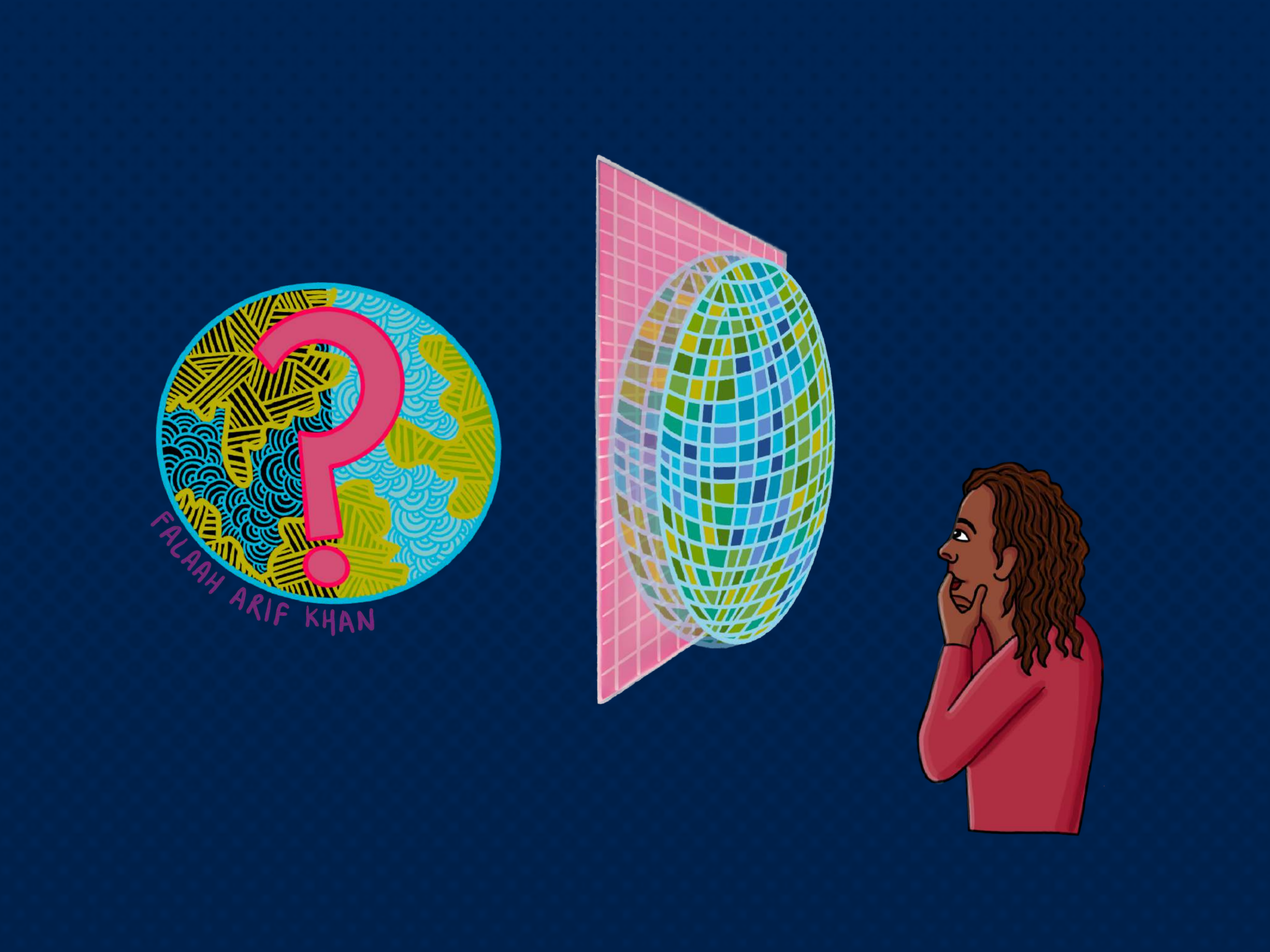
**Technical:** introduced or exacerbated by the technical properties of an ADS

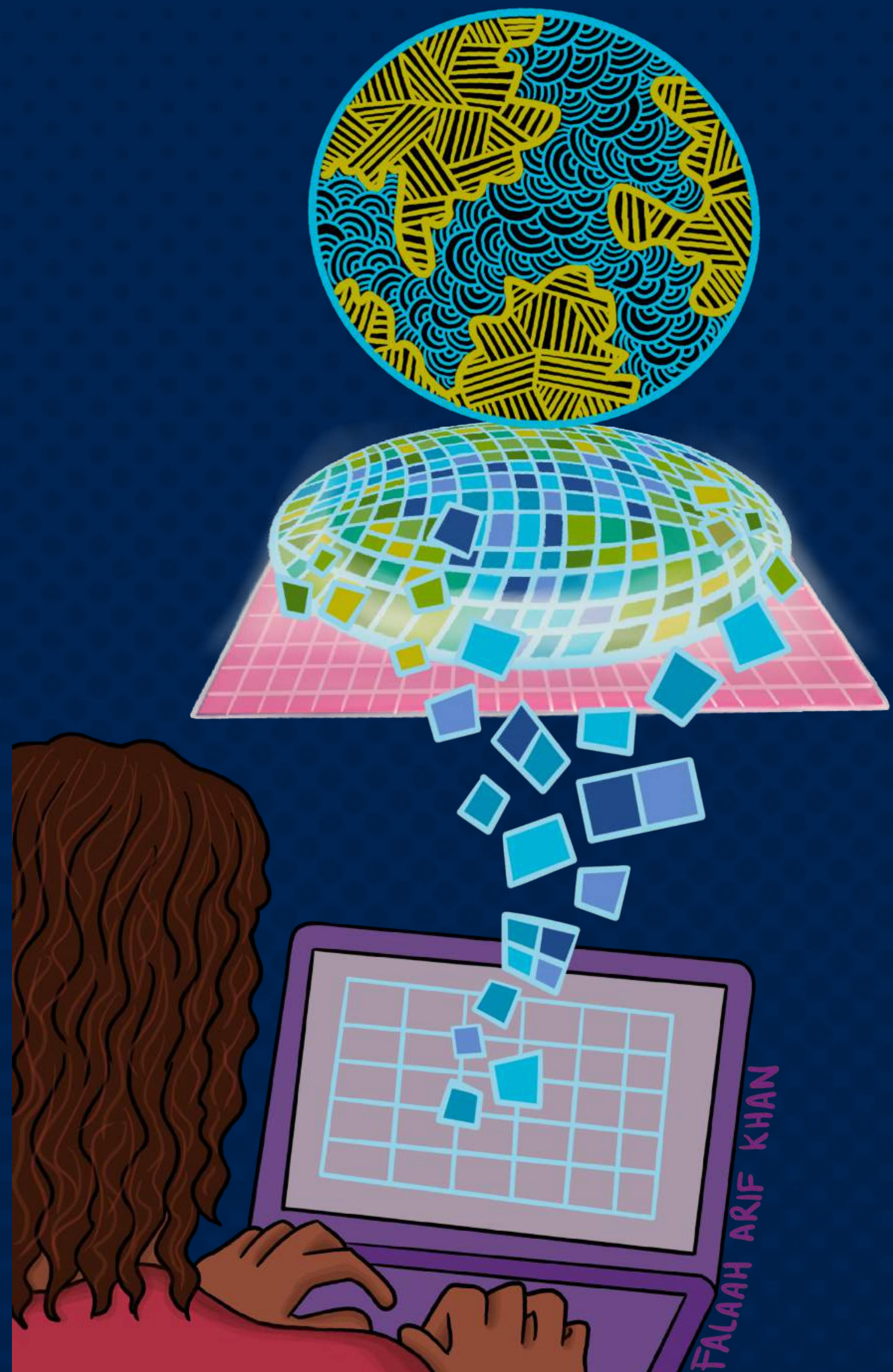
**Emergent:** arises due to context of use



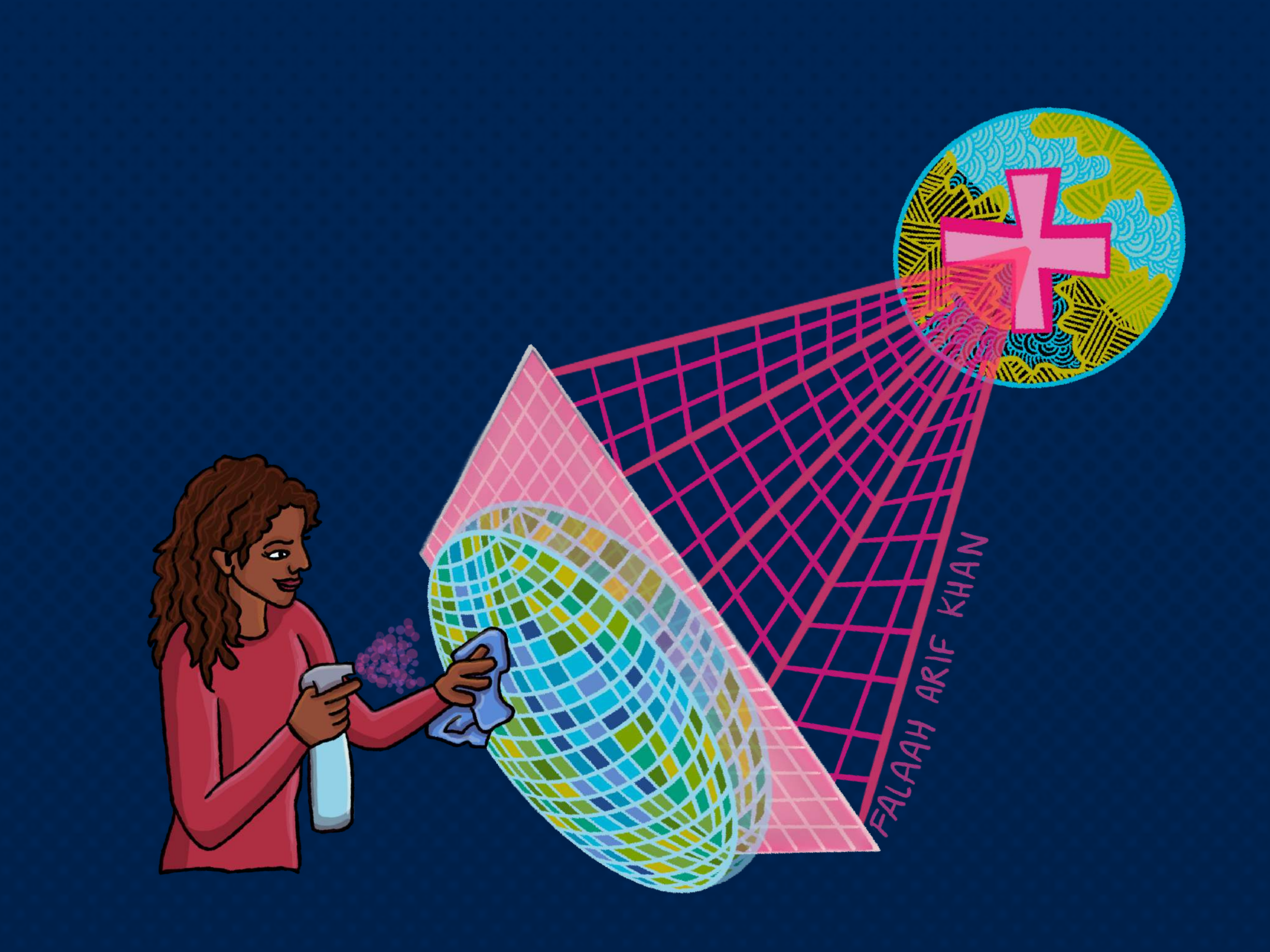


[Friedman & Nissenbaum (1996)]


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

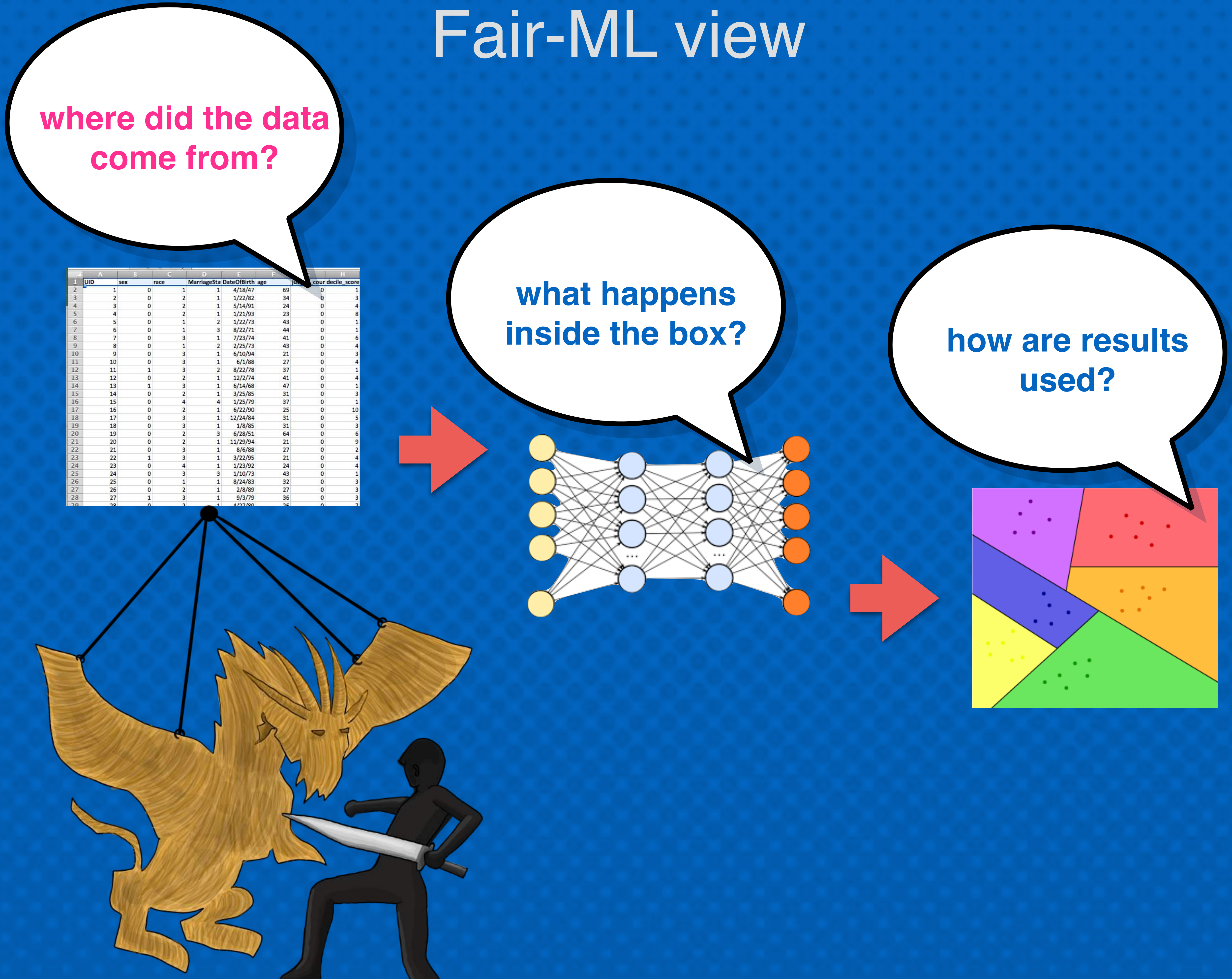
**Technical**: introduced or exacerbated by the technical properties of an ADS

**Emergent:** arises due to context of use

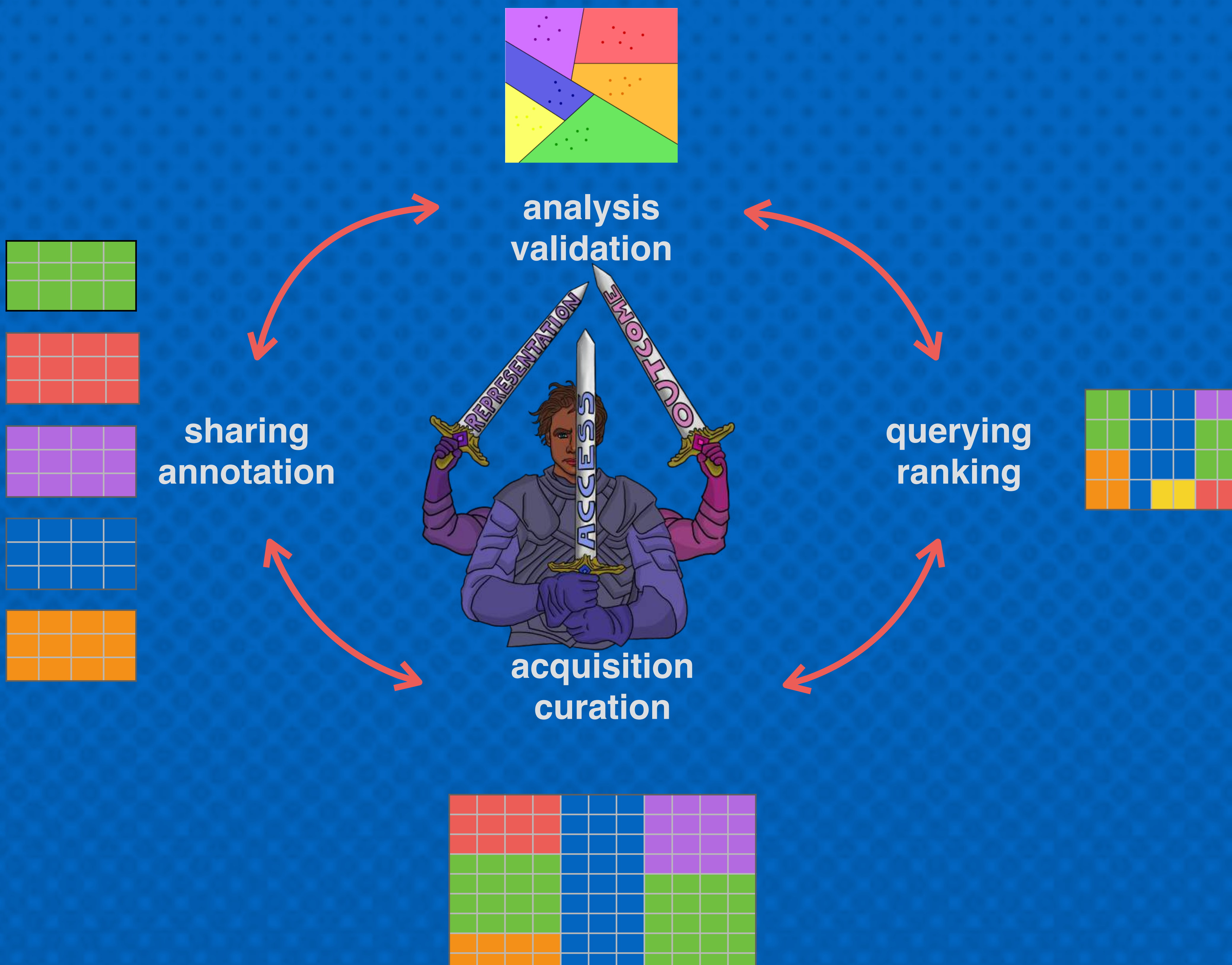
**to fight bias, state beliefs and assumptions explicitly**


[Friedman & Nissenbaum (1996)]

# Fair-ML view


where did the data come from?
UID sex race MarriageSta DateOfBirth age decile_score
what happens inside the box?
how are results used?

# Lifecycle view




[Stoyanovich, Howe, Jagadish (2020)]

# Models and assumptions

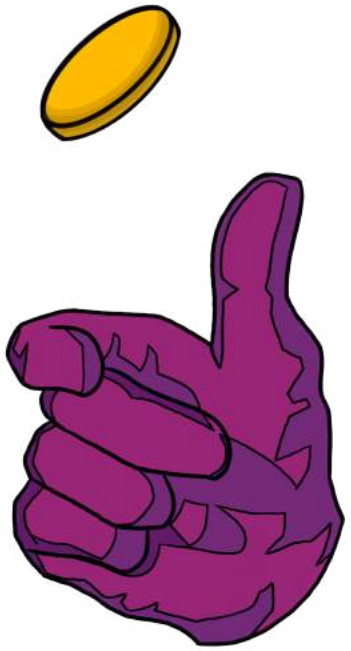
[Stoyanovich, Howe, Jagadish (2020)]

# **module 3:** data protection & privacy

# Privacy: two sides of the same coin

**Did you go out drinking over the weekend?**

$$\frac{\text{protecting an individual}}{\text{plausible deniability}}$$



$$\frac{\text{learning about the population}}{\text{noisy estimates}}$$

# 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

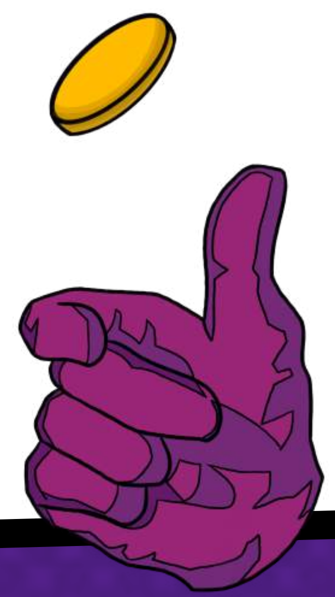
thus, we estimate **p** as:

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

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

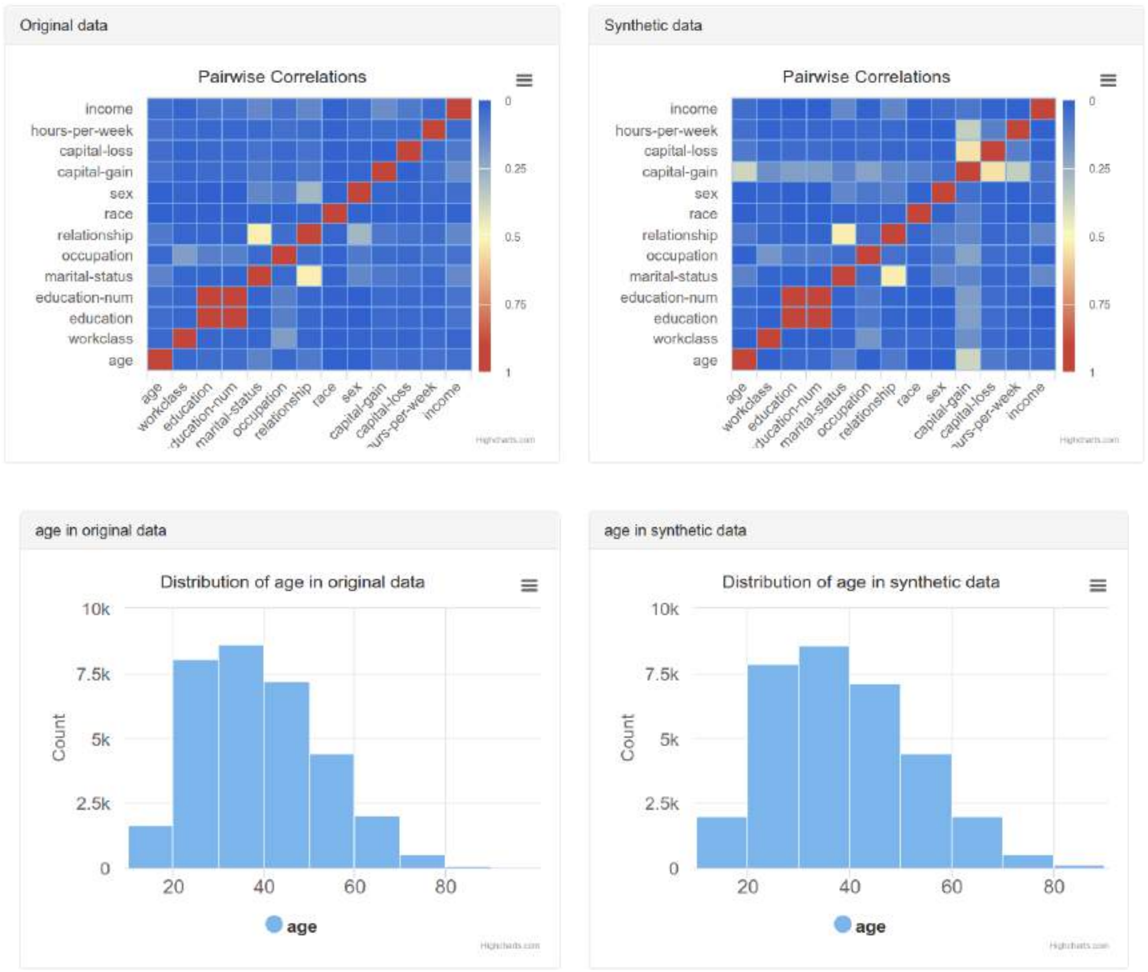
**What does it mean to preserve privacy?**

BY CYNTHIA DWORK

# A Firm Foundation for Private Data Analysis

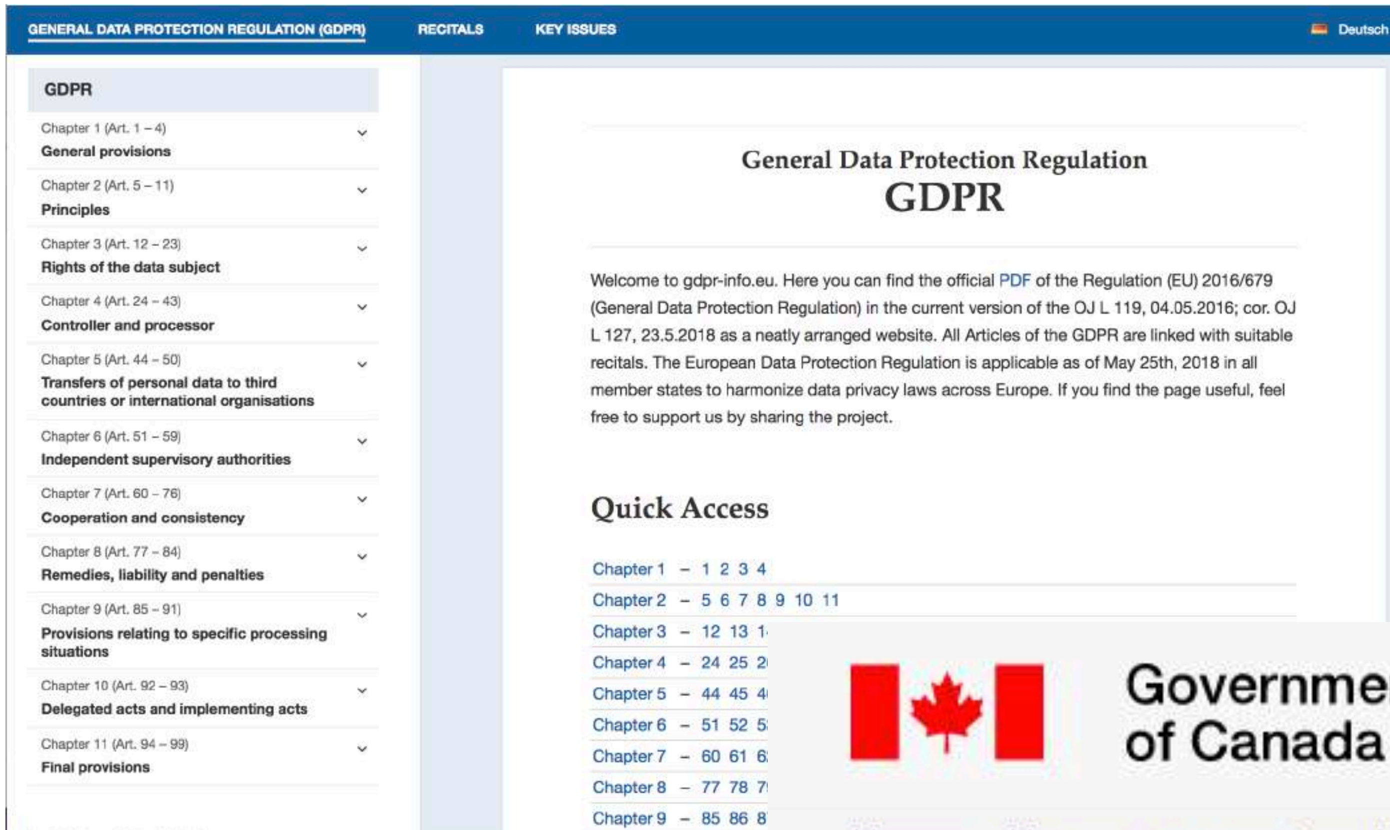
Communications of the ACM CACM Homepage archive

Volume 54 Issue 1, January 2011
Pages 86-95



# Regulating ADS?

# Legal frameworks




# **module 4:**
transparency & interpretability

# The evils of discrimination

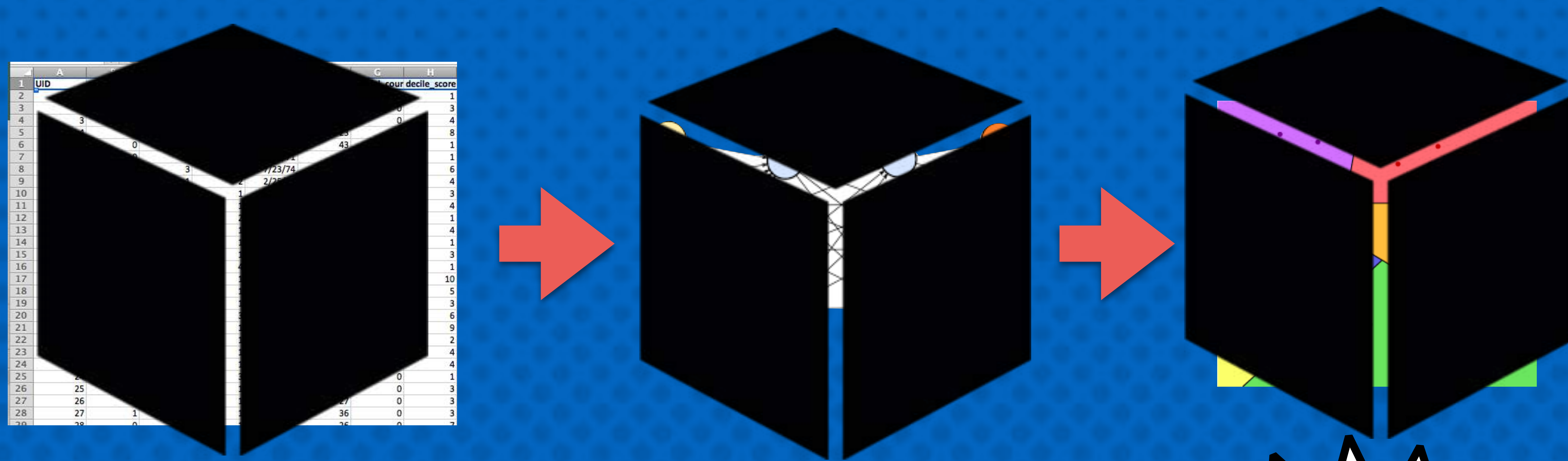
**Disparate treatment**

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

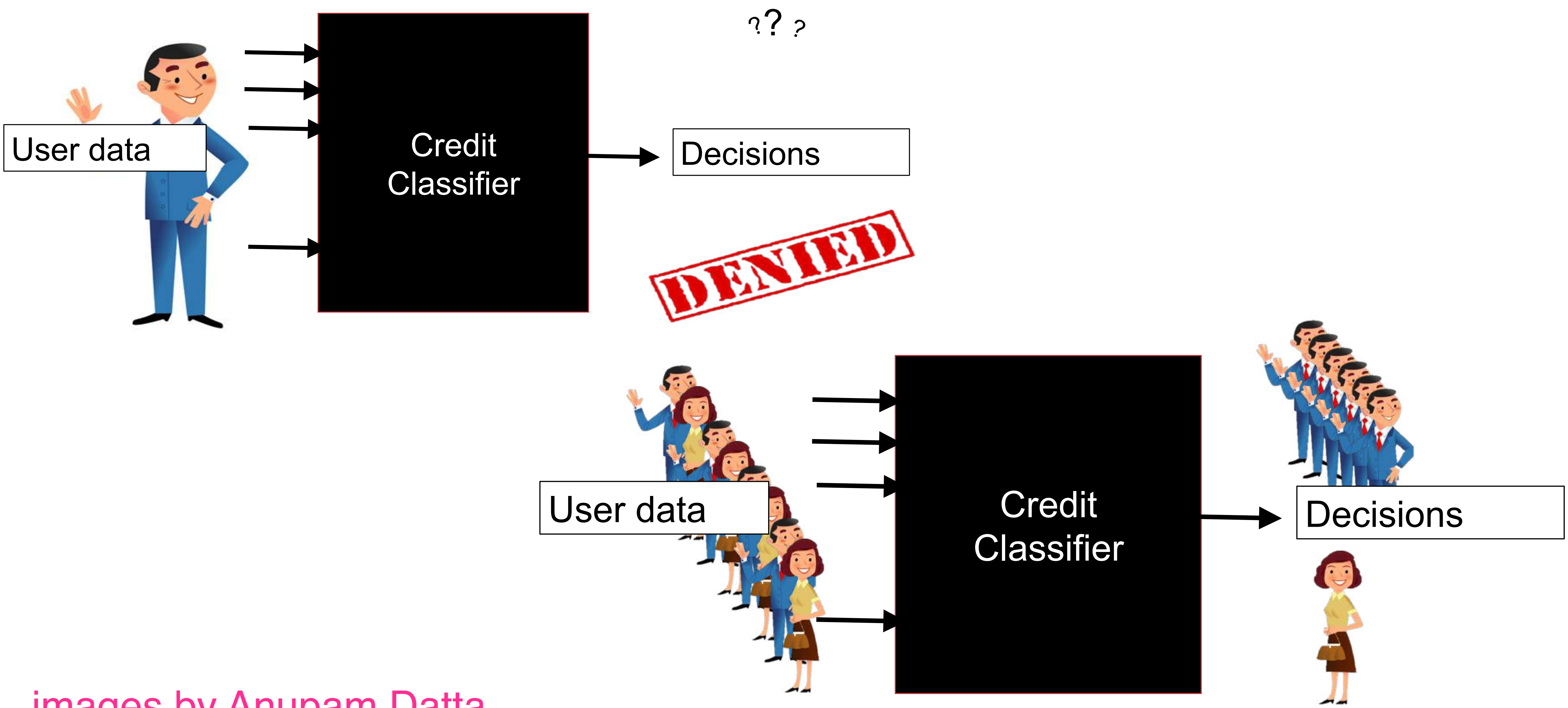


**Fair Housing Act**

**Equal Credit Opportunity Act, 1964**

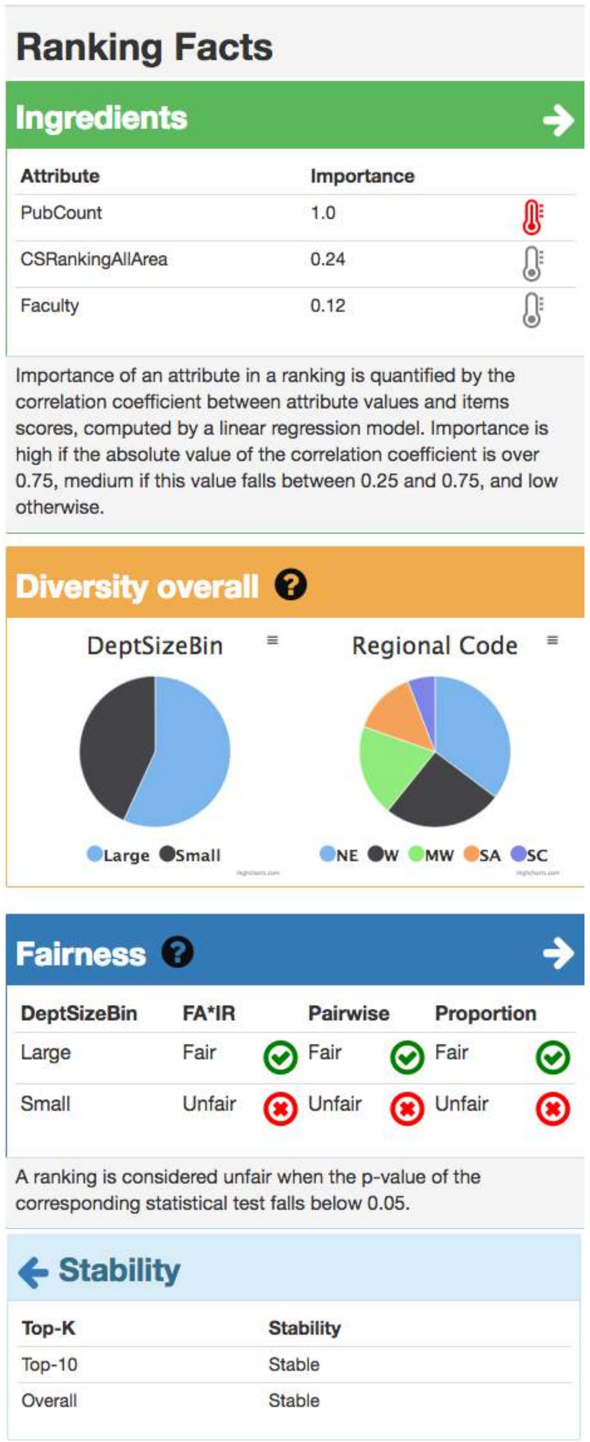
**Civil Rights Act, 1964**

# Auditing black-box models




images by Anupam Datta

# Nutritional labels




**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:

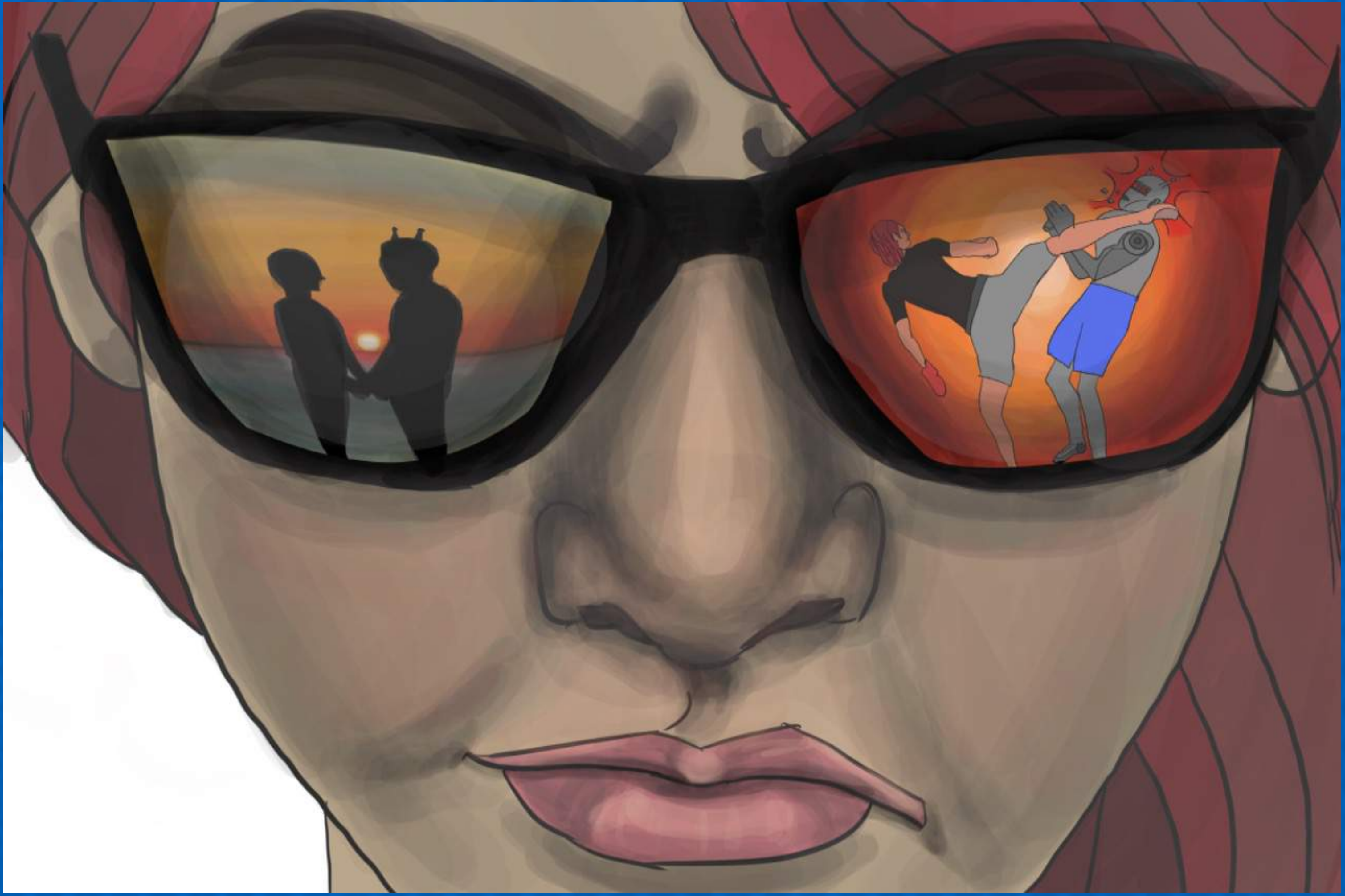
1. fairness, accountability and transparency
2. data engineering
3. privacy & data protection

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

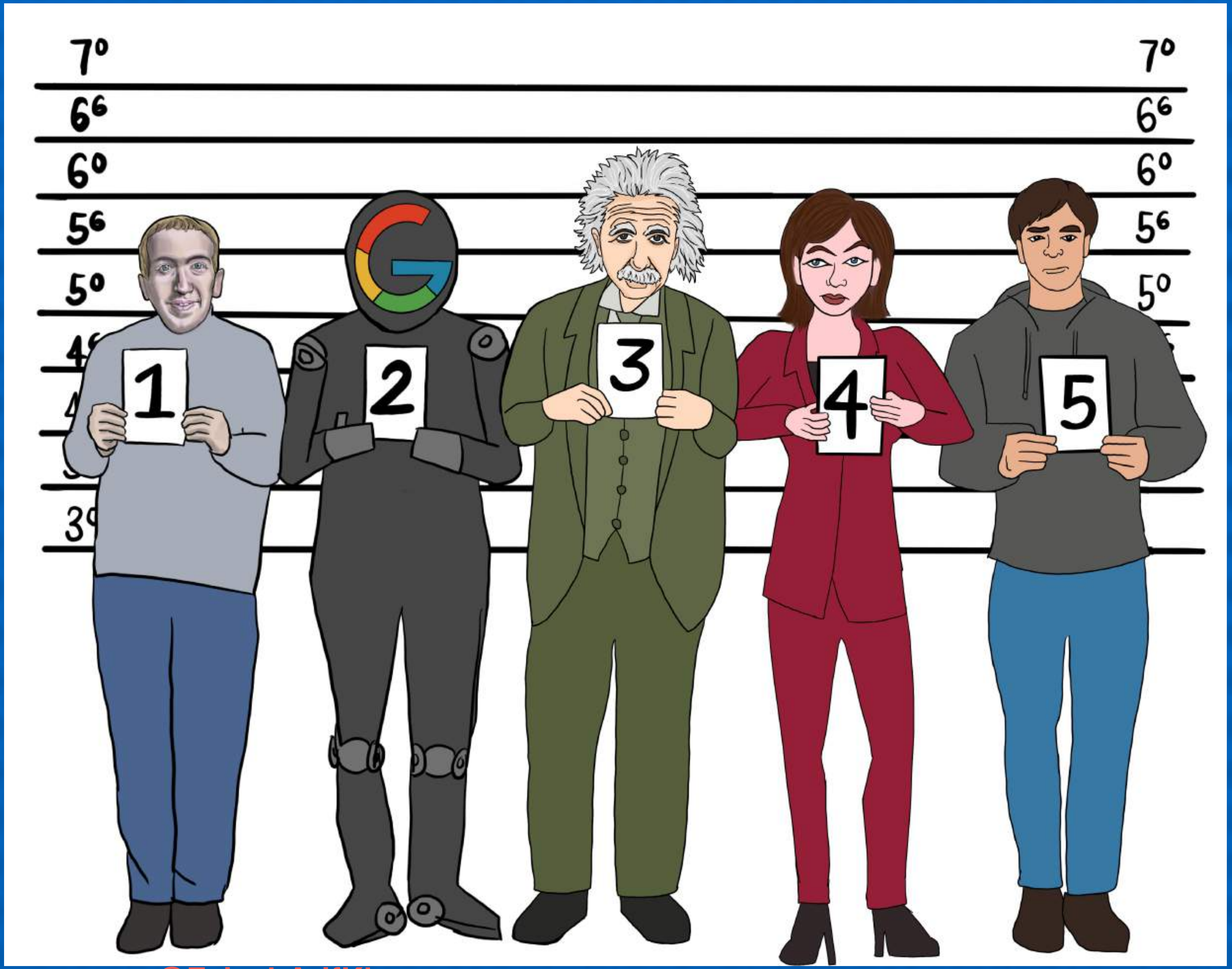
Bottom line: we will learn that many of the problems are **socio-technical**, and so cannot be "solved" with technology alone.

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

# Nuance, please!



# We all are responsible




@FalaahArifKhan

# Responsible Data Science

Introduction and Overview

# Thank you!

NYU TANDON SCHOOL OF ENGINEERING

NYU Center for Data Science

r/ai