Responsible Data Science

Introduction and Overview

January 23, 2023

Prof. Elisha Cohen
Center for Data Science

course logistics

Instructor: Elisha Cohen

Faculty Fellow at the Center for Data Science

Ph.D. in Political Science from Emory University
M.A. Hunter College CUNY
B.A. University of Wisconsin - Madison

Research:

- Develop statistical methods for assessing fairness and bias
- Discrimination in Policing, police use of force, officer involved shootings, hiring

Office hours: Wednesdays noon - 1pm EST and by appointment

Course Staff

Section Leader & Grader: **Nan Wu**
Office hours: Thursdays 9–10am, via Zoom

Section Leader: **Raphael Meyer**
Office hours: TBD

Grader: **Buz Galbraith, Akshit Gandhi, and Anirudh Ghildiyal**

Meeting times

Lectures: Mondays and Wednesdays, 4:55 - 6:10 PM
Room 121, Meyer Hall

Lab 002: Fridays, 8:00 - 8:50 AM
Lab 003: Fridays, 9:00 - 9:50 AM
Lab 004: Fridays, 3:45 - 4:35 PM
Lab 005: Fridays, 2:45 - 3:35 PM

Lab location: 60 Fifth Avenue, Room 110

Assignments and grading

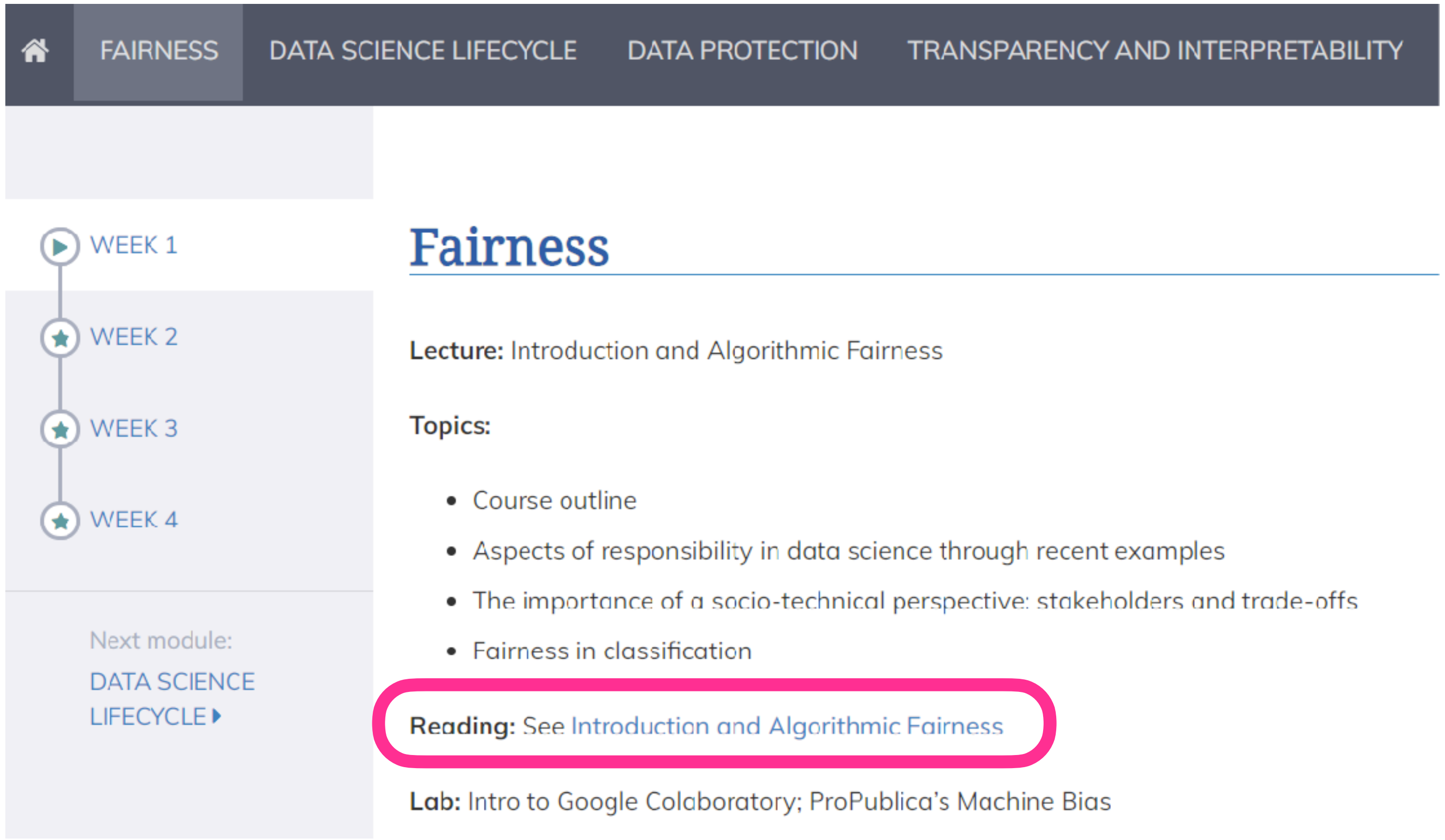
Grading: homeworks: 10% x 3 = 30%
project: 30%
final exam: 20%
labs: 10%
Quizzes: 10%

Late policy: For homeworks, 2 late days over the term, no questions asked. If a homework is submitted a few hours late — a day is used in full. No credit for late submissions once your late days have been used up. No late days for project.

Assignment schedule posted to Brightspace, see calendar or assignments tab.

Where to find information

Website: **https://dataresponsibly.github.io/rds/** slides, lab notebooks, reading



Brightspace: everything assignment-related, Zoom links for lectures and labs, announcements, discussion board

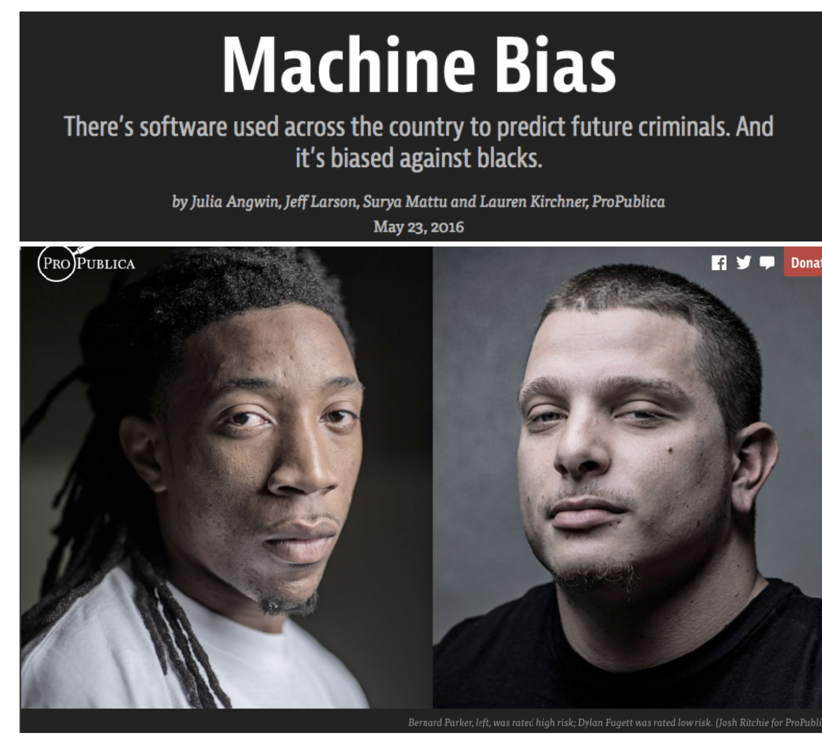
This week's reading


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



Comics by Professor Stoyanovich, Falaah Arif Khan, Mona Sloane

Requirements of this course

This course will require that you:

- **Program in Python**: importing data, inspecting data, wrangling, visualization, writing technical programs

- **Don't leave homeworks and project to the last minute**; they are substantial pieces of work and you are given 3+ weeks for each homework

- **Work in partnerships**: the project is a joint submission with a fellow student what is responsible DS?

Algorithms, AI, ML, Data Science

Algorithm

a step-by-step **set of instructions** that tell a computer what to do with a given input

Data Science

Can involve AI, always involves algorithms. Usually covers the **pipeline** from data processing to modeling

Artificial Intelligence (AI)

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

Machine learning (ML) is a subset of AI in which a program learns patterns from data

[Kearns and Roth, 2019]

The DS chef

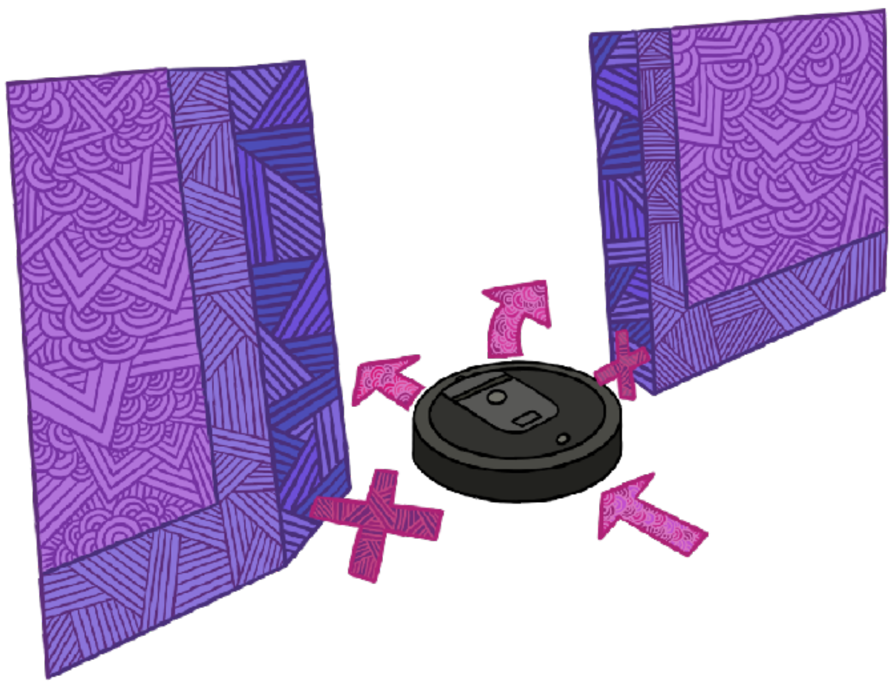
Data: flour, sugar, baking powder, baking soda, butter, milk, eggs

Algorithm 1: remove recipes that include ingredients not in data

Algorithm 2 (AI, ML): check if system recommended pancakes recently. If no: recommend pancakes. If yes: recommend waffles.

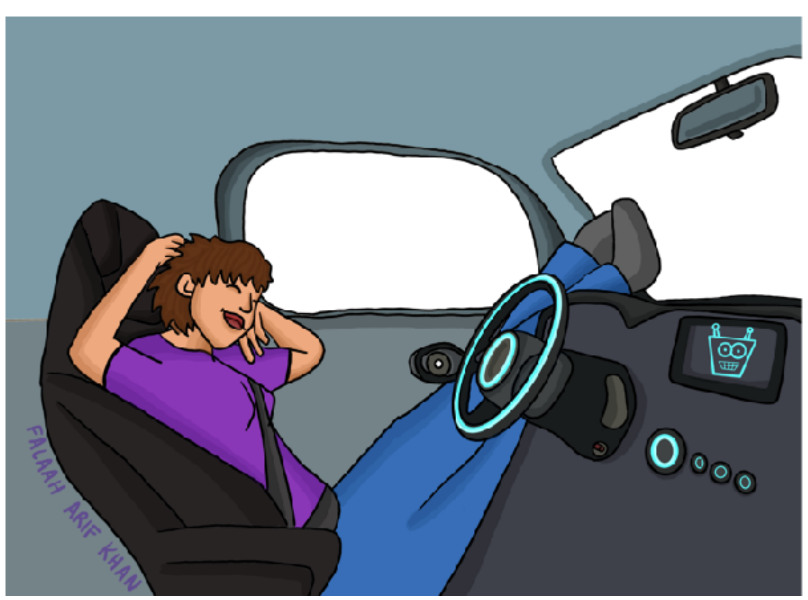
Does the **DS chef** have enough information to make good (or safe) recommendations?

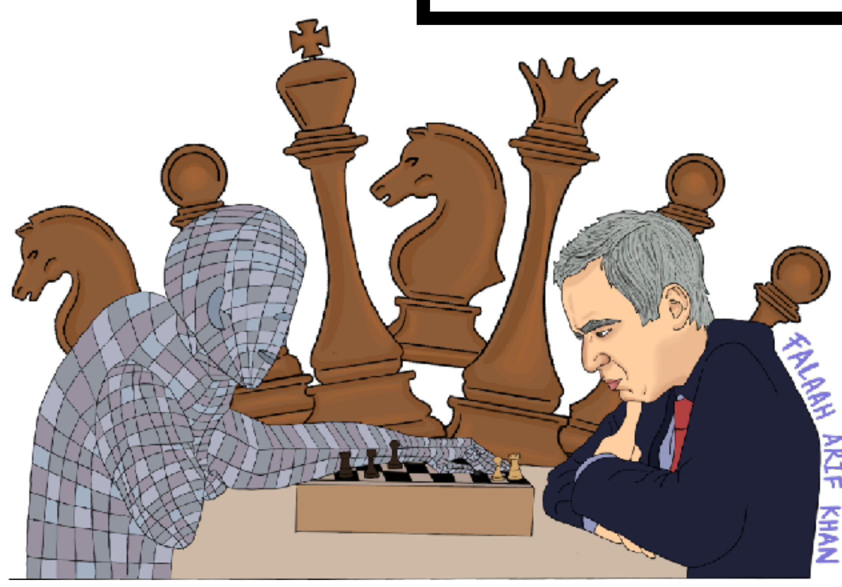
AI: algorithms, data, decisions

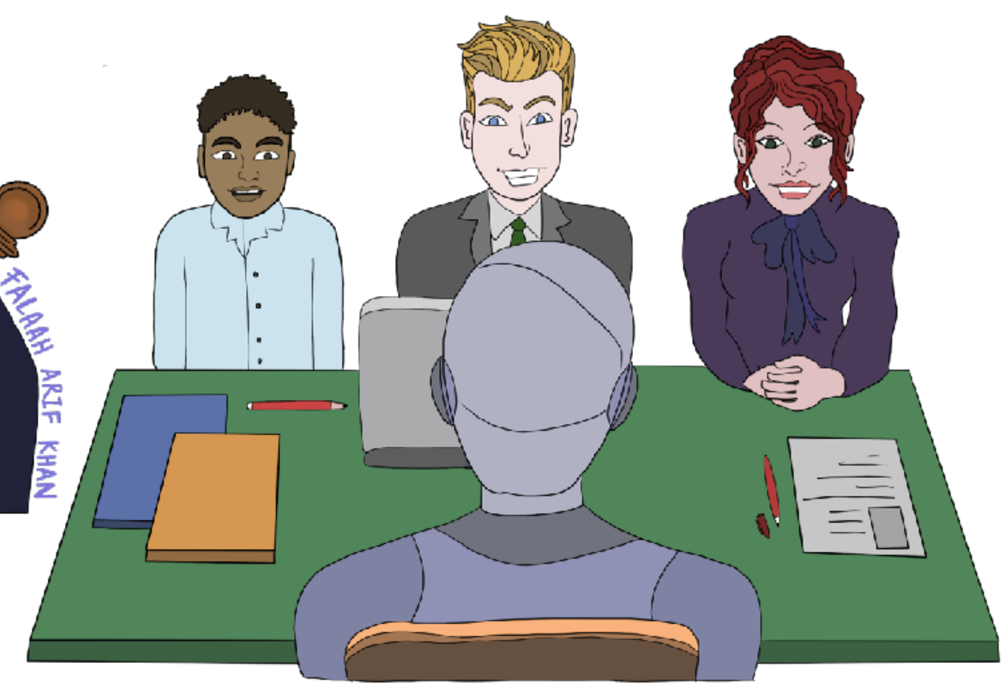


Artificial Intelligence (AI)

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







The promise of AI

Opportunity

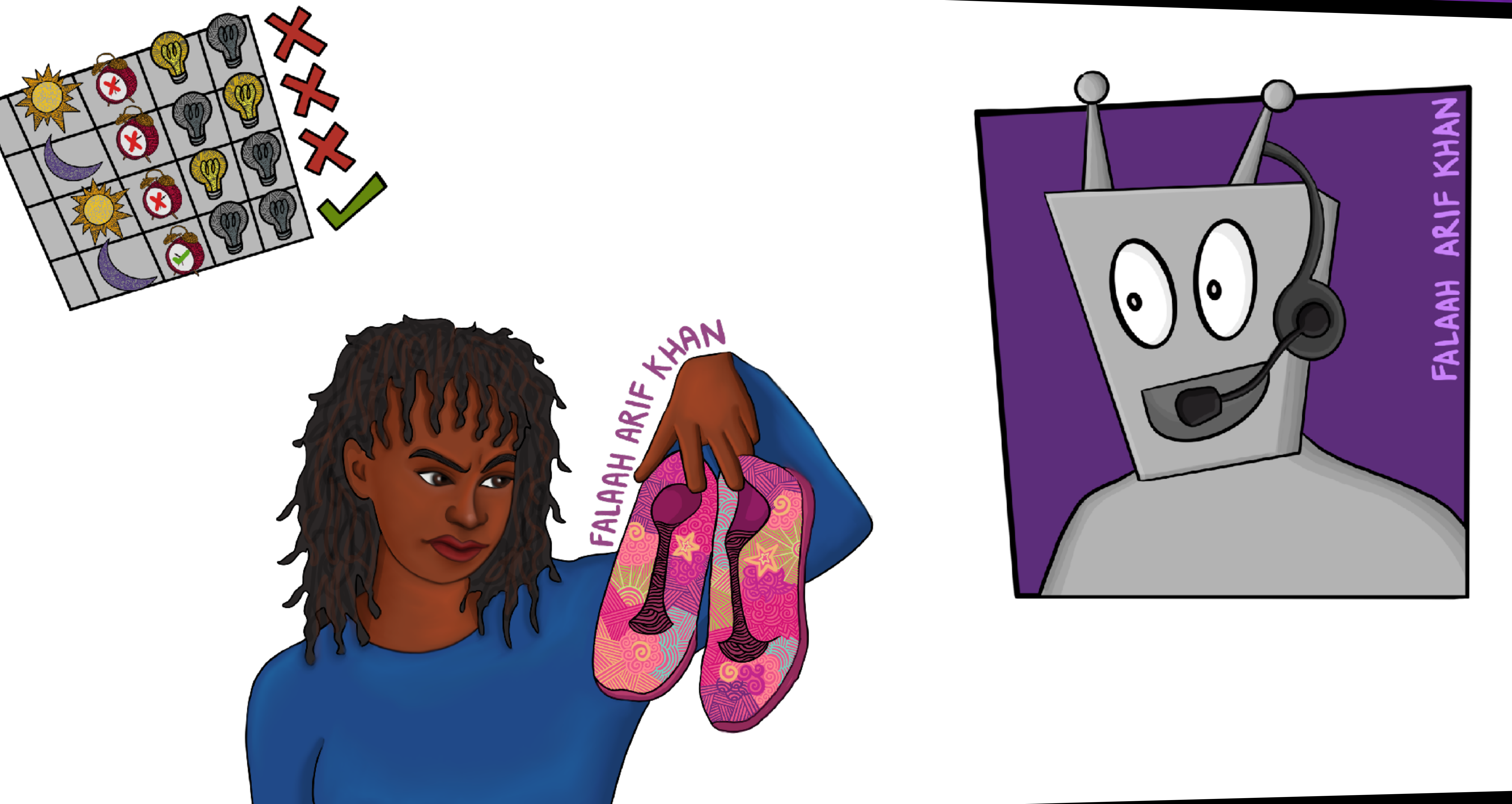
make our lives convenient

accelerate science

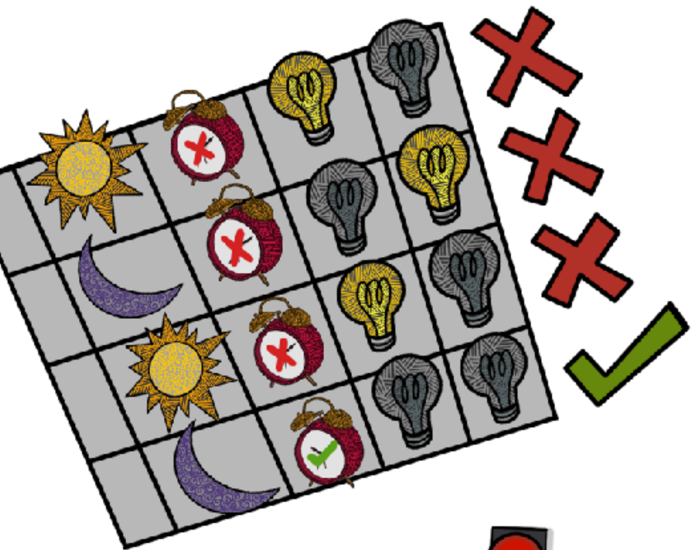
boost innovation

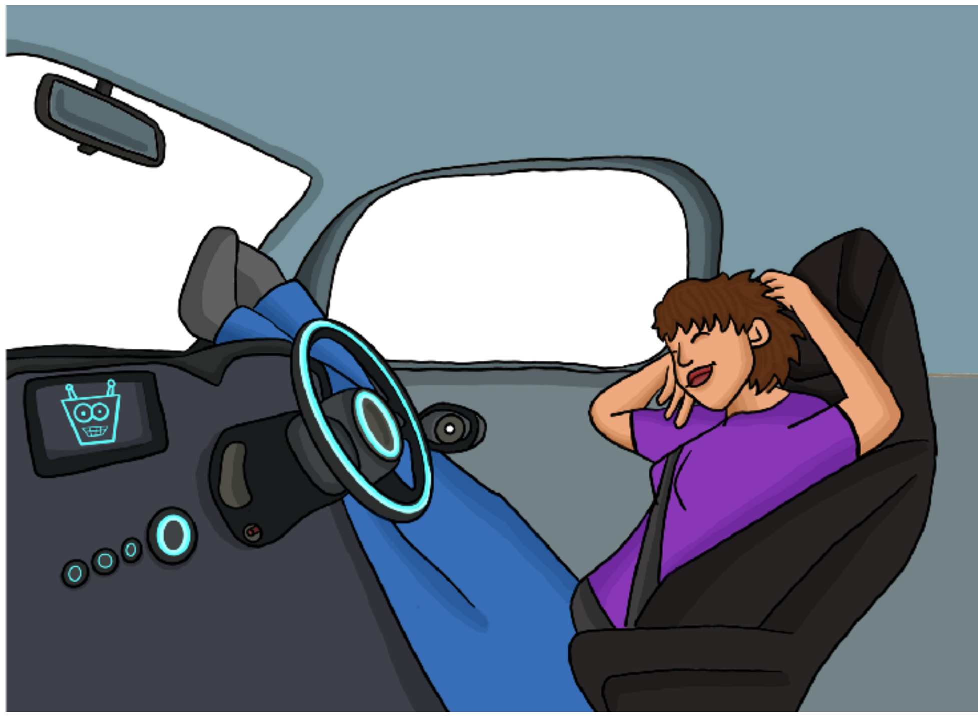
transform government

Machines make mistakes

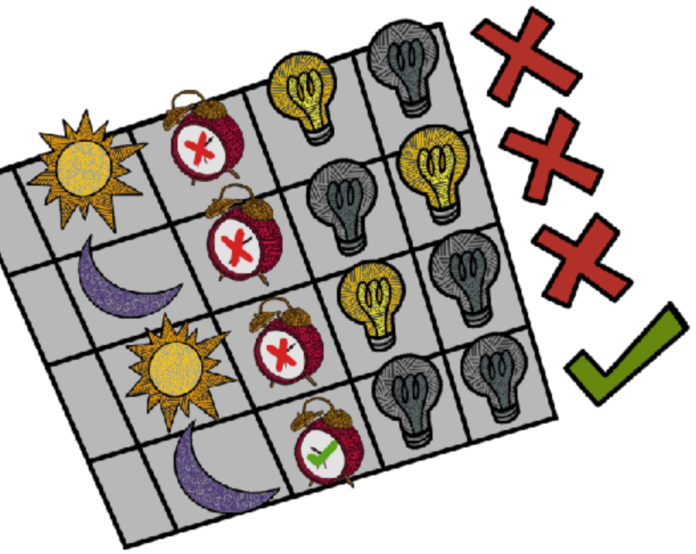


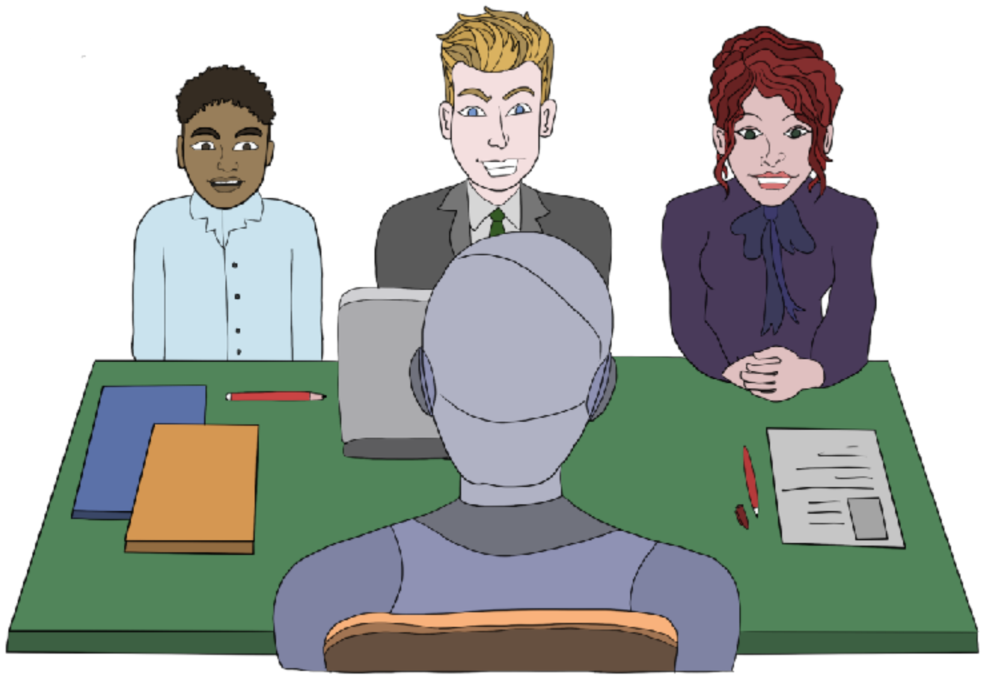
Mistakes lead to harms

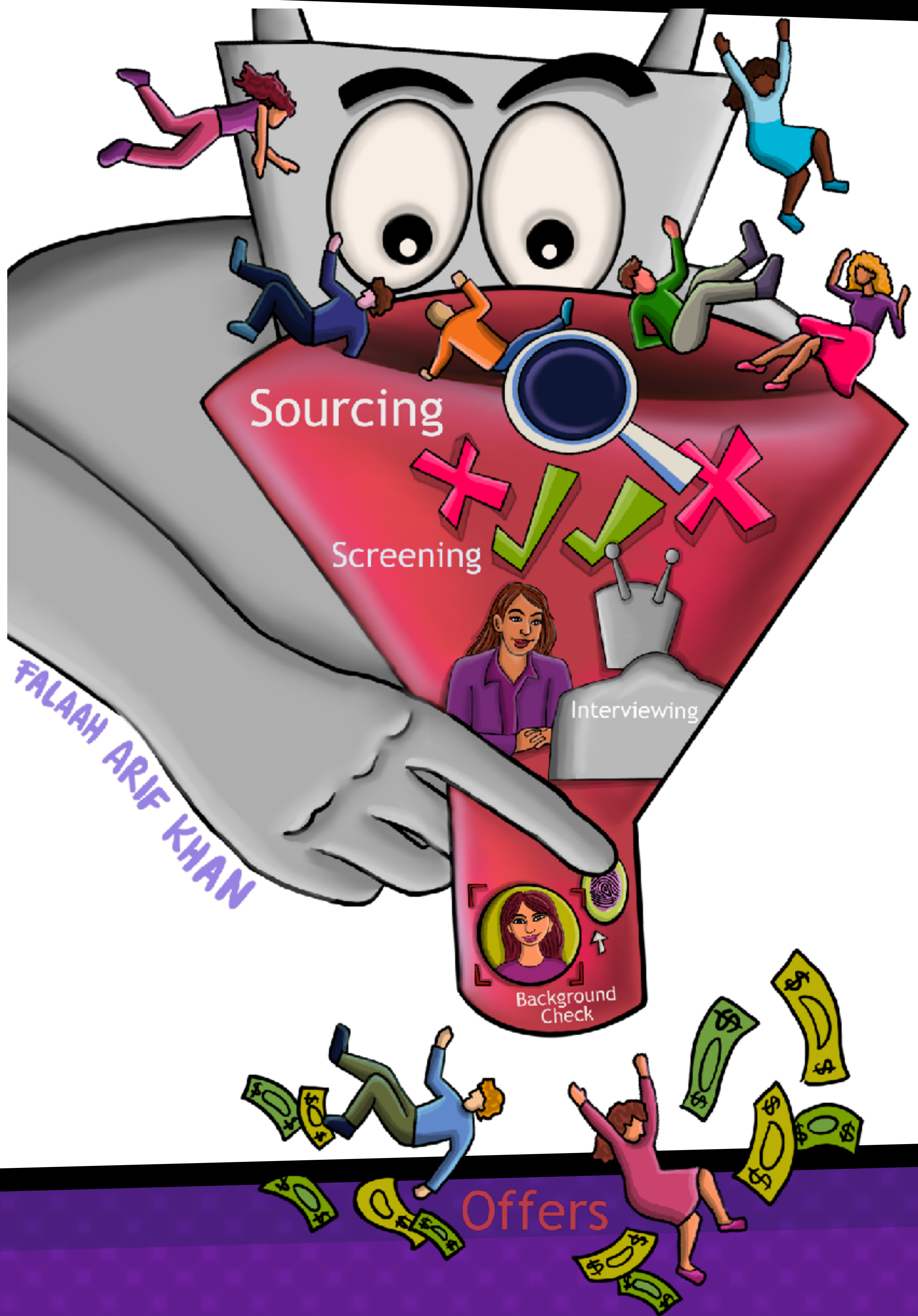




Harms can be cumulative







what is responsible DS?

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 investig
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 datase 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

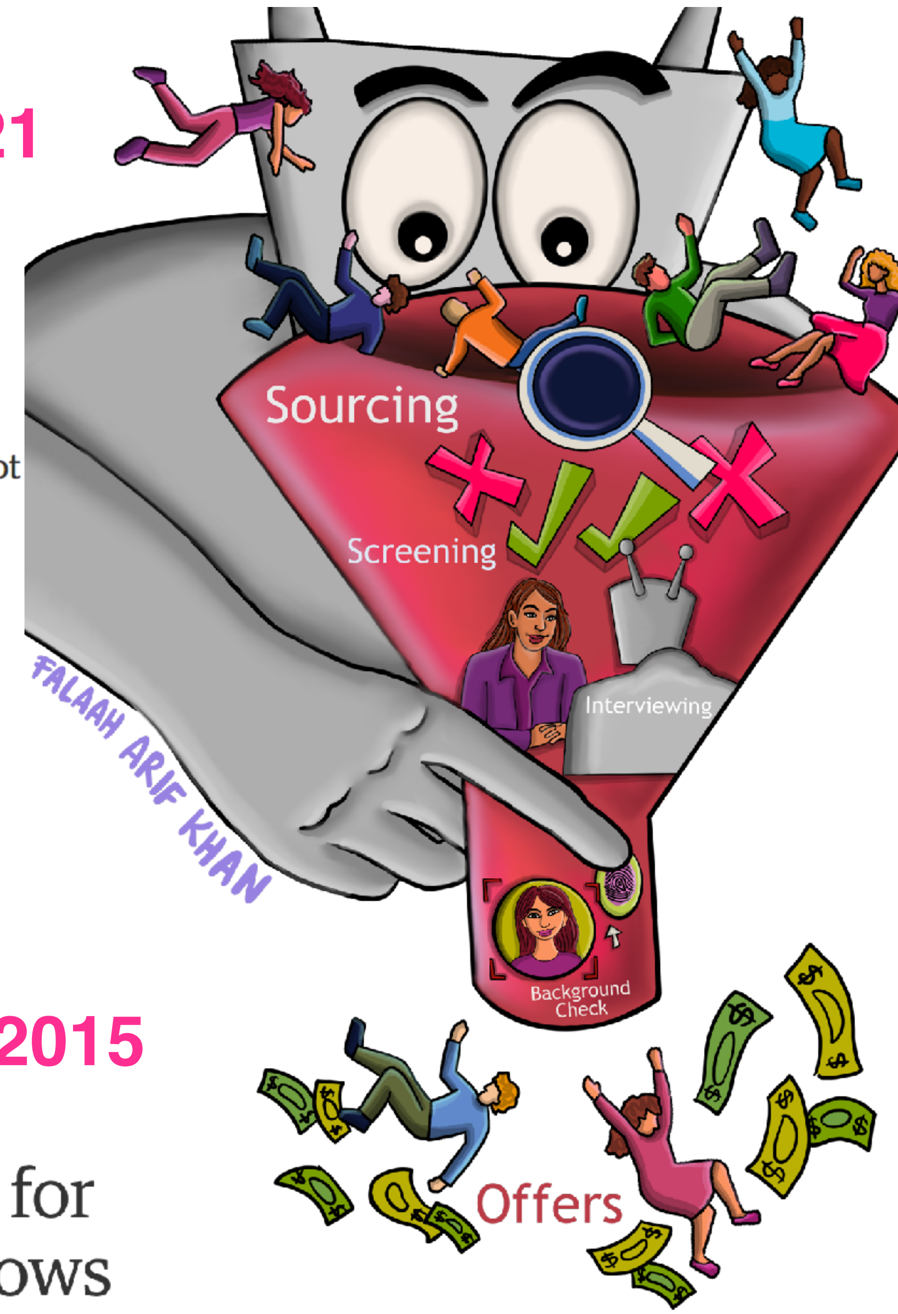
Artificial intelligence used to evaluate job candidates must not become a tool that exacerbates discrimination.

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 DS 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

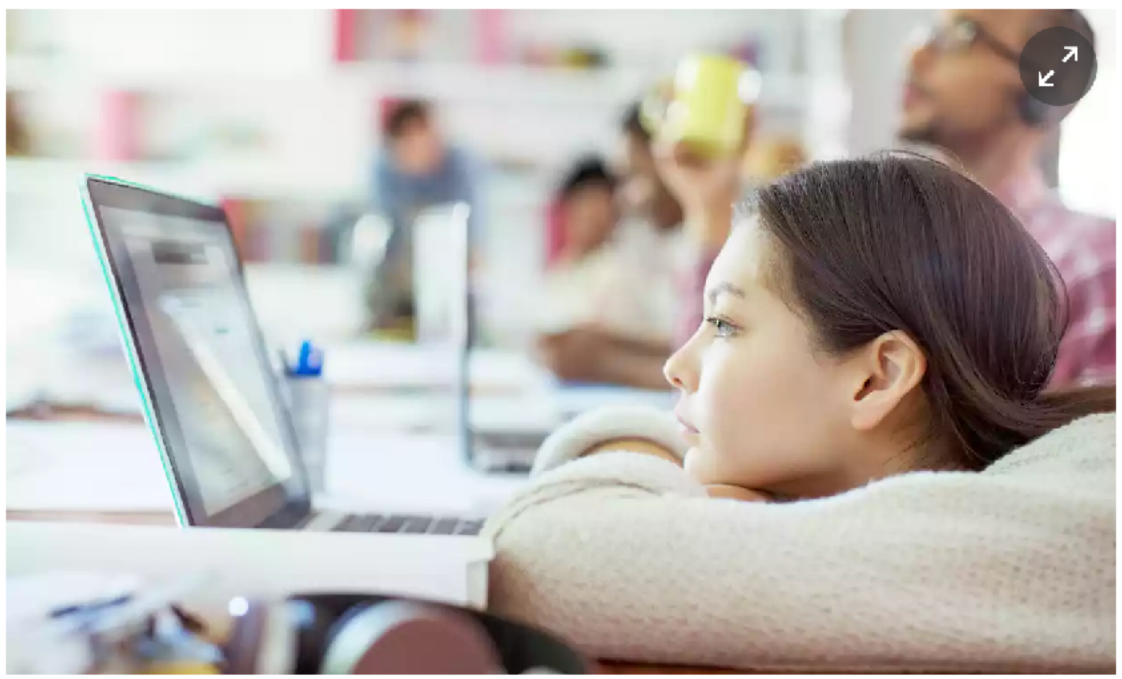
July 2015

Samuel Gibbs

Wednesday 8 July 2015 11.29 BST

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

r/ai

Use case: Resume screening

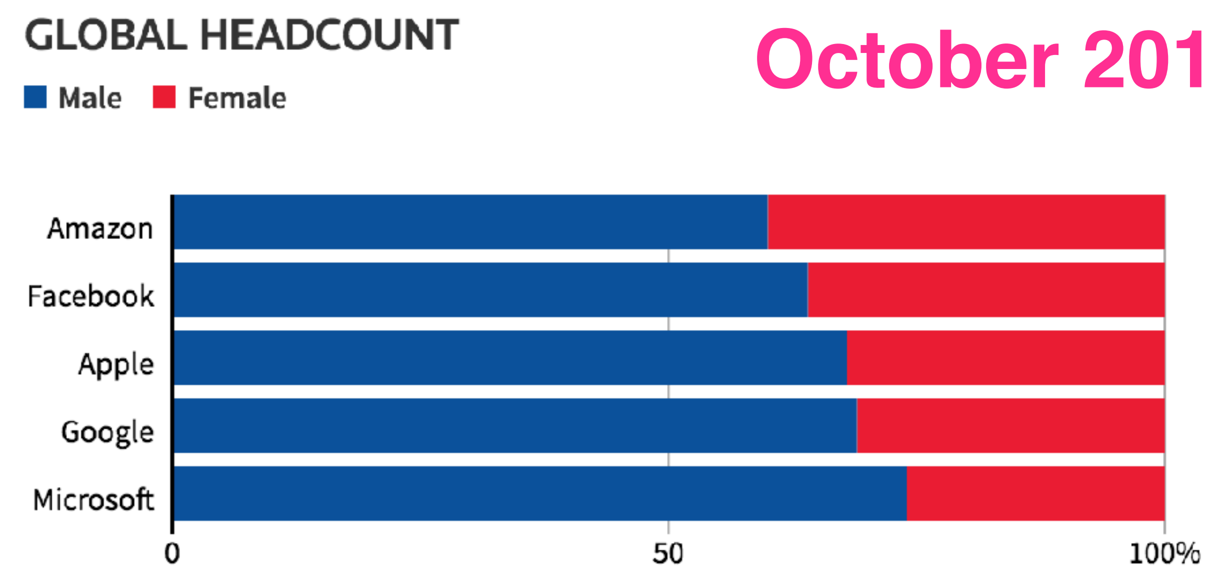
REUTERS Jeffrey Dastin

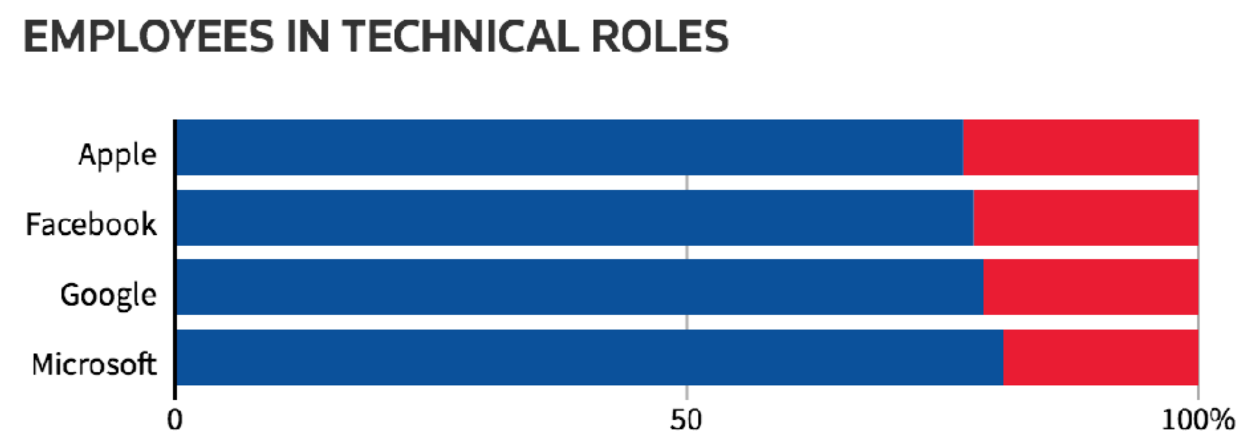
BUSINESS NEWS OCTOBER 9, 2018 / 11:12 PM / 6 MONTHS AGO

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



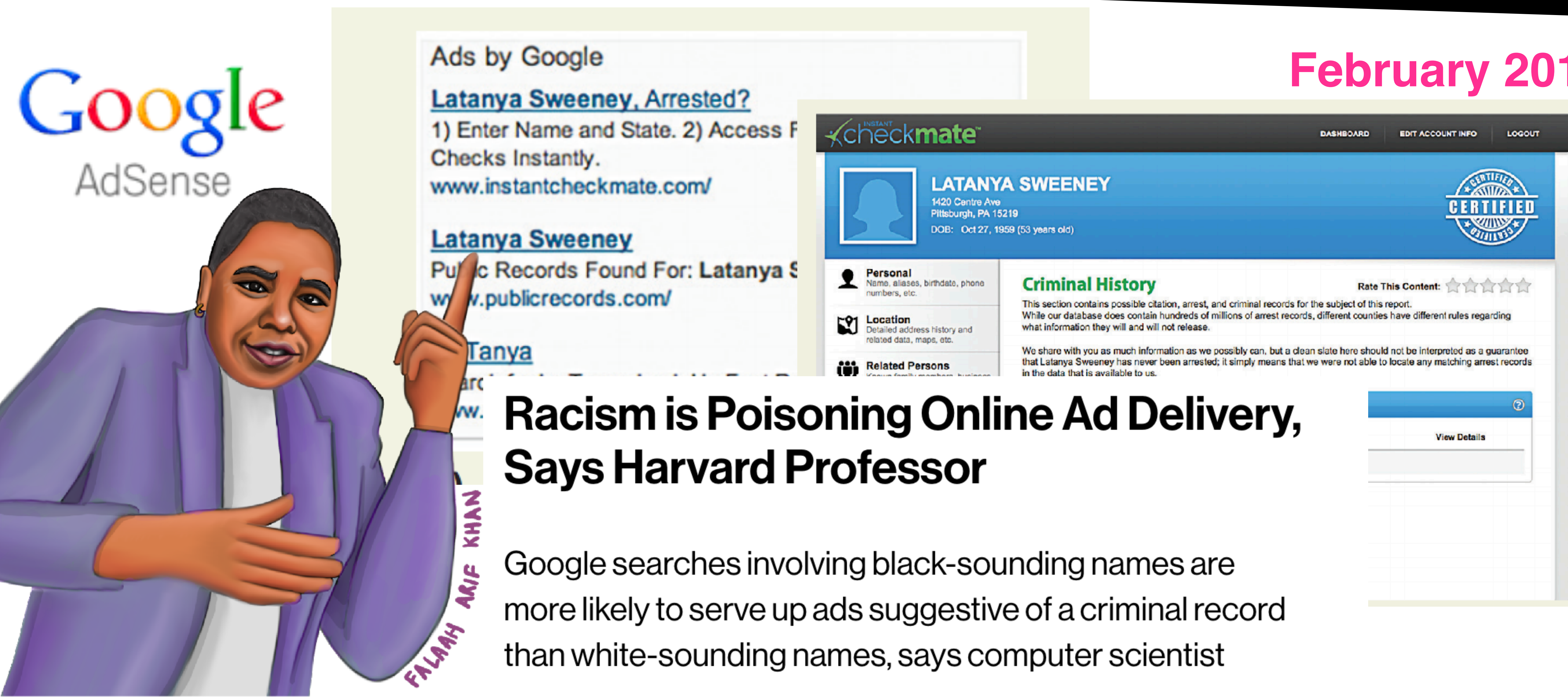


"Note: Amazon does not disclose the gender breakdown of its technical workforce."

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 self-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

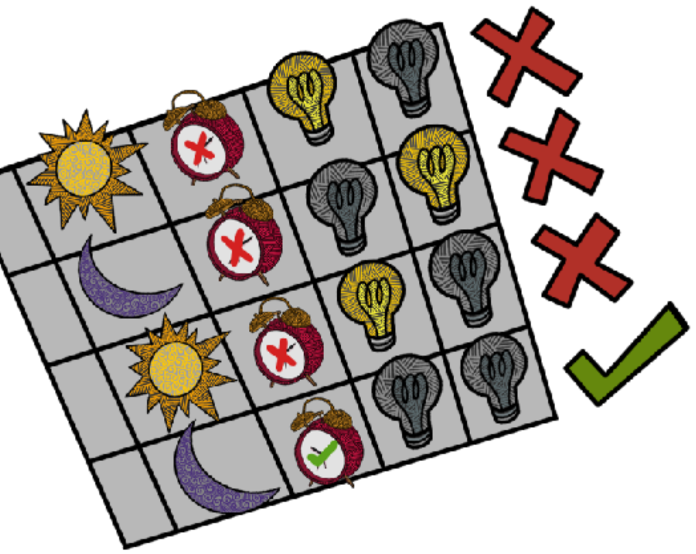
Amazon Doesn't Consider the Race of Its Customers. Should It?

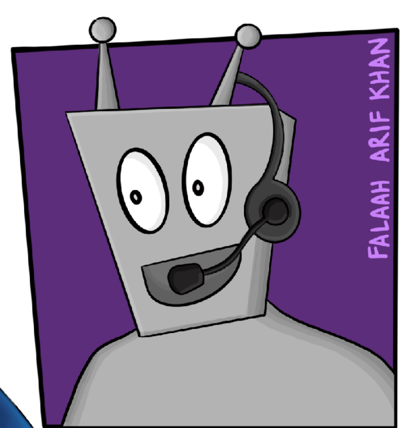
"The most striking gap in Amazon's same-day service is in Boston, where **three ZIP codes encompassing the primarily black neighborhood of Roxbury are excluded** from same-day service, while the neighborhoods that surround it on all sides are eligible."

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

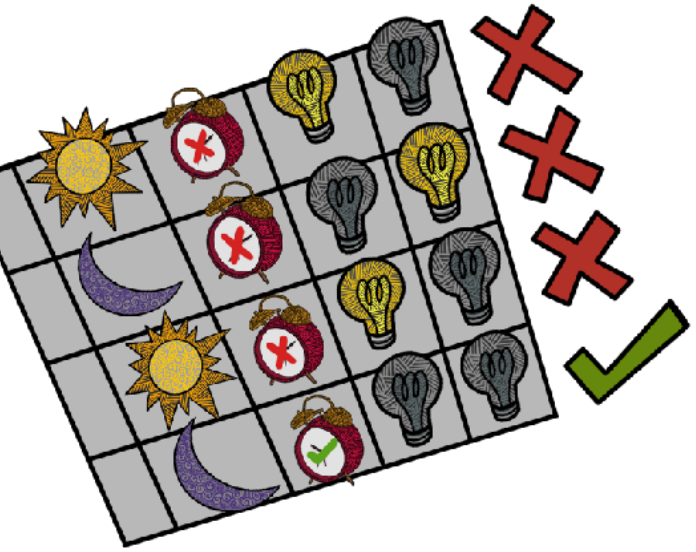
data science
ethics teaser

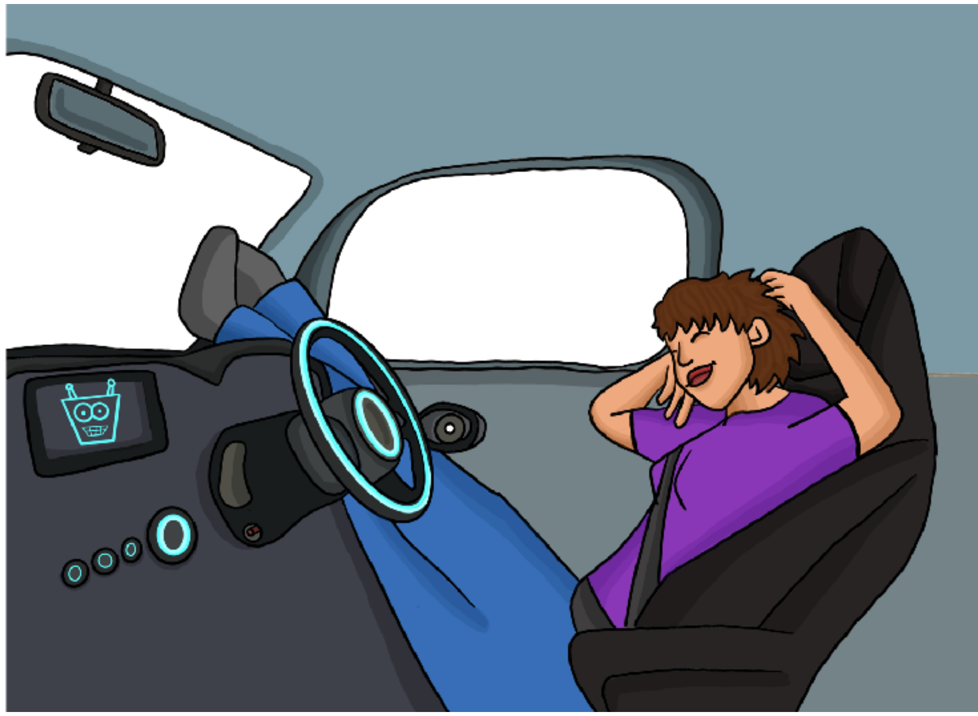
Mistakes lead to harms

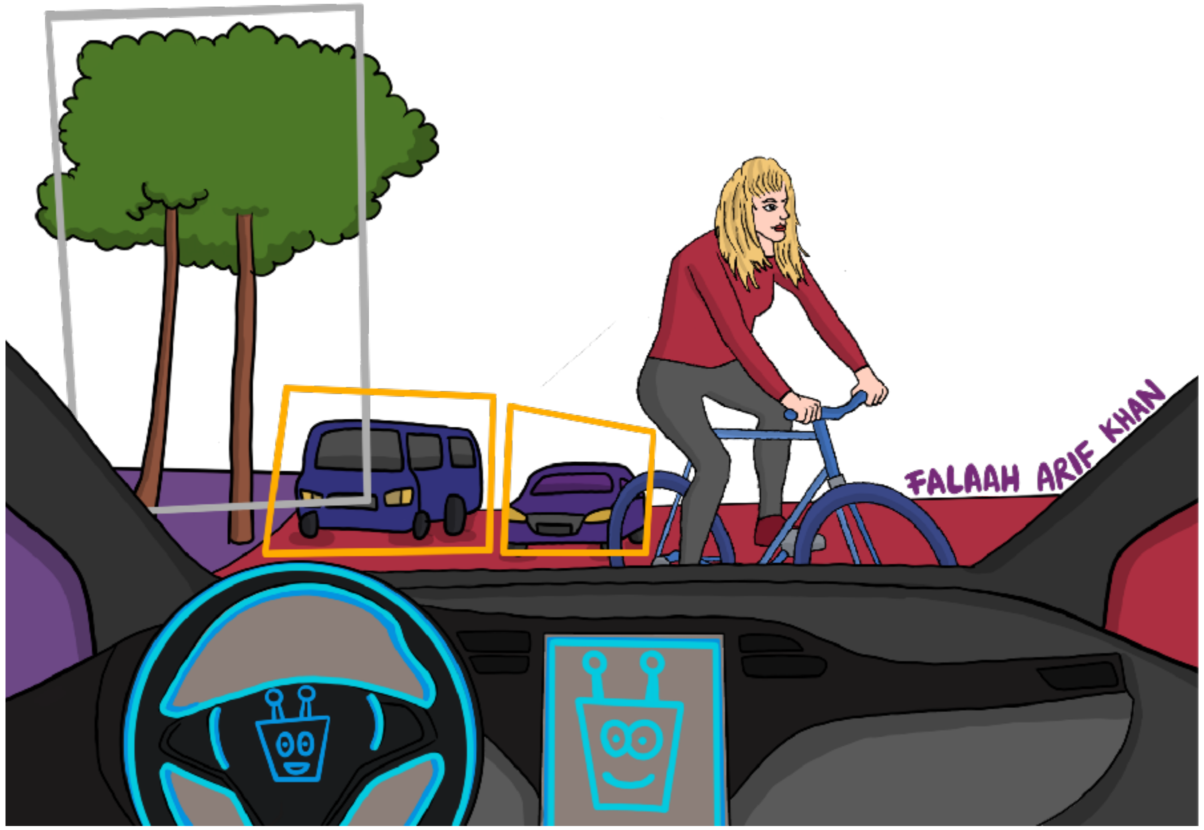




Mistakes lead to harms

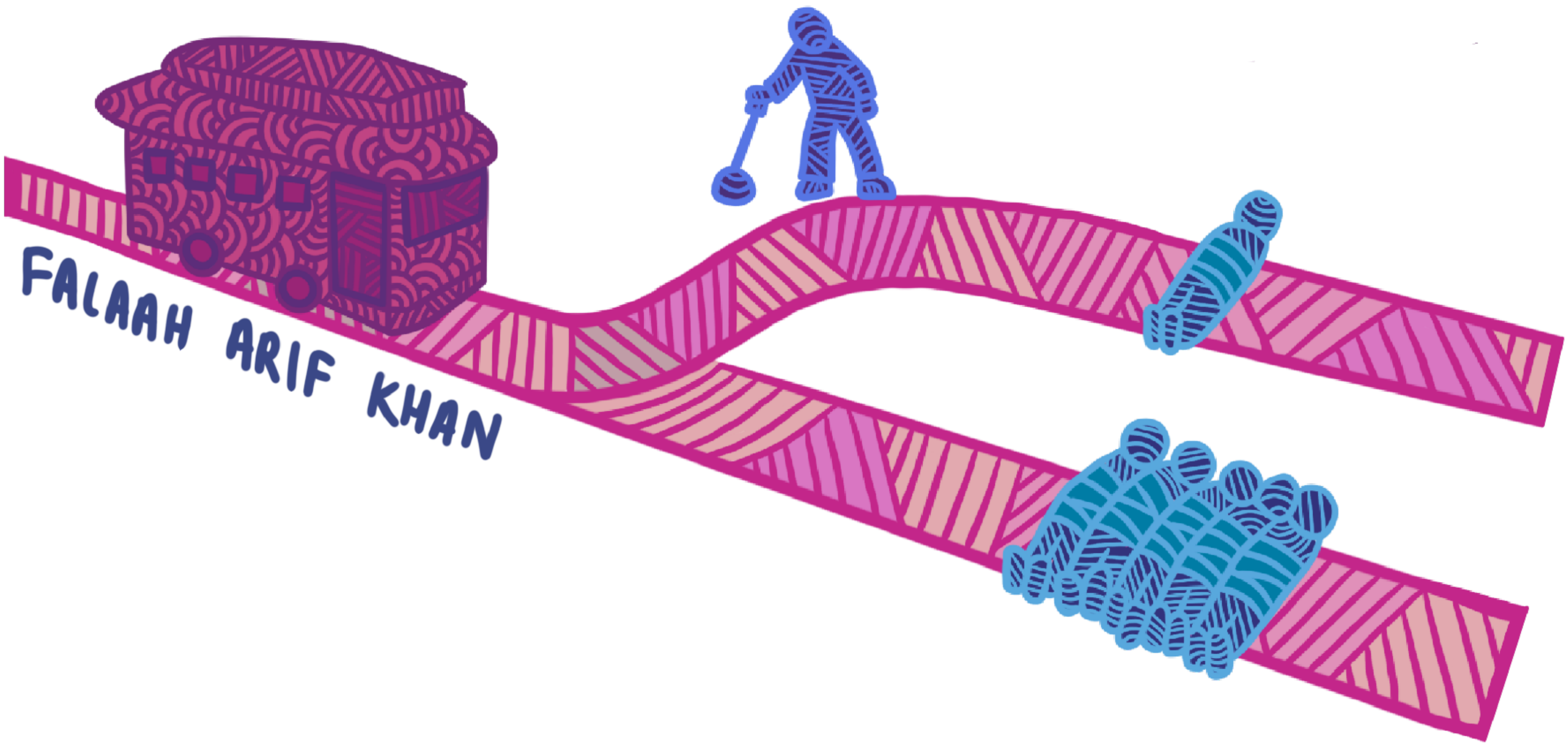






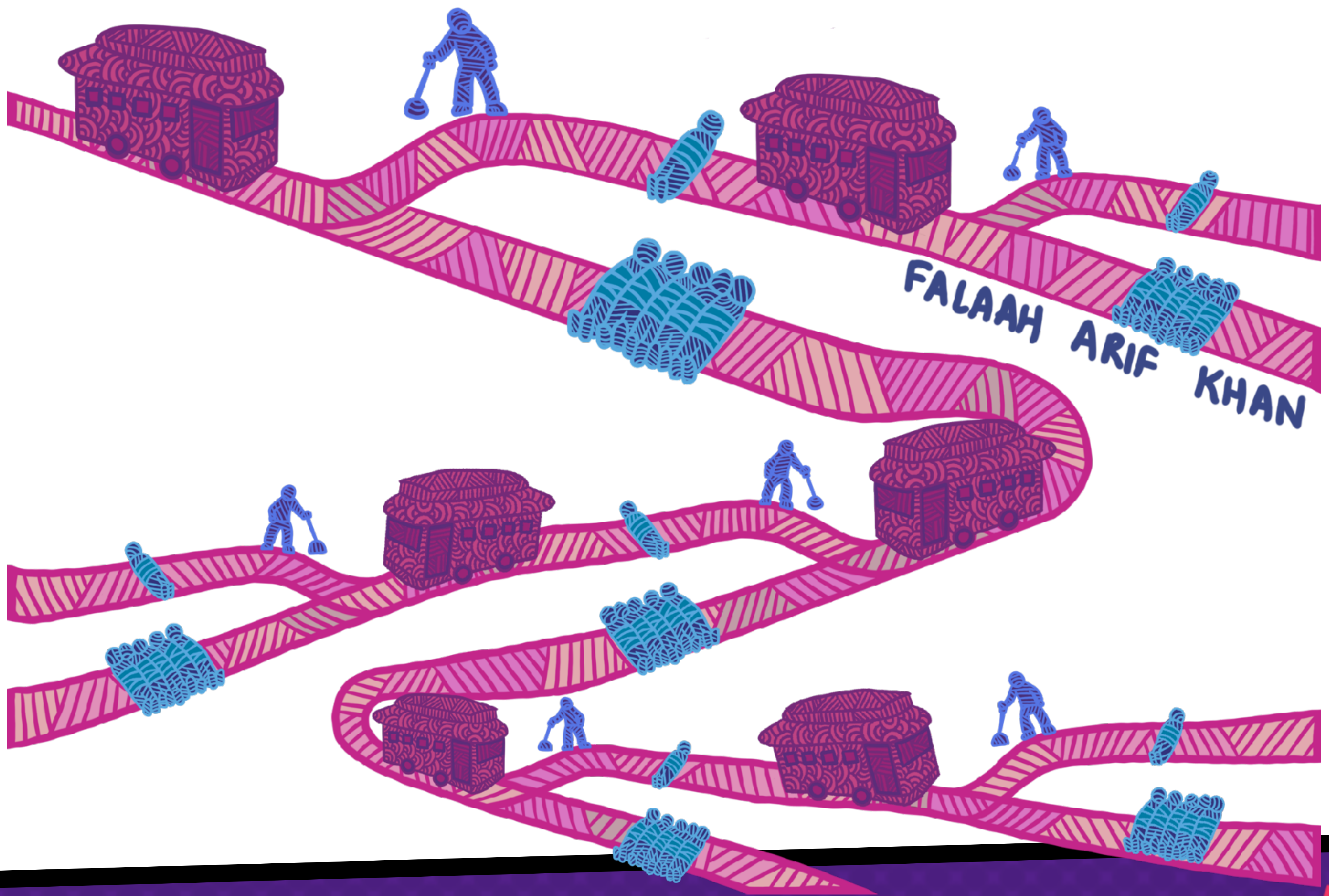
r/ai

The trolley problem



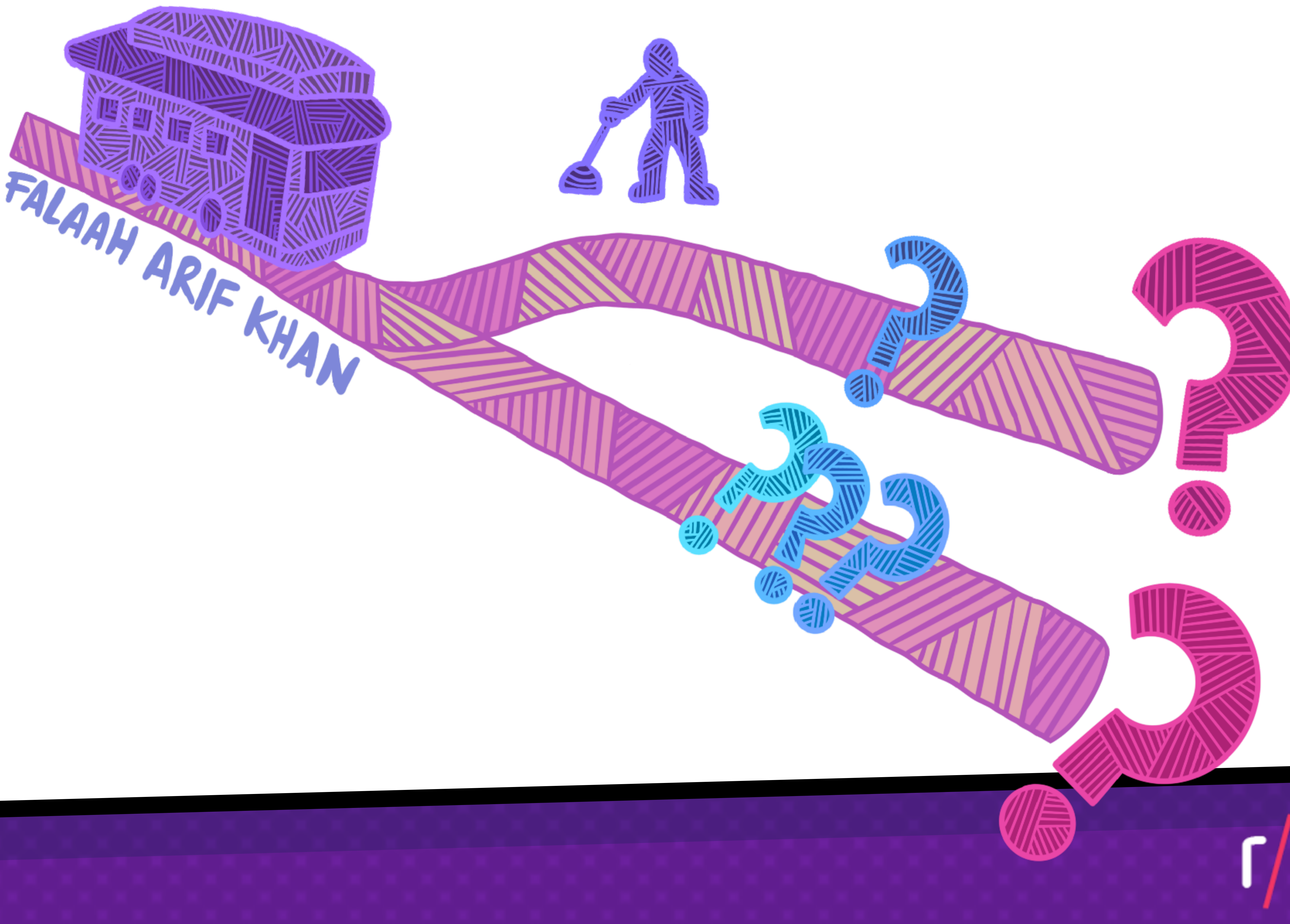
r/ai

The trolley problem



r/ai

Dealing with uncertainty

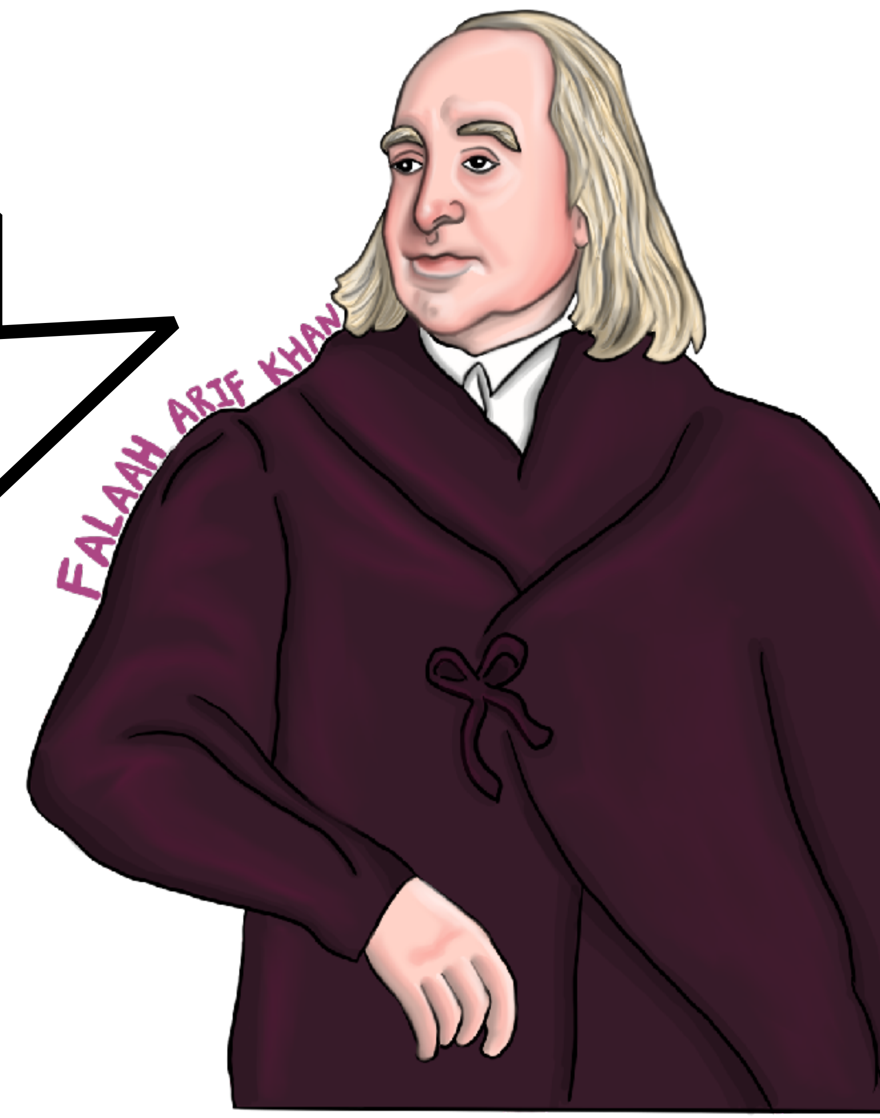


r/ai

Utilitarianism

"It is the greatest happiness of the greatest number that is the measure of right and wrong."

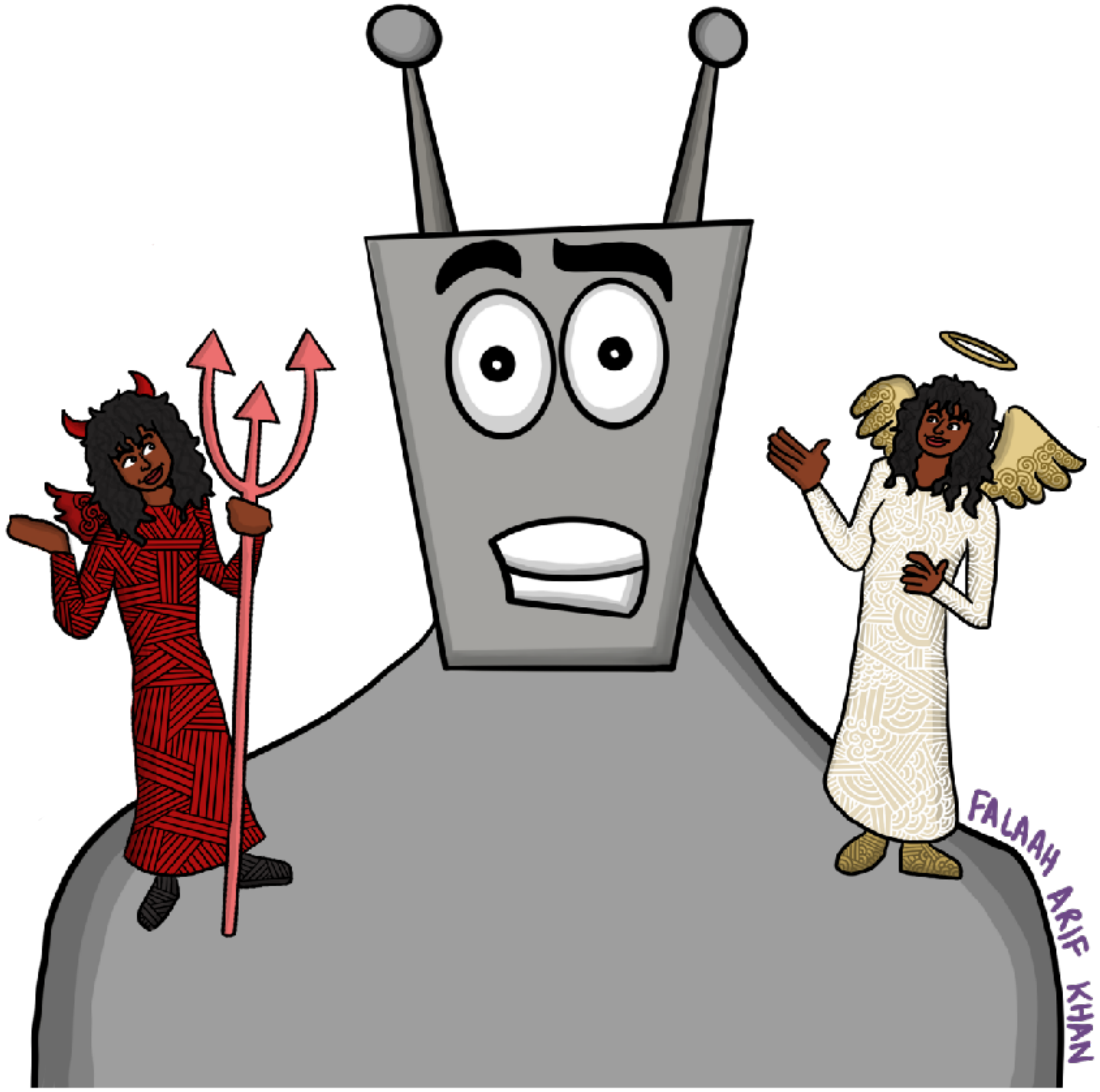
Jeremy Bentham



Algorithmic morality?

Algorithmic morality

is the act of attributing moral reasoning to algorithmic systems



r/ai

Algorithmic morality?

r/ai

Tech rooted in people

r/ai

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.**

Black people are almost twice as likely as white people 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 black people 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

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

course overview

module 1:

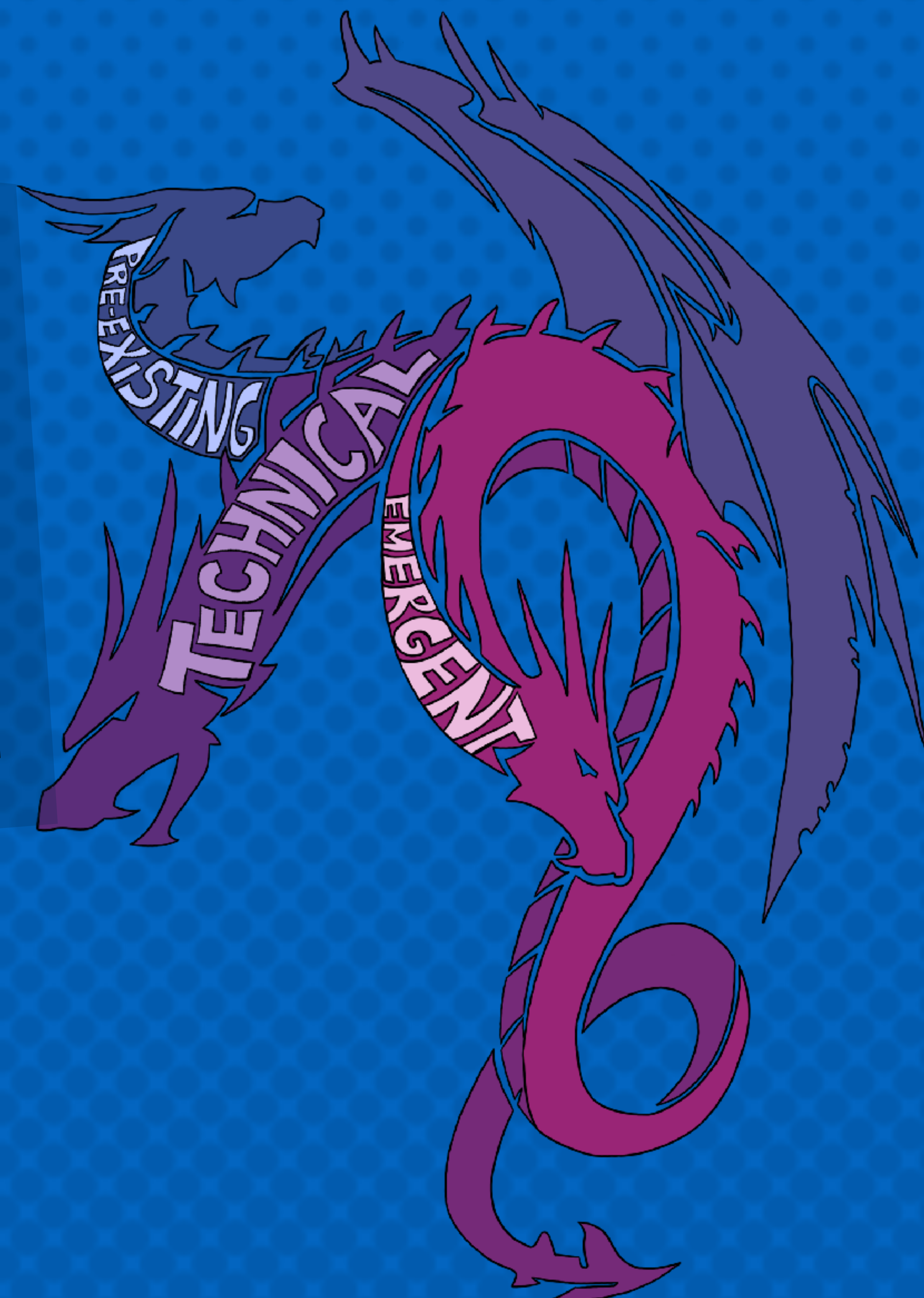
algorithmic fairness

Bias in computer systems

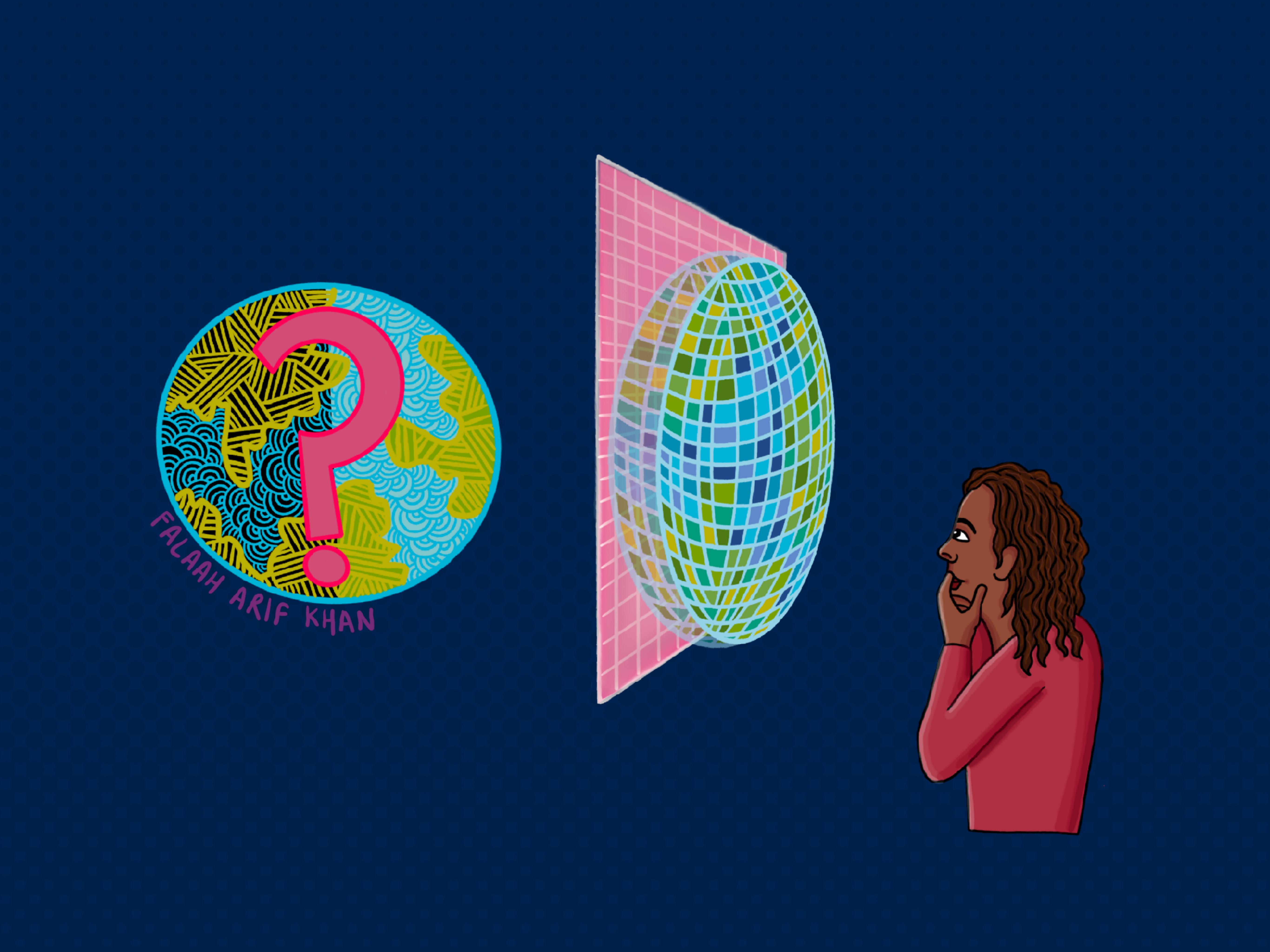
Pre-existing: exists independently of algorithm, has origins in society

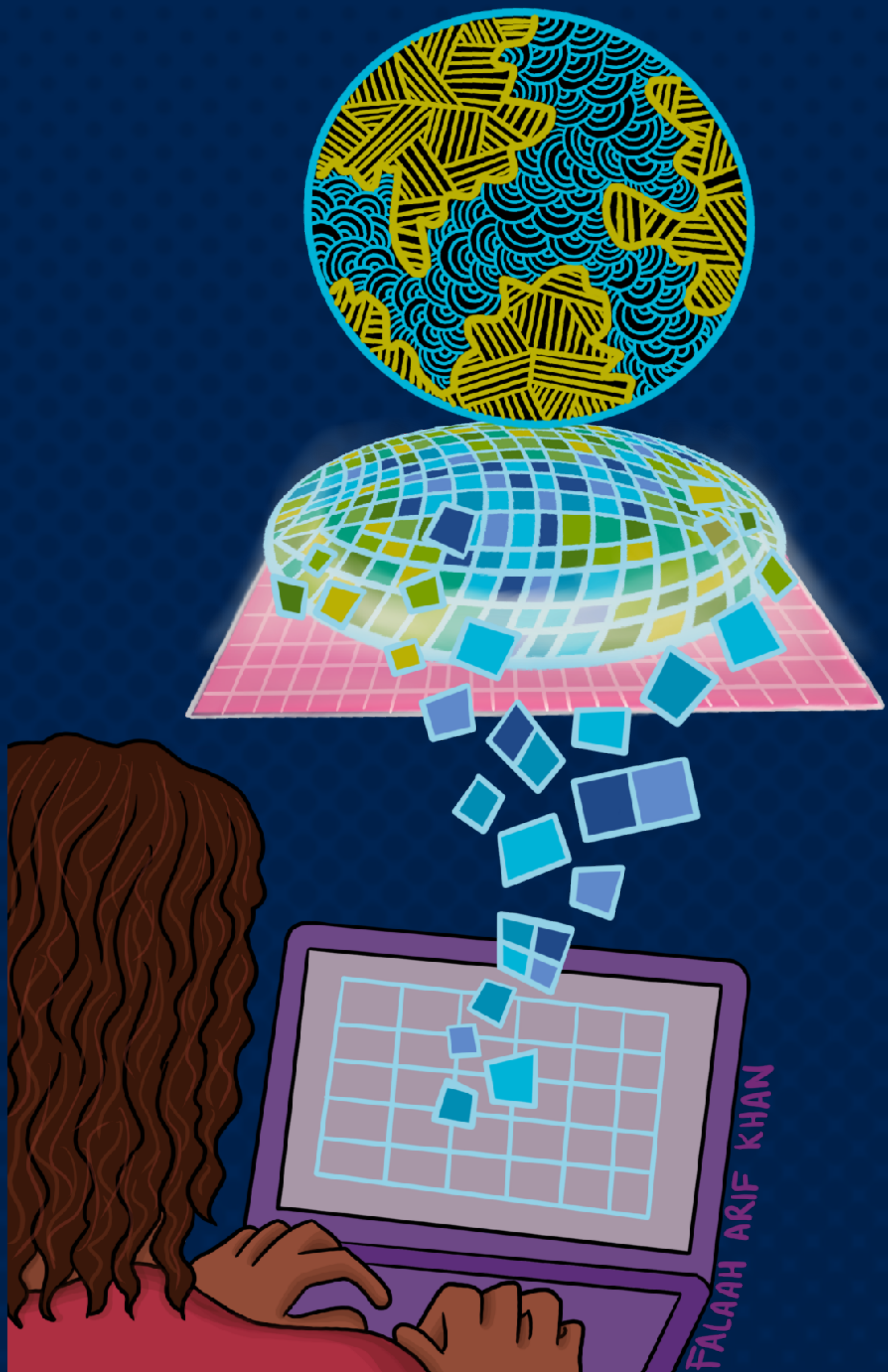
Technical: introduced or exacerbated by the technical properties of an ADS

Emergent: arises due to context of use

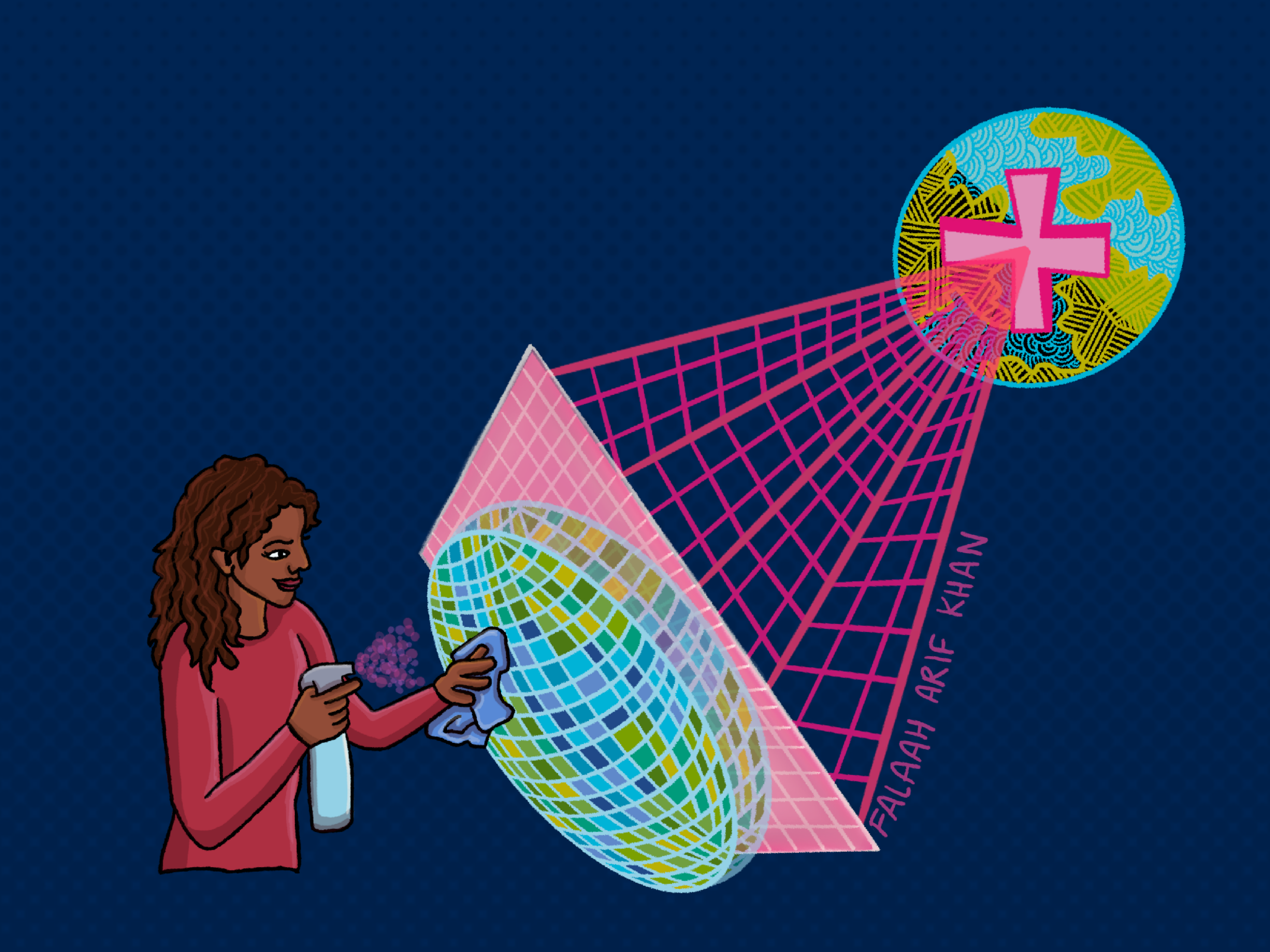


[Friedman & Nissenbaum (1996)]


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

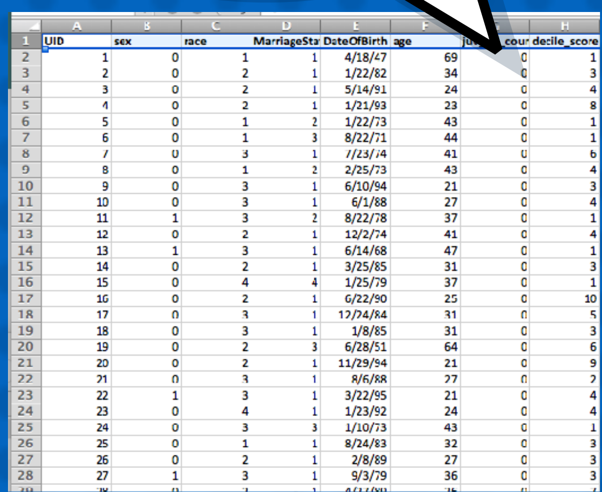
Emergent: arises due to context of use

to fight bias, state beliefs and assumptions explicitly

[Friedman & Nissenbaum (1996)]

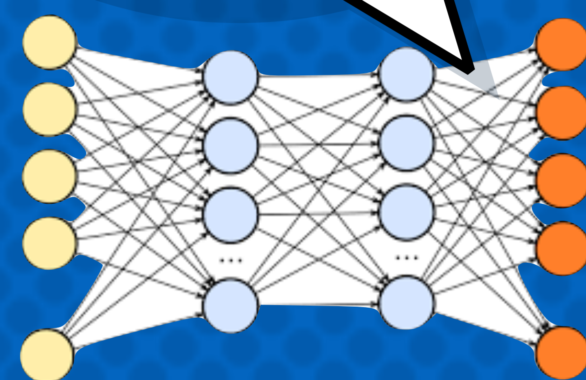
Fair-ML view

where did the data come from?

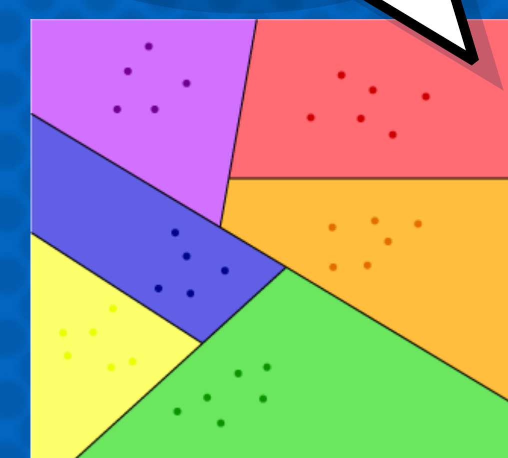


UID	sex	race	MarriageSta	DateOfBirth	age	ju[illegible]_cour	decile_score
1	0	1	1	4/18/47	69	0	1
2	0	2	1	1/22/82	34	0	3
3	0	2	1	5/14/91	24	0	4
4	0	2	1	1/21/93	23	0	8
5	0	1	2	1/22/73	43	0	1
6	0	1	3	8/22/71	44	0	1
7	0	3	1	7/23/74	41	0	6
8	0	1	2	2/25/73	43	0	4
9	0	3	1	6/10/94	21	0	3
10	0	3	1	6/1/88	27	0	4
11	1	3	2	8/22/78	37	0	1
12	0	2	1	12/2/74	41	0	4
13	1	3	1	6/14/68	47	0	1
14	0	2	1	3/25/85	31	0	3
15	0	4	4	1/25/79	37	0	1
16	0	2	1	6/22/90	25	0	10
17	0	3	1	12/24/84	31	0	5
18	0	3	1	1/8/85	31	0	3
19	0	2	3	6/28/51	64	0	6
20	0	2	1	11/29/94	21	0	9
21	0	3	1	8/6/88	27	0	2
22	1	3	1	3/22/95	21	0	4
23	0	4	1	1/23/92	24	0	4
24	0	3	3	1/10/73	43	0	1
25	0	1	1	8/24/83	32	0	3
26	0	2	1	2/8/89	27	0	3
27	1	3	1	9/3/79	36	0	3

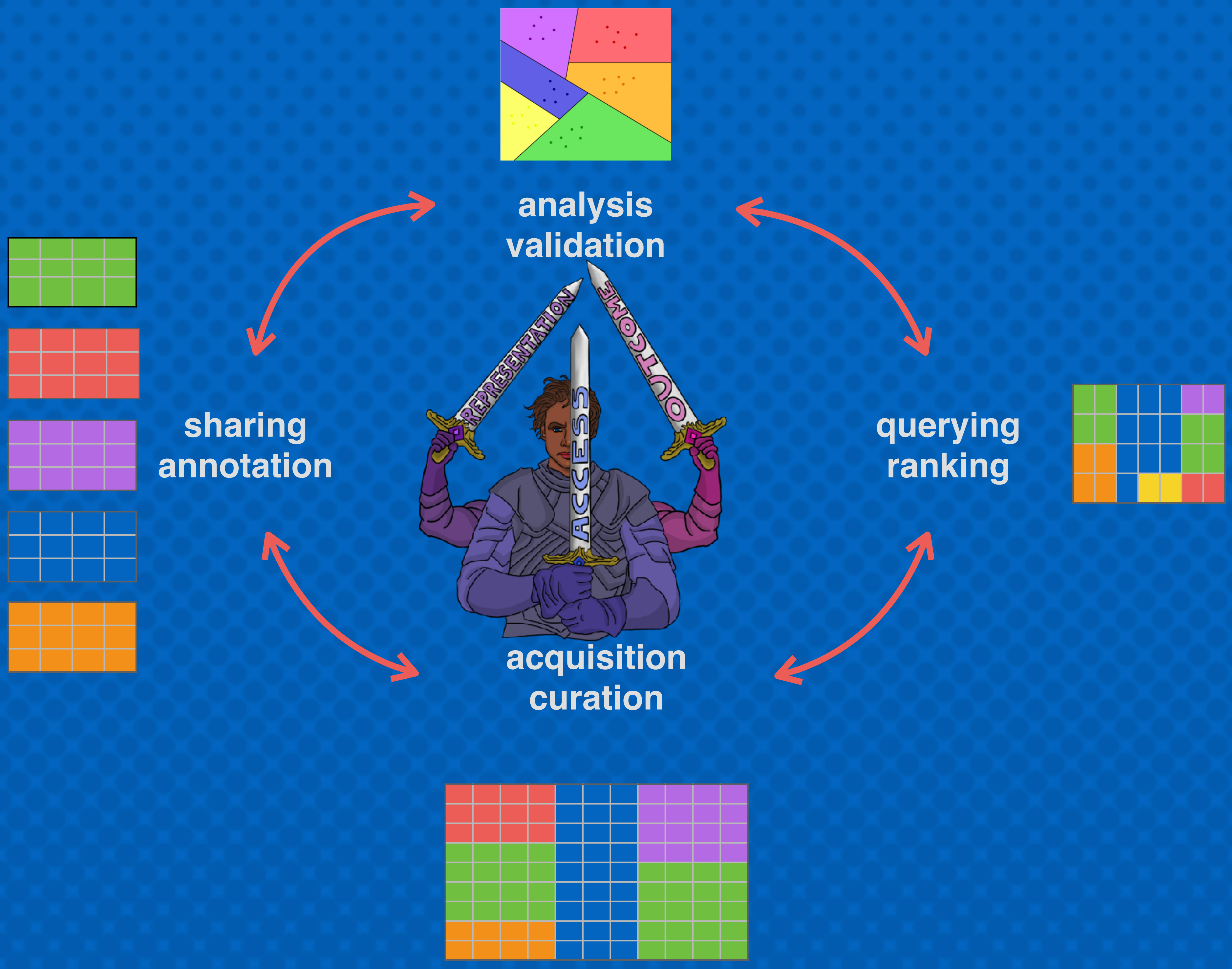
what happens inside the box?



how are results used?



Lifecycle view



[Stoyanovich, Howe, Jagadish (2020)]

Models and assumptions

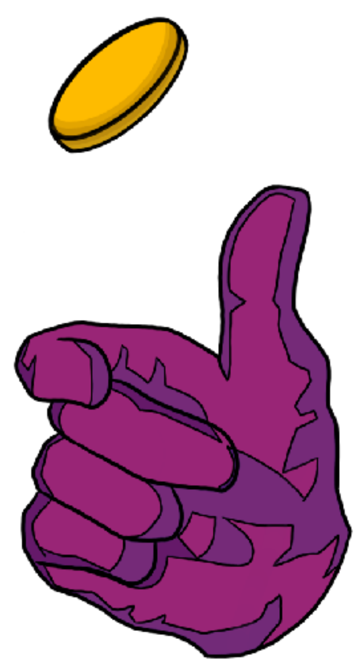
[Stoyanovich, Howe, Jagadish (2020)]

module 3:
data protection & privacy

Privacy: two sides of the same coin

Did you go out drinking over the weekend?

protecting an individual / plausible deniability



learning about the population / 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

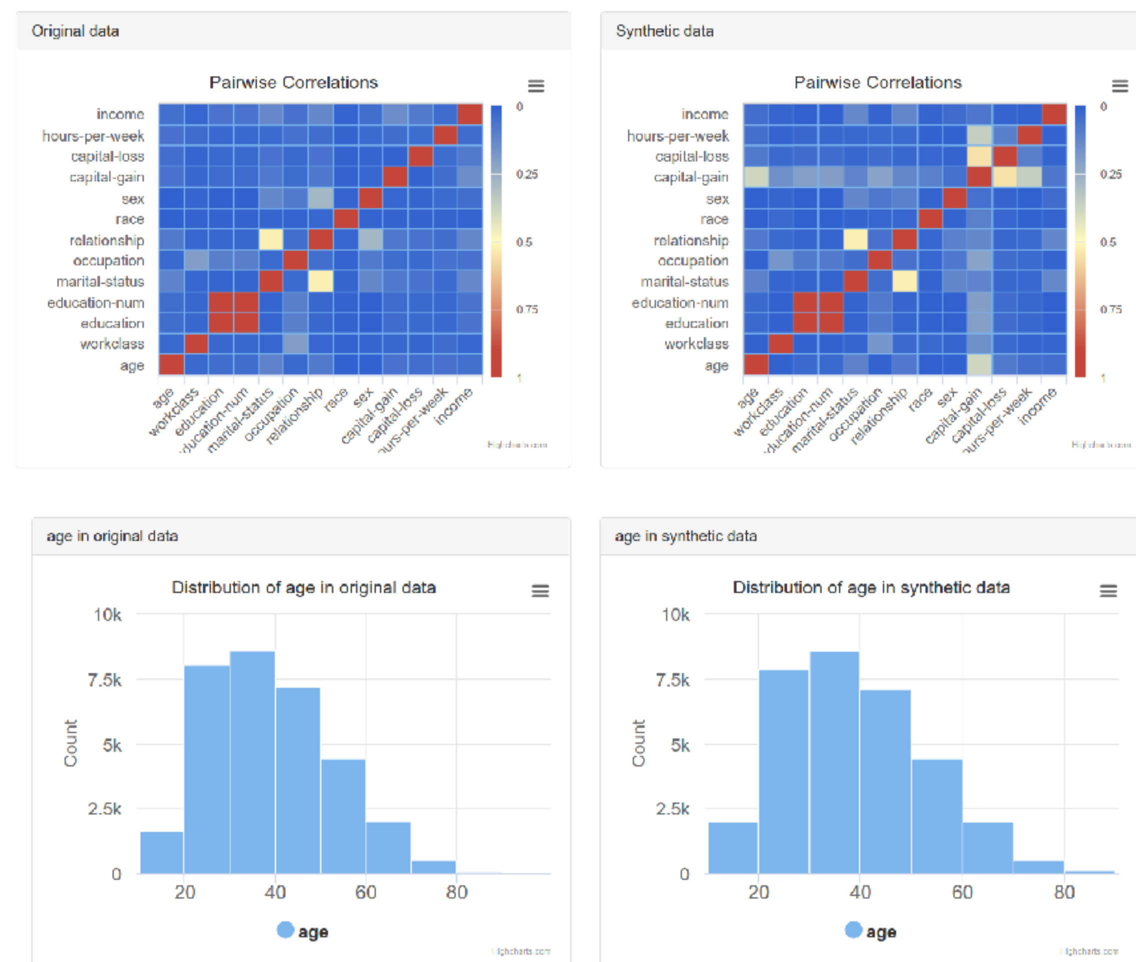
What does it mean to preserve privacy?

BY CYNTHIA DWORK

A Firm Foundation for Private Data Analysis

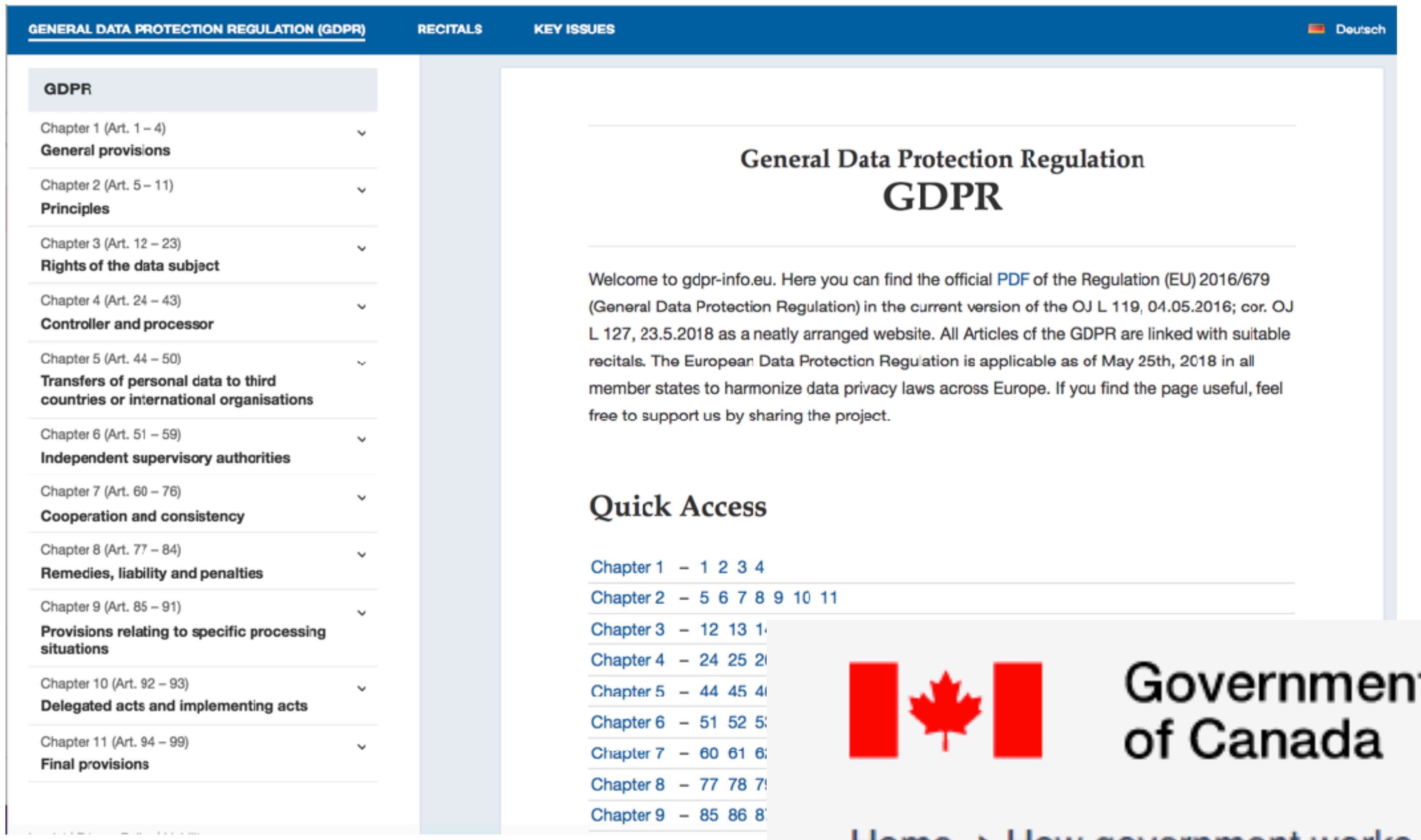
Communications of the ACM CACM Homepage archive

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

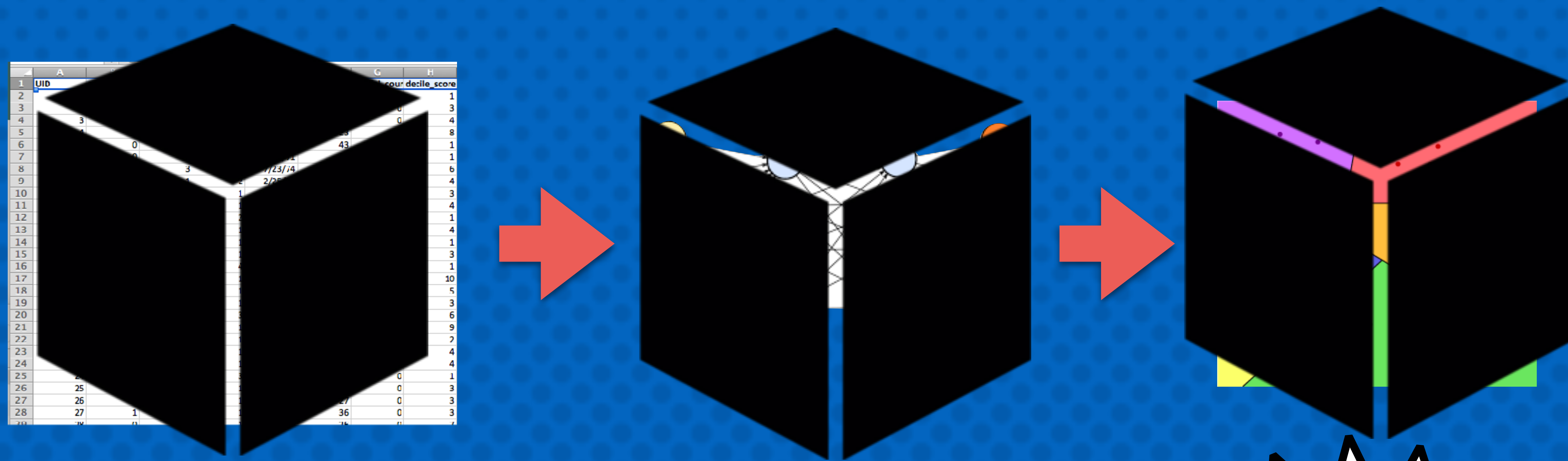
Disparate treatment

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

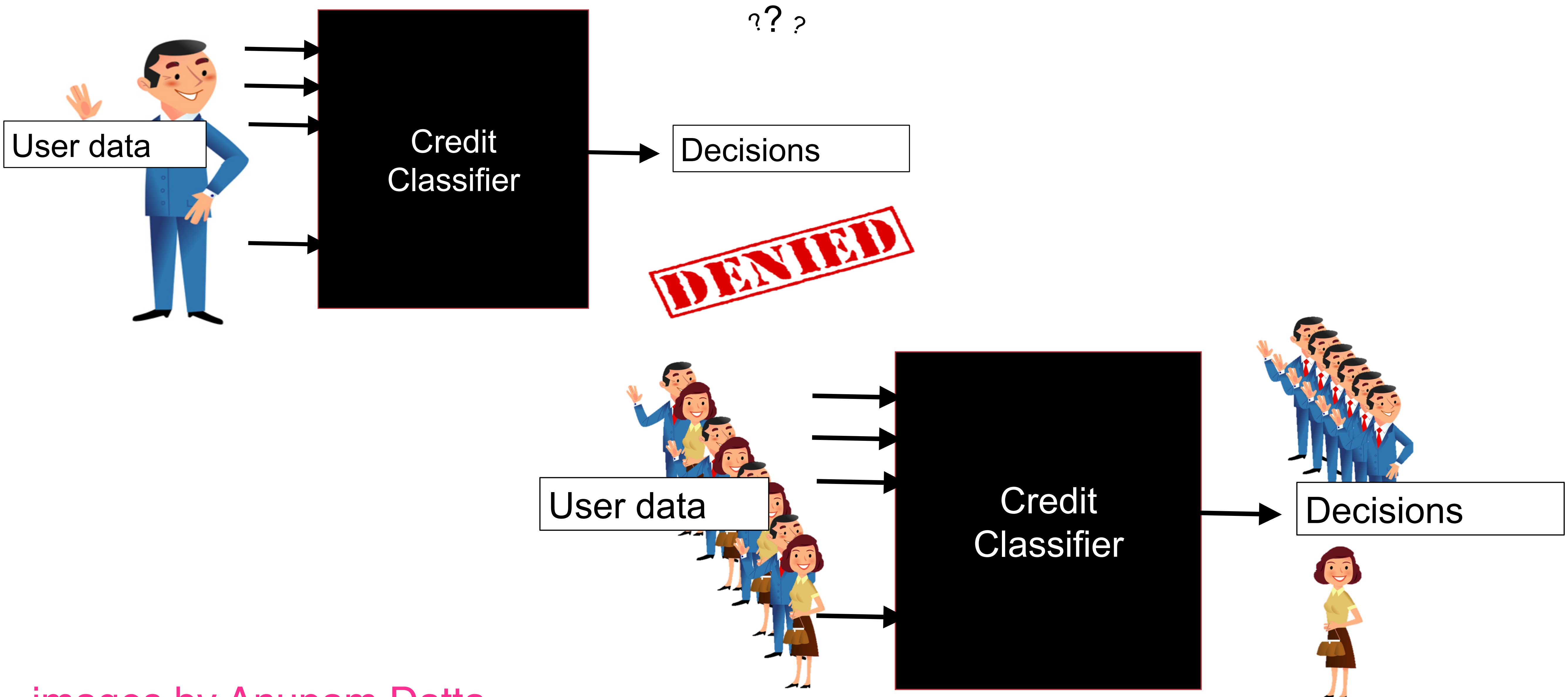


Fair Housing Act

Equal Credit Opportunity Act, 1964

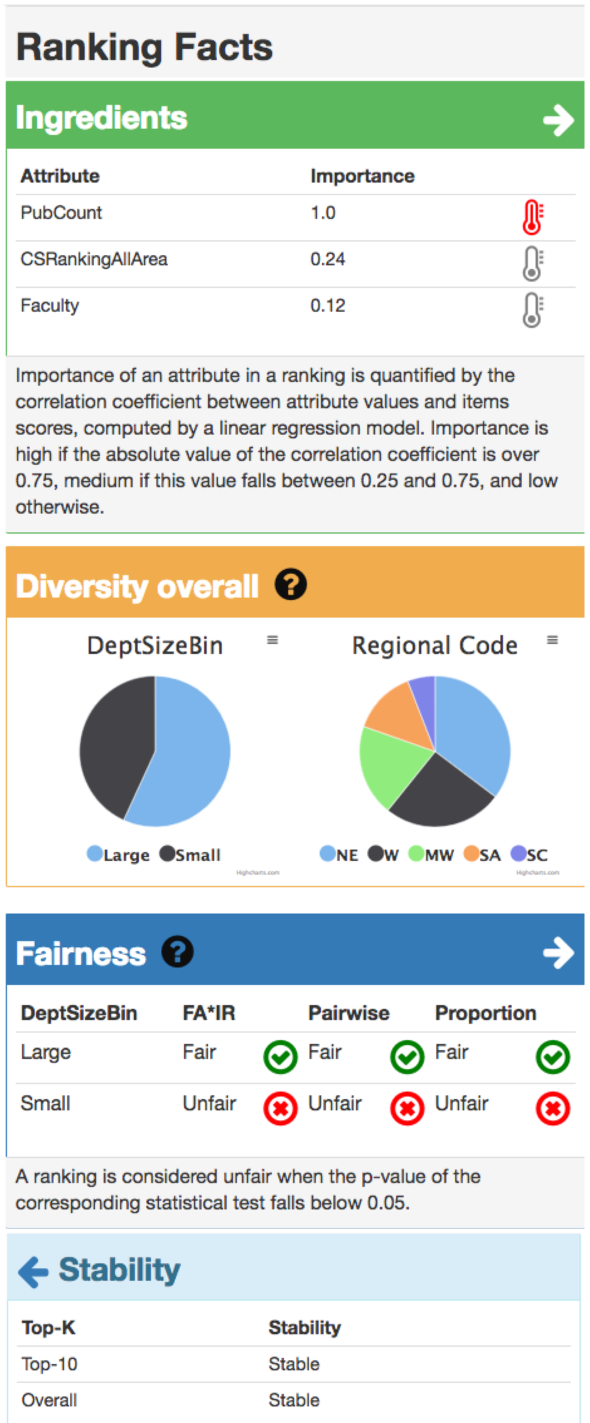
Civil Rights Act, 1964

Auditing black-box models



images by Anupam Datta

Nutritional labels



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. security and privacy

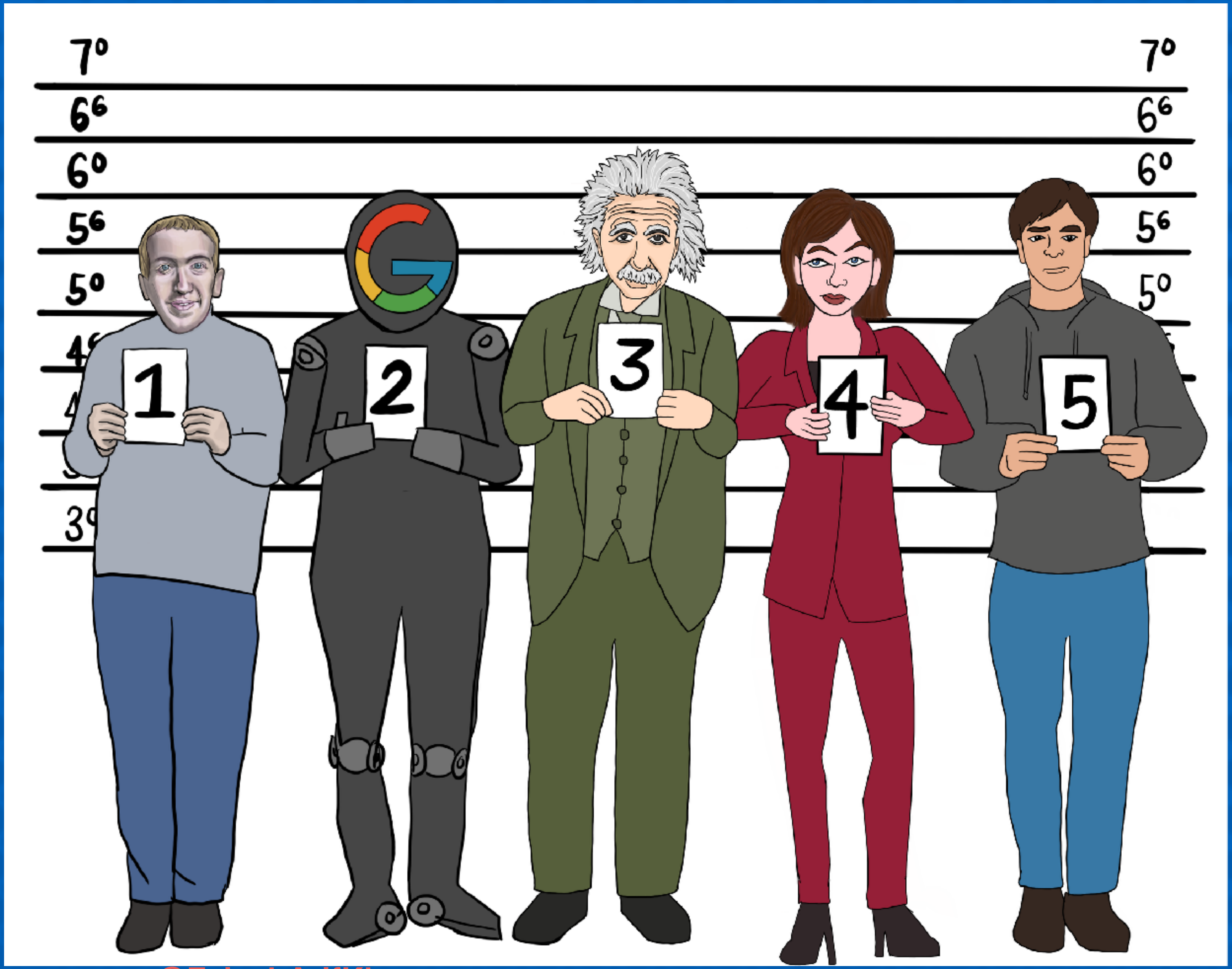
We will learn **algorithmic techniques** for data analysis.
We will also learn about recent **laws** / **regulatory frameworks**.

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!