Responsible Data Science

Introduction and Overview

Prof. Julia Stoyanovich

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

data RESPONSIBLY

NYU

course logistics

Instructor: Julia Stoyanovich

Assistant Prof. of Data Science at the Center for Data Science
Assistant Prof. of Computer Science & Engineering at Tandon
Director, Center for Responsible AI (R/AI)

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

Research: data and knowledge management ("databases")

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

@stoyanoj

@AIResponsibly

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

Instructor: George Wood

Moore Sloan Faculty Fellow at the Center for Data Science

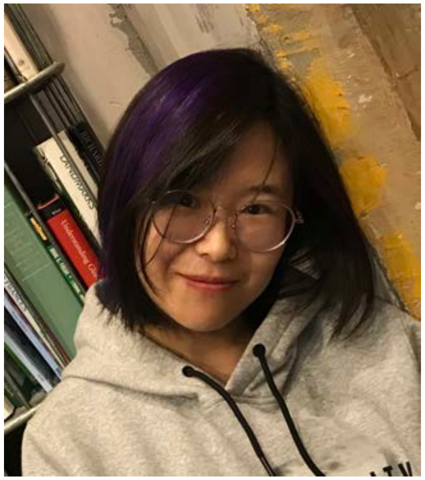
Ph.D. in Sociology from University of Oxford
B.S. in Sociology from University of Sheffield

My research examines inequalities in public health and criminal justice. As part of this work, I evaluate the effects of social programs and interventions that aim to reduce gunshot victimization, police misconduct, and police use of force. I also develop tools to enhance transparency and accountability in policing.

Office hours: Tuesdays 1-2pm EST and by appointment

DS-GA 1017 Course Staff

Section Leader: Prasanthi Gurumurthi
Office hours: Wednesdays 3-4pm and by appointmen



Grader: Nan Wu
Office hours: Fridays, 10-11am and by appointment

Grader: Evaristus Ezekwem
Office hours: Thursdays, 1-2pm and by appointment

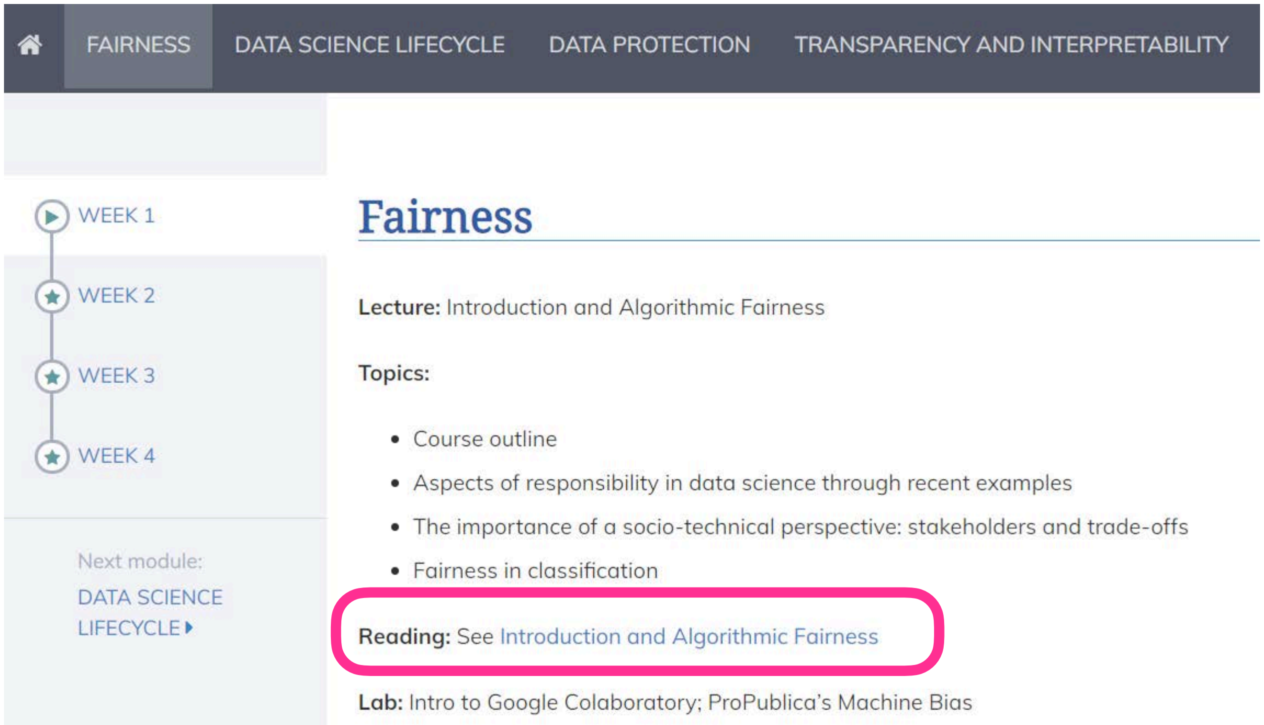
DS-UA 202 Course Staff

Section Leader / Grader: Apurva Bhargava
Office hours: Wednesdays 4-5pm and by appointment

Section Leader / Grader: Jeewon Ha
Office hours: Thursdays, 3-4pm and by appointment

Where to find information

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



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

Assignments and grading

Grading: homeworks - 10% x 3 = 30%
project - 30%
final exam - 30%
attendance & participation - 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 NYU Classes (under Overview), subject to change.

Meeting times

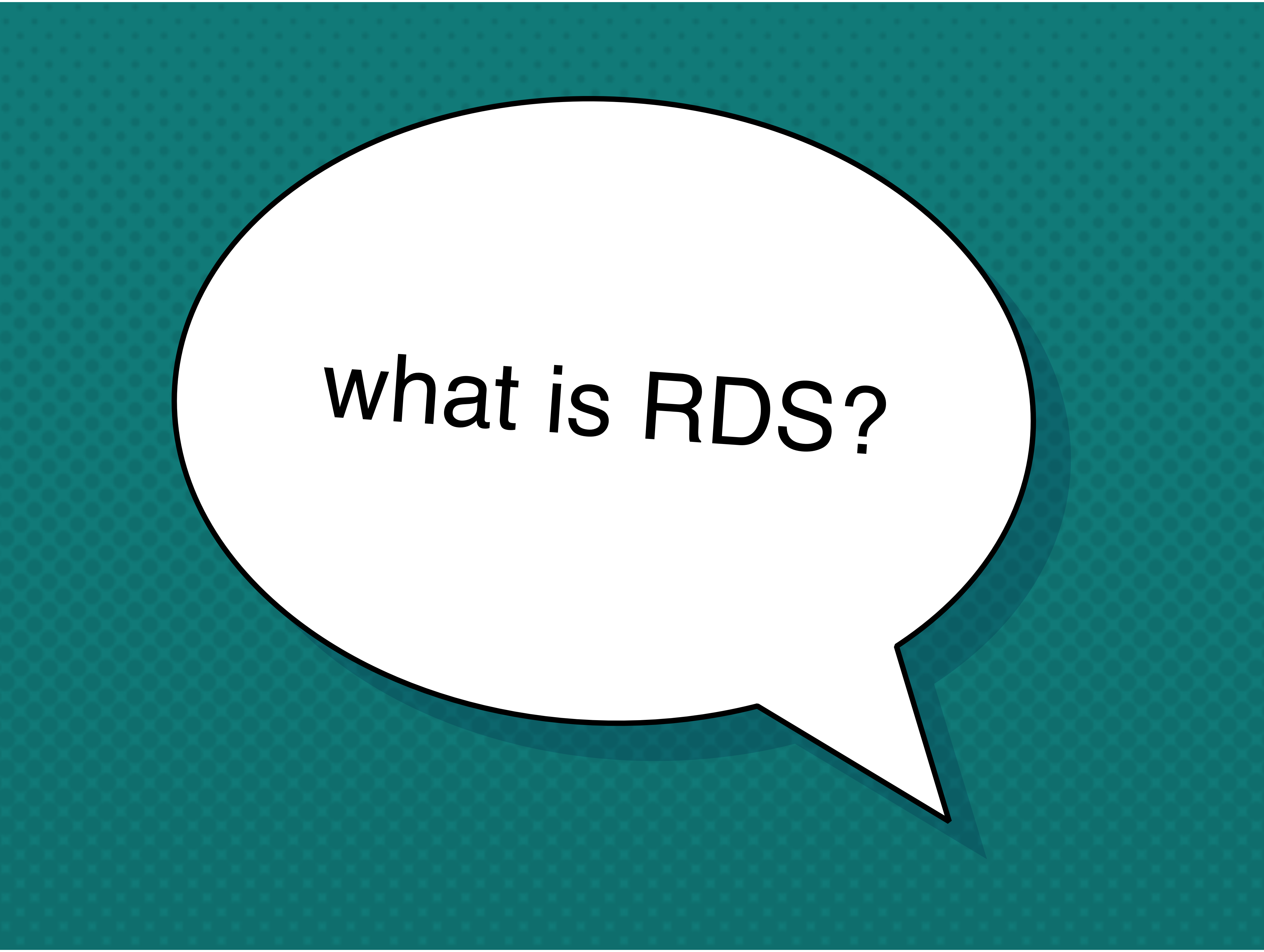
Meeting Times

DS-UA 202:

	Day	Time	Format
Lecture A	Tuesdays	9:30am – 12pm	Blended
Lab A	Wednesdays	9:30am – 10:20am	Online
Lab B	Wednesdays	10:25am – 11:15am	Blended

DS-GA 1017:

	Day	Time	Format
Lecture B	Mondays	9:15am – 10:55am	Blended
Lab C	Mondays	11:35am – 12:25pm	Blended
Lab D	Wednesdays	9:30am – 10:20am	Online


what is RDS?

The promise of “AI”

Power

unprecedented data collection

enormous computational power

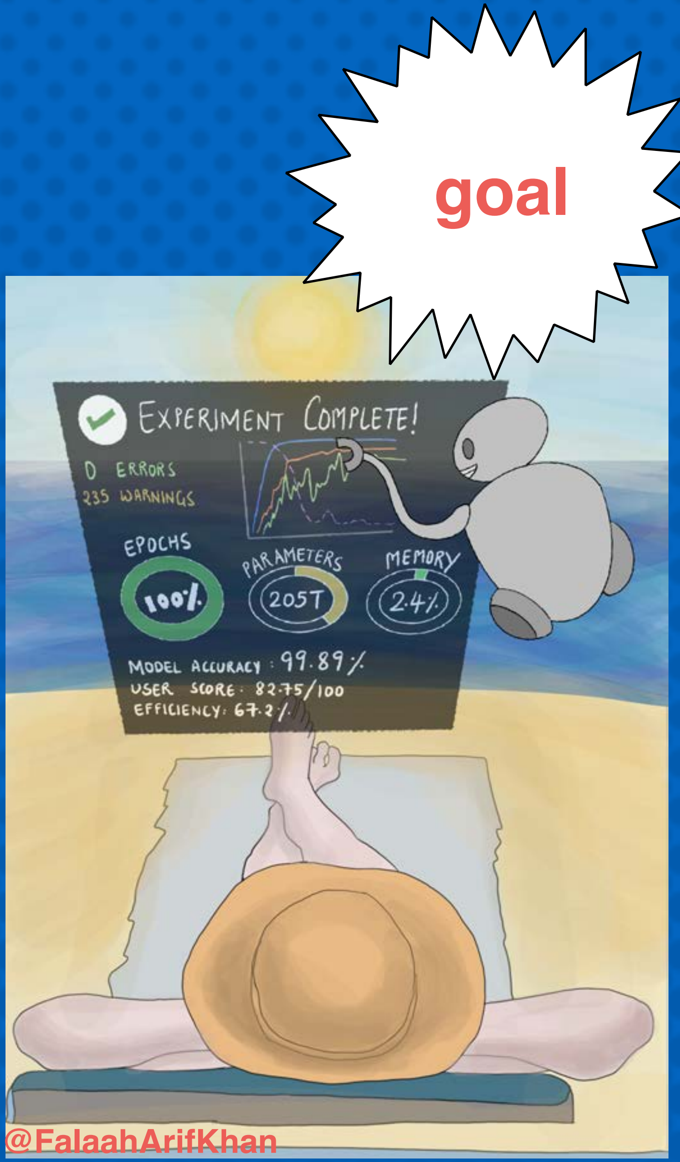
ubiquity and broad acceptance

Opportunity

accelerate science

boost innovation

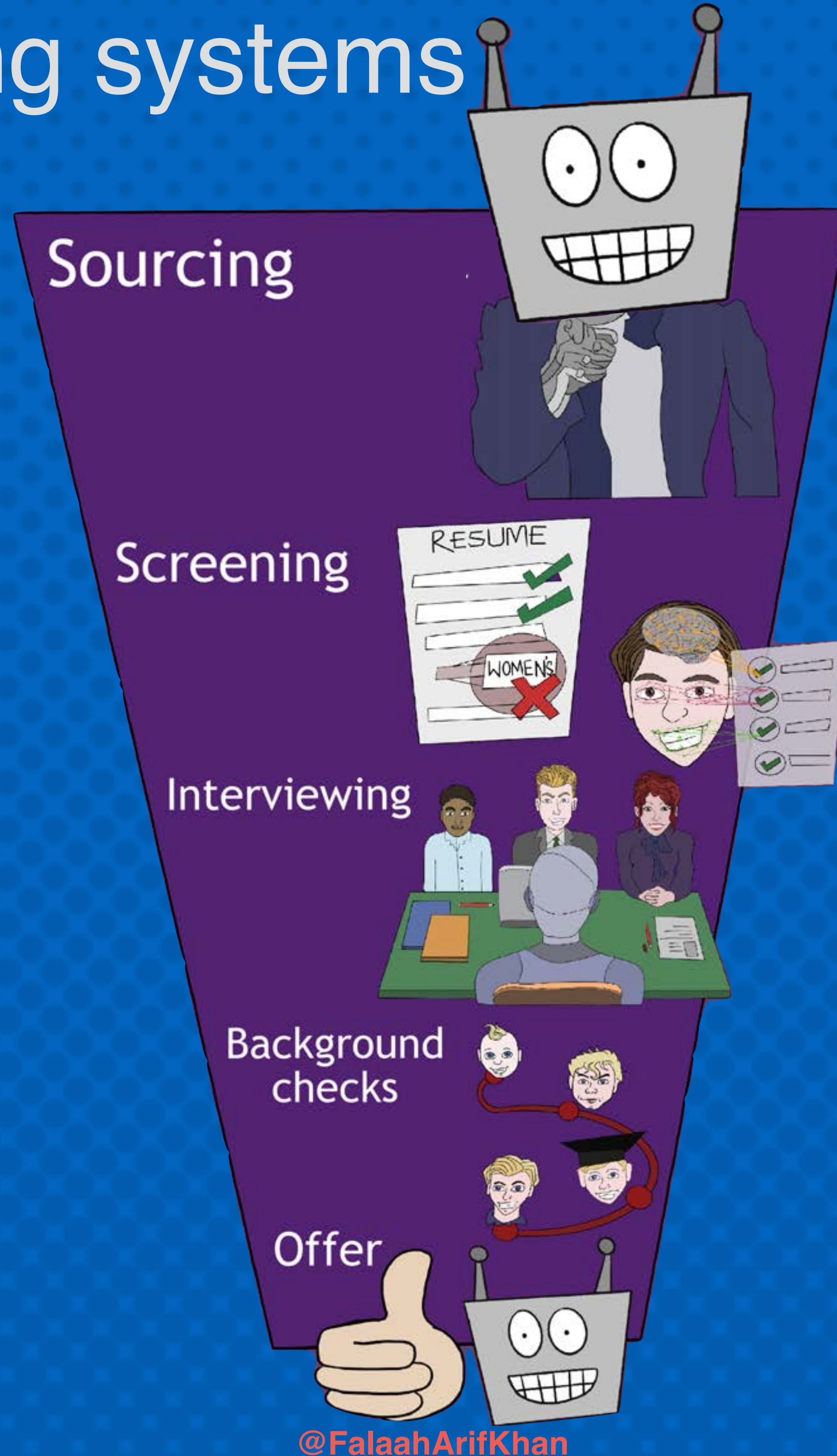
transform government



Automated hiring systems

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

Jenny Yang



and now…
some bad news

Online job ads

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

Gender bias in recruiting

REUTERS Jeffrey Dastin

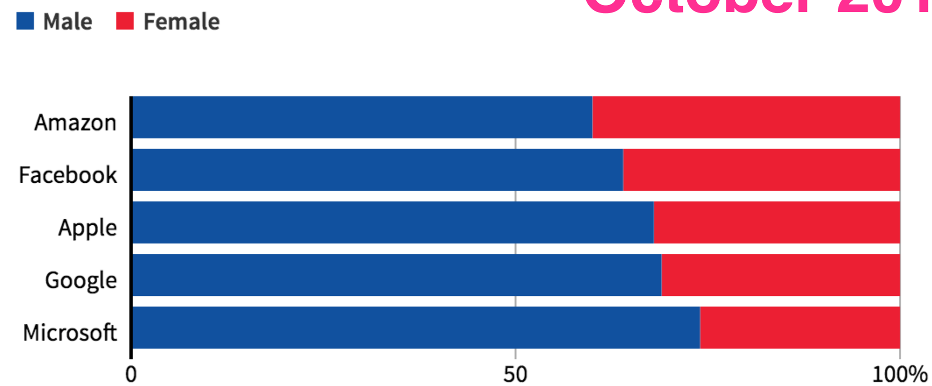
BUSINESS NEWS OCTOBER 9, 2018 / 11:12 PM / 6 MONTHS AGO

Amazon scraps secret AI recruiting tool that showed bias against women

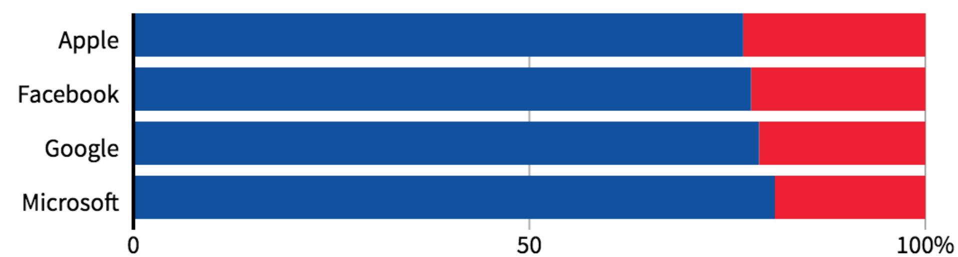
"In effect, **Amazon's system taught itself that male candidates were preferable**. It penalized resumes that included the word "women's," as in "women's chess club captain." And it **downgraded graduates of two all-women's colleges**, according to people familiar with the matter. They did not specify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

GLOBAL HEADCOUNT

October 2018



EMPLOYEES IN TECHNICAL ROLES



"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

Job-screening personality tests

THE WALL STREET JOURNAL.

September 2014

Are Workplace Personality Tests Fair?

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

Kyle Behm accused Kroger and six other companies of discrimination against the mentally ill through their use of personality tests. *TROY STAINS FOR THE WALL STREET JOURNAL*

The Equal Employment Opportunity Commission is **investigating whether personality tests discriminate against people with disabilities**.

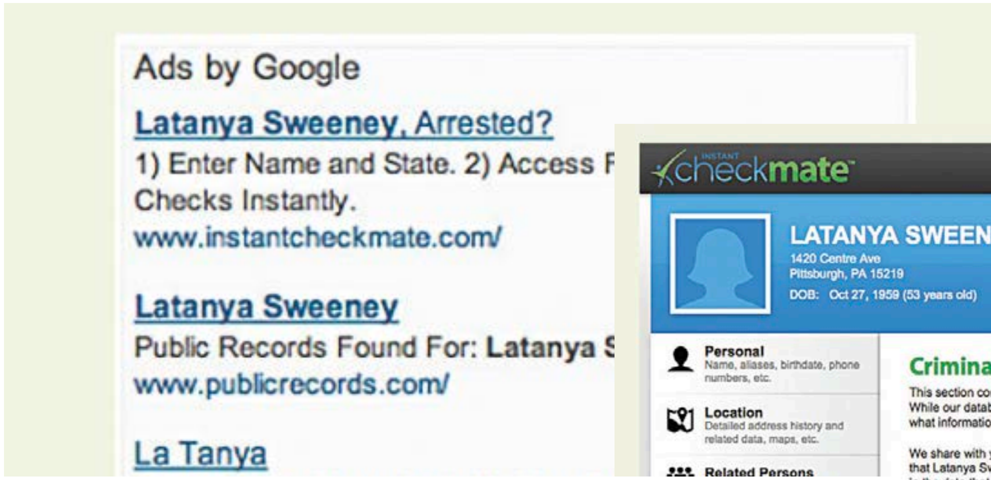
As part of the investigation, officials are trying to determine if the tests **shut out people suffering from mental illnesses** such as depression or bipolar disorder, even if they have the right skills for the job.

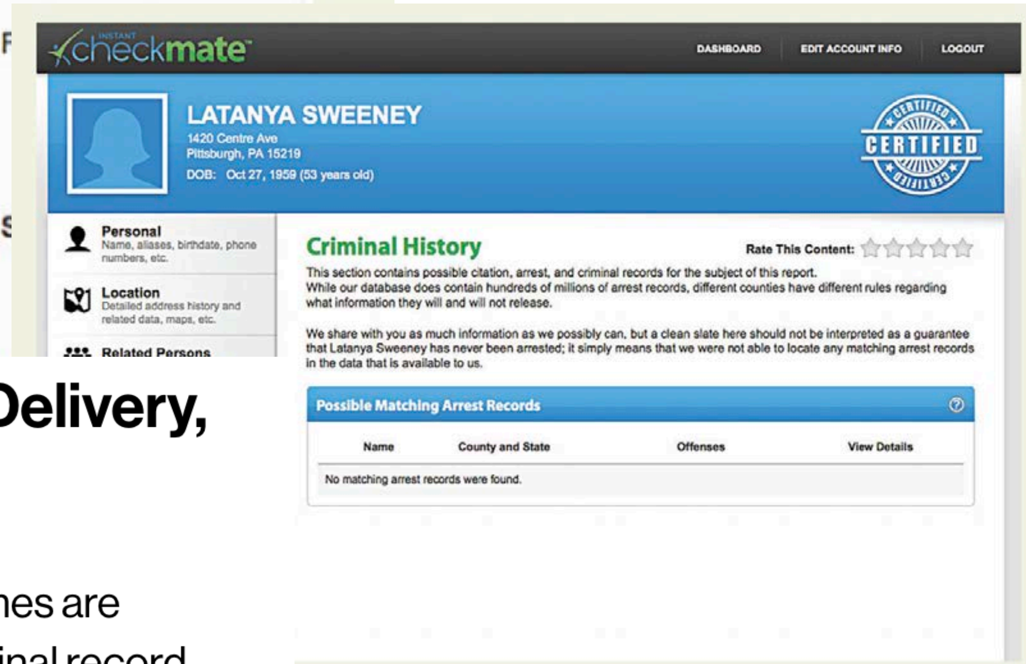
http://www.wsj.com/articles/are-workplace-personality-tests-fair-1412044257

Racially identifying names

February 2013

Google
AdSense





Racism is Poisoning Online Ad Delivery, Says Harvard Professor

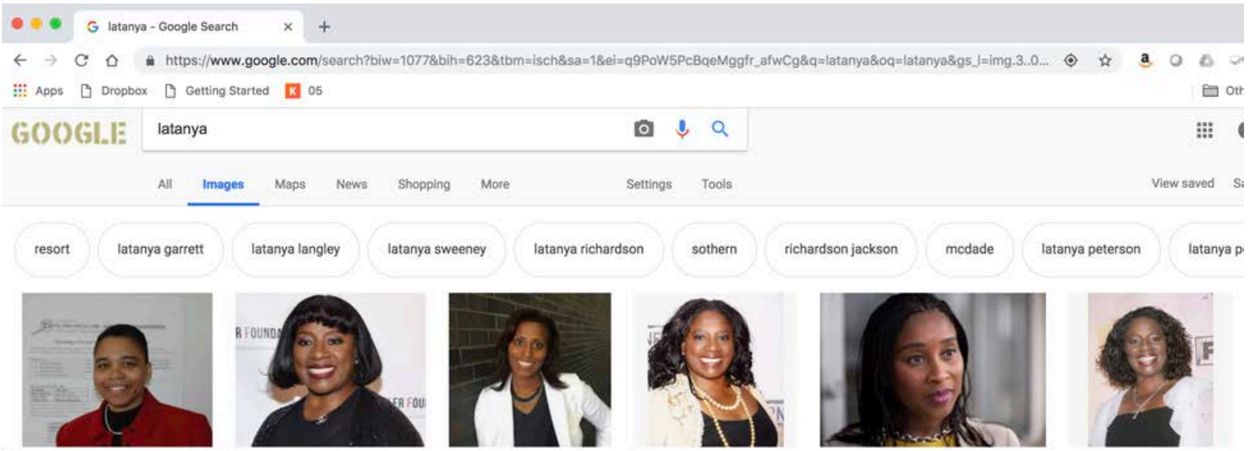
Google searches involving black-sounding names are more likely to serve up ads suggestive of a criminal record than white-sounding names, says computer scientist

racially identifying names trigger ads suggestive of a criminal record

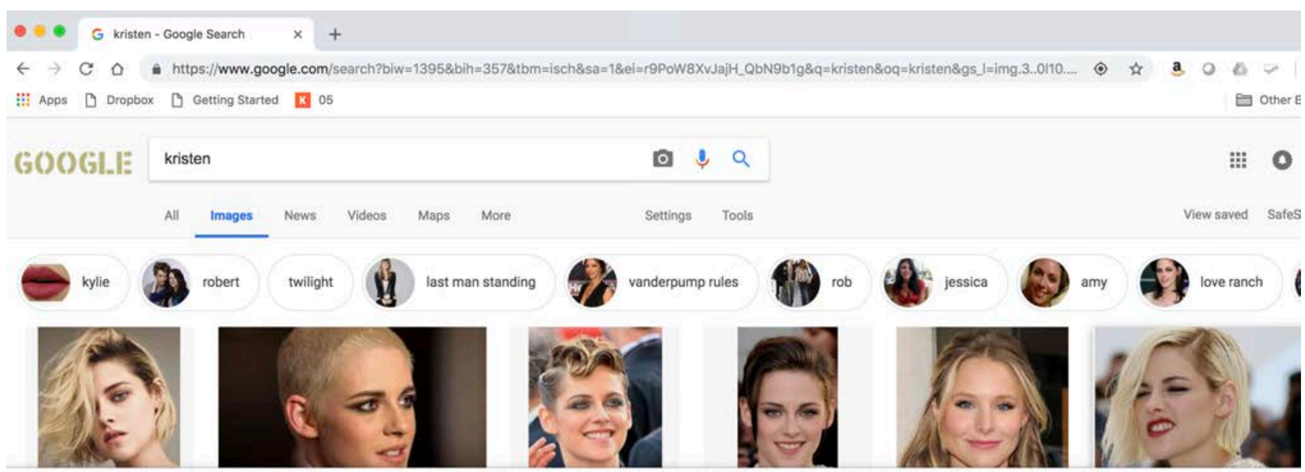
https://www.technologyreview.com/s/510646/racism-is-poisoning-online-ad-delivery-says-harvard-professor/

Try it!

Latanya



Kristen



a slight detour: more on racial bias

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 health-care algorithms

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.

Racial bias in health-care 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

Racial bias in health-care 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

back to hiring

Racial bias in resume screening

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

September 2004

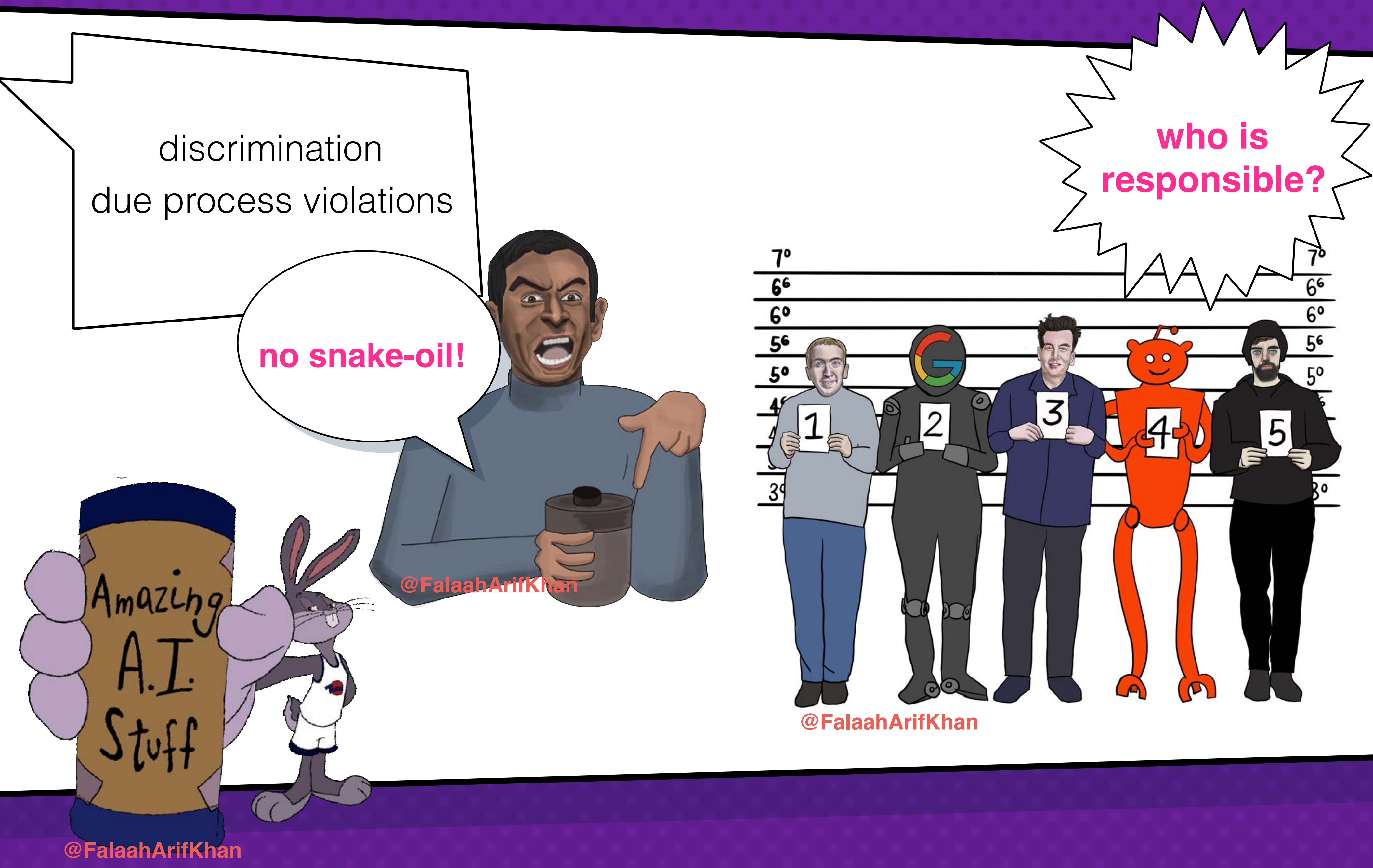
Marianne Bertrand

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.

In summary…


discrimination
due process violations
no snake-oil!
who is responsible?
Amazing A.I. Stuff
@FalaahArifKhan
@FalaahArifKhan
@FalaahArifKhan
1
2
3
4
5

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**

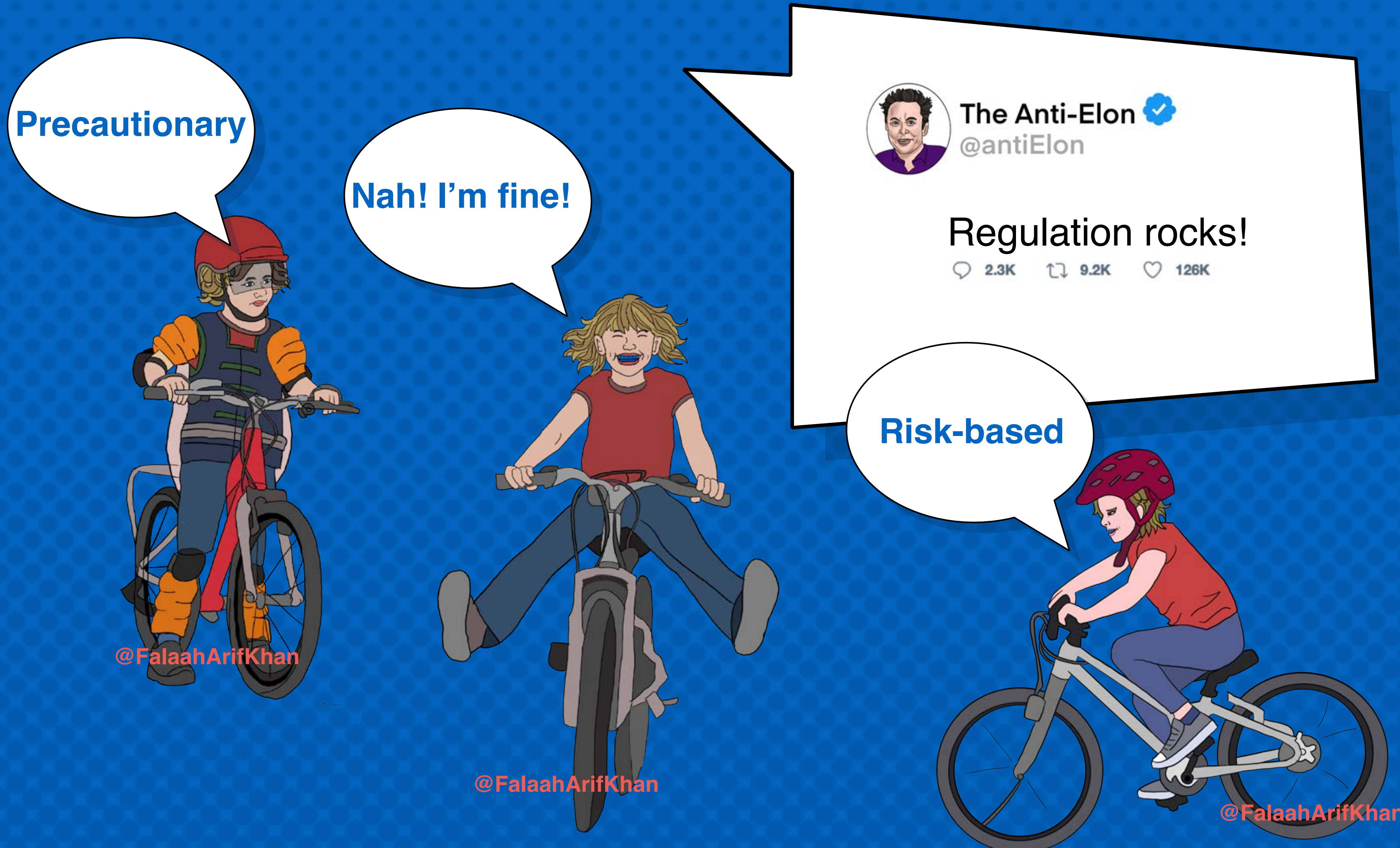
are subject to **auditing** and **public disclosure**

may or may not use AI

may or may not have autonomy

rely heavily on data

Regulating ADS?



ADS regulation in NYC: take 1

@FalaahArifKhan

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

Regulating hiring ADS: Int 1894-2020

THE NEW YORK CITY COUNCIL
Corey Johnson, Speaker

This bill would **regulate the use of automated employment decision tools**, which, for the purposes of this bill, encompass certain systems that use algorithmic methodologies to filter candidates for hire or to make decisions regarding any other term, condition or privilege of employment. This bill would prohibit the sale of such tools if they were not the **subject of an audit for bias** in the past year prior to sale, were not sold with a yearly bias audit service at no additional cost, and were not accompanied by a notice that the tool is subject to the provisions of this bill. This bill would also require any person who uses automated employment assessment tools for hiring and other employment purposes to **disclose to candidates, within 30 days, when such tools were used** to assess their candidacy for employment, and the **job qualifications or characteristics** for which the tool was used to screen. Violations of the provisions of the bill would incur a penalty.

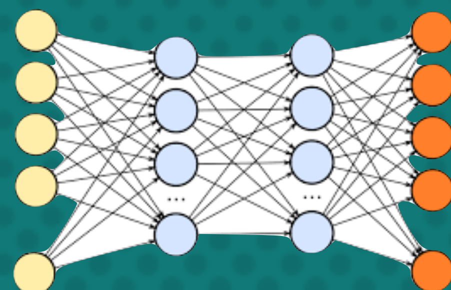
great!
now what?

Framing technical solutions

module 1:
algorithmic fairness

“Bias” in predictive analytics

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



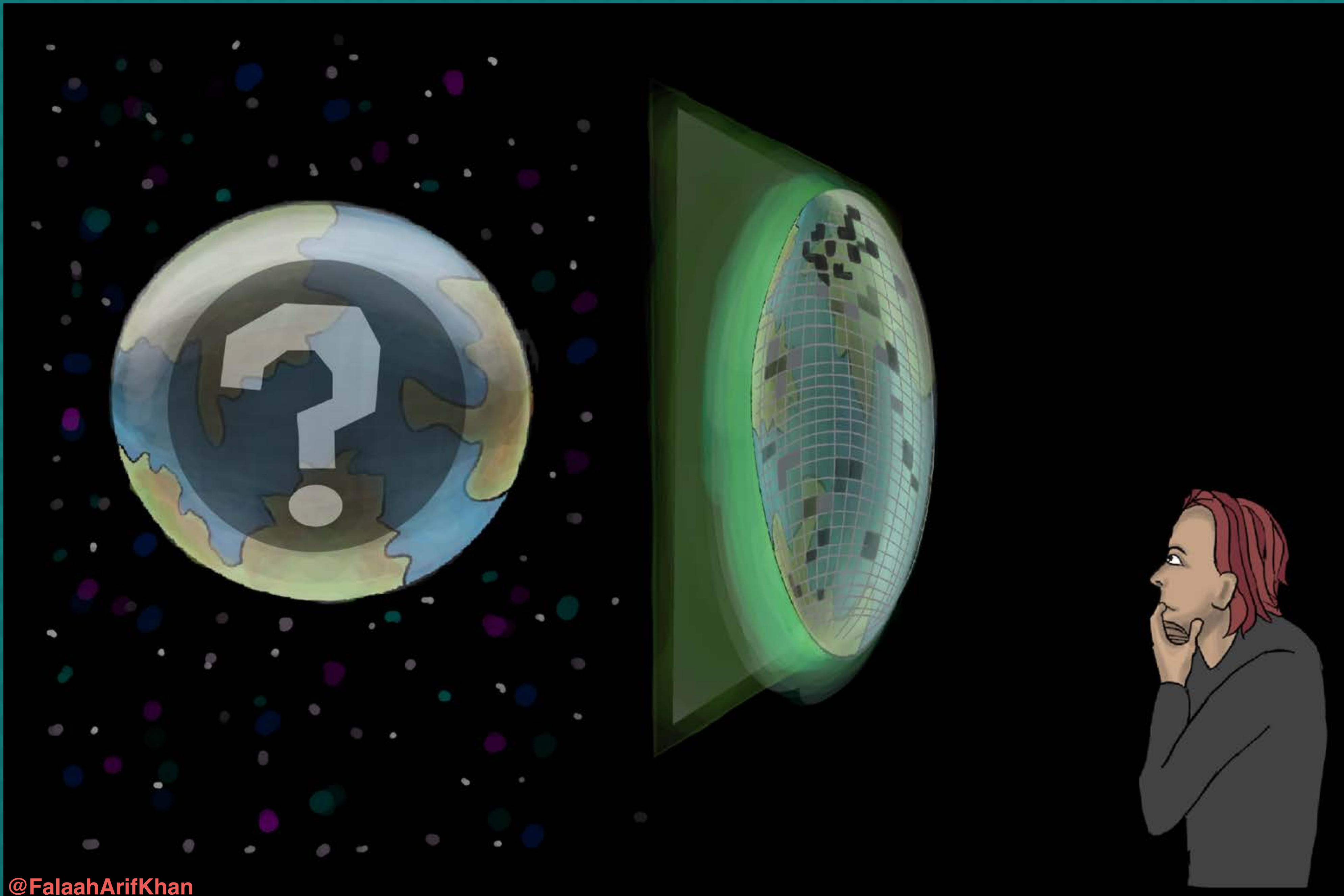
Statistical

model does not summarize the data correctly

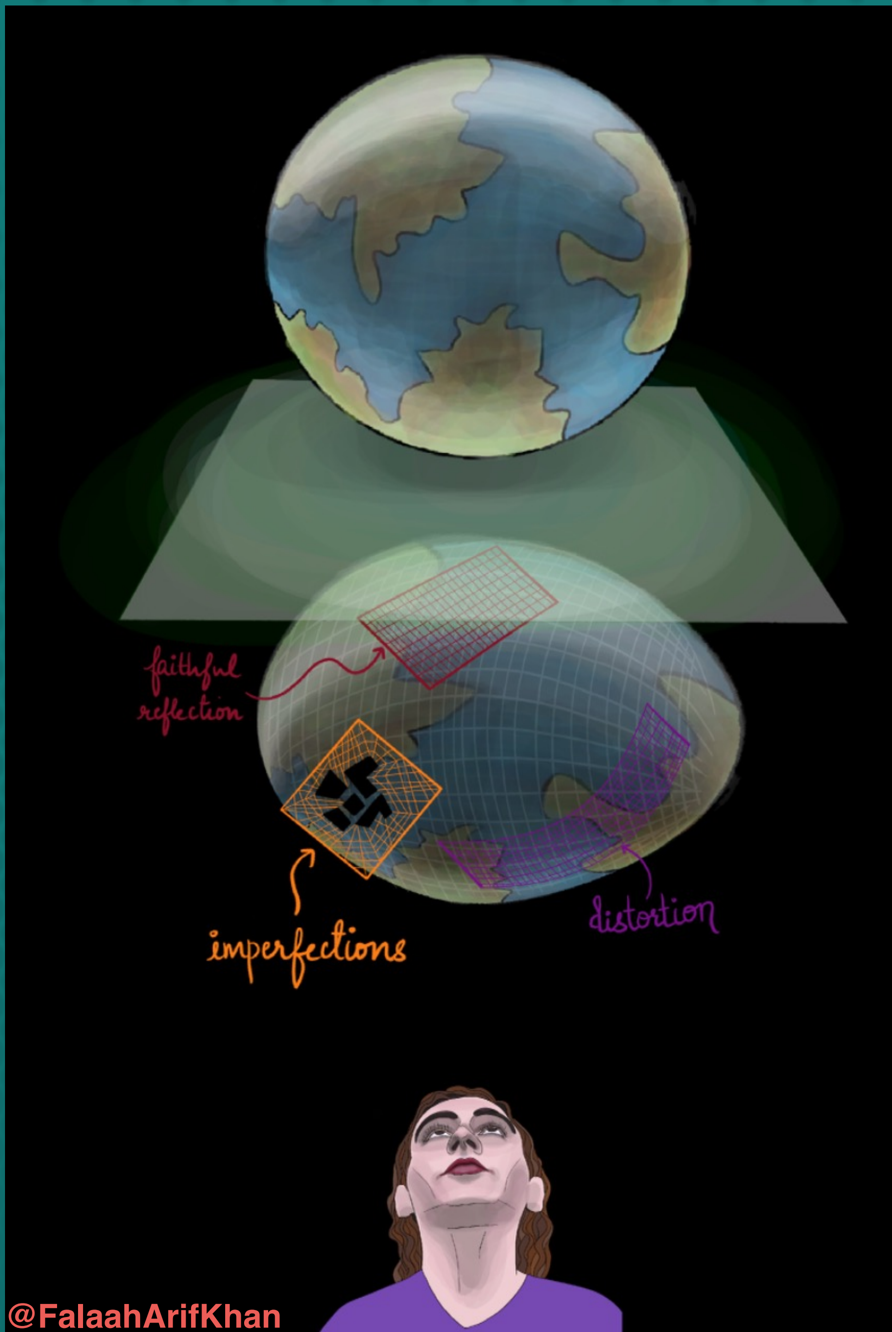
Societal

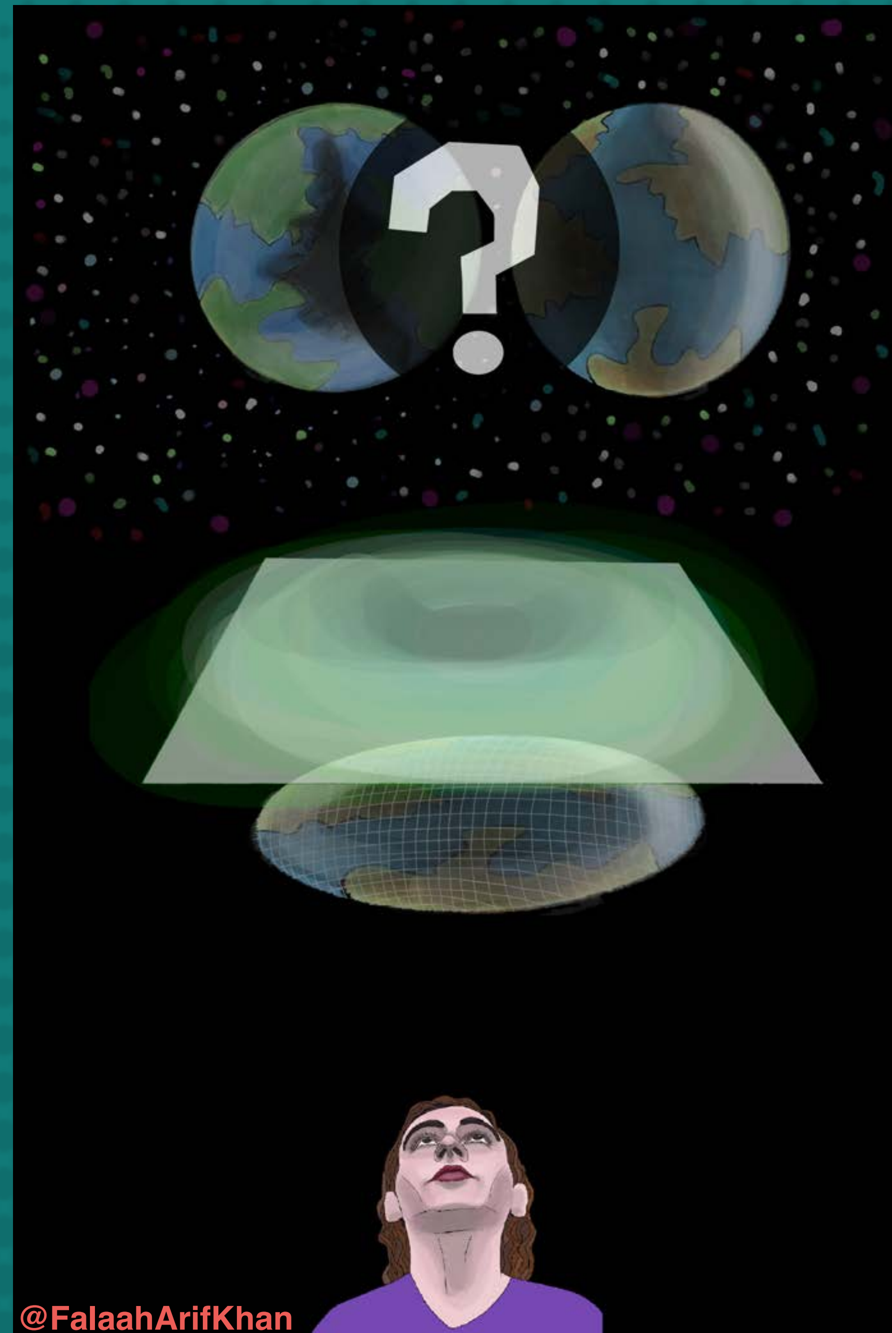
data does not represent the world correctly

Data, a reflection of the world

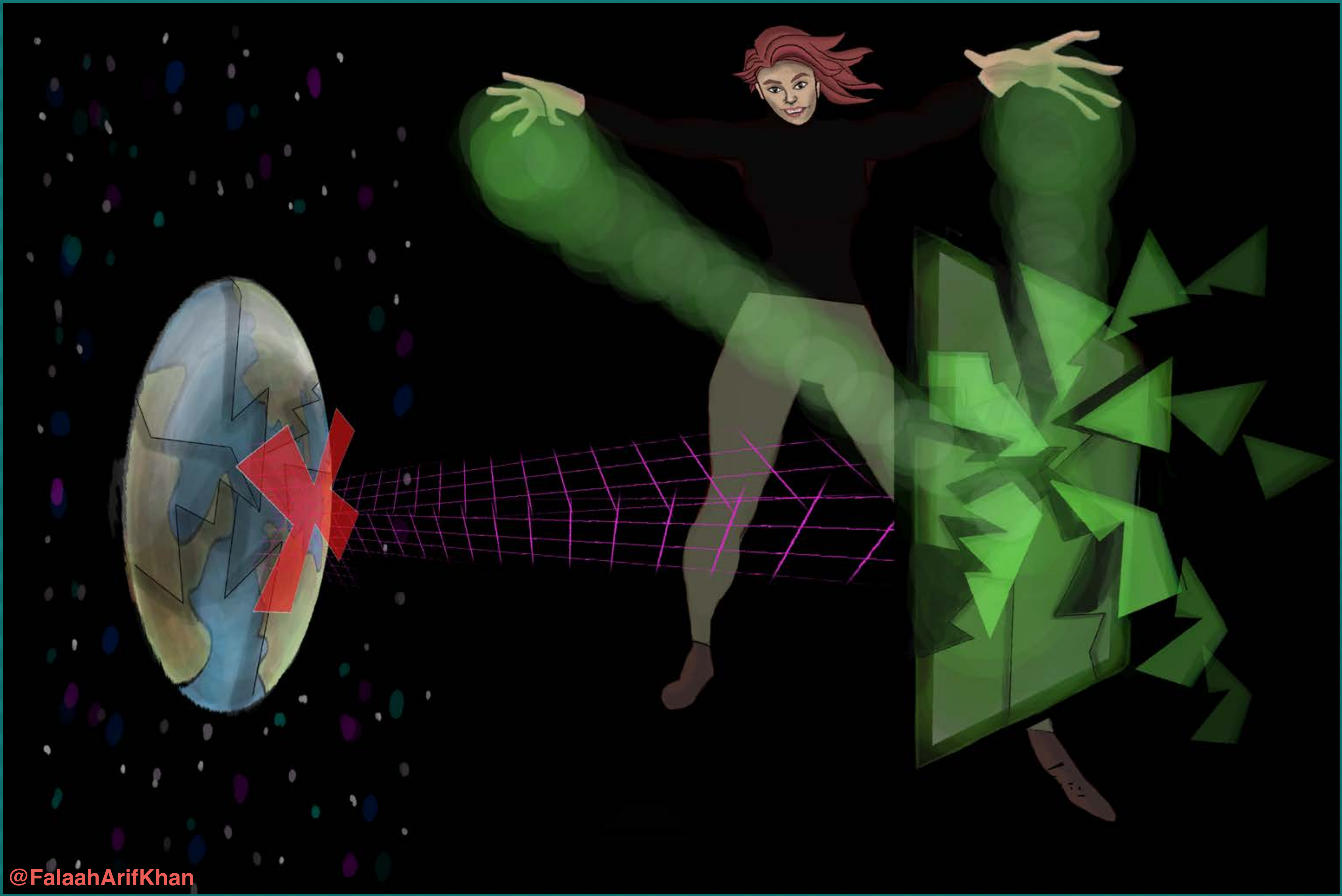


Data, a reflection of the world





Changing the reflection won't change the world



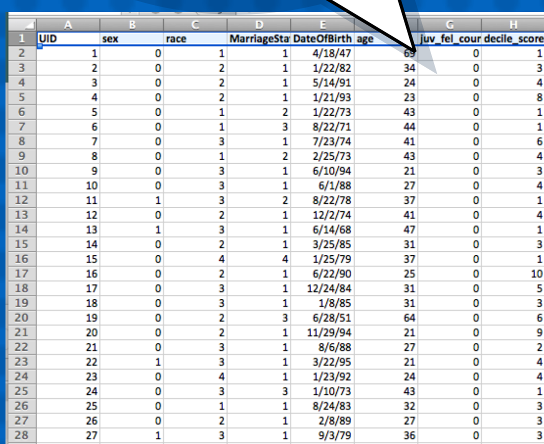
module 2:
the data science lifecycle

Frog’s eye view

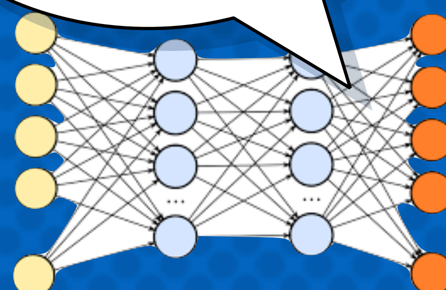
where did the data come from?

what happens inside the box?

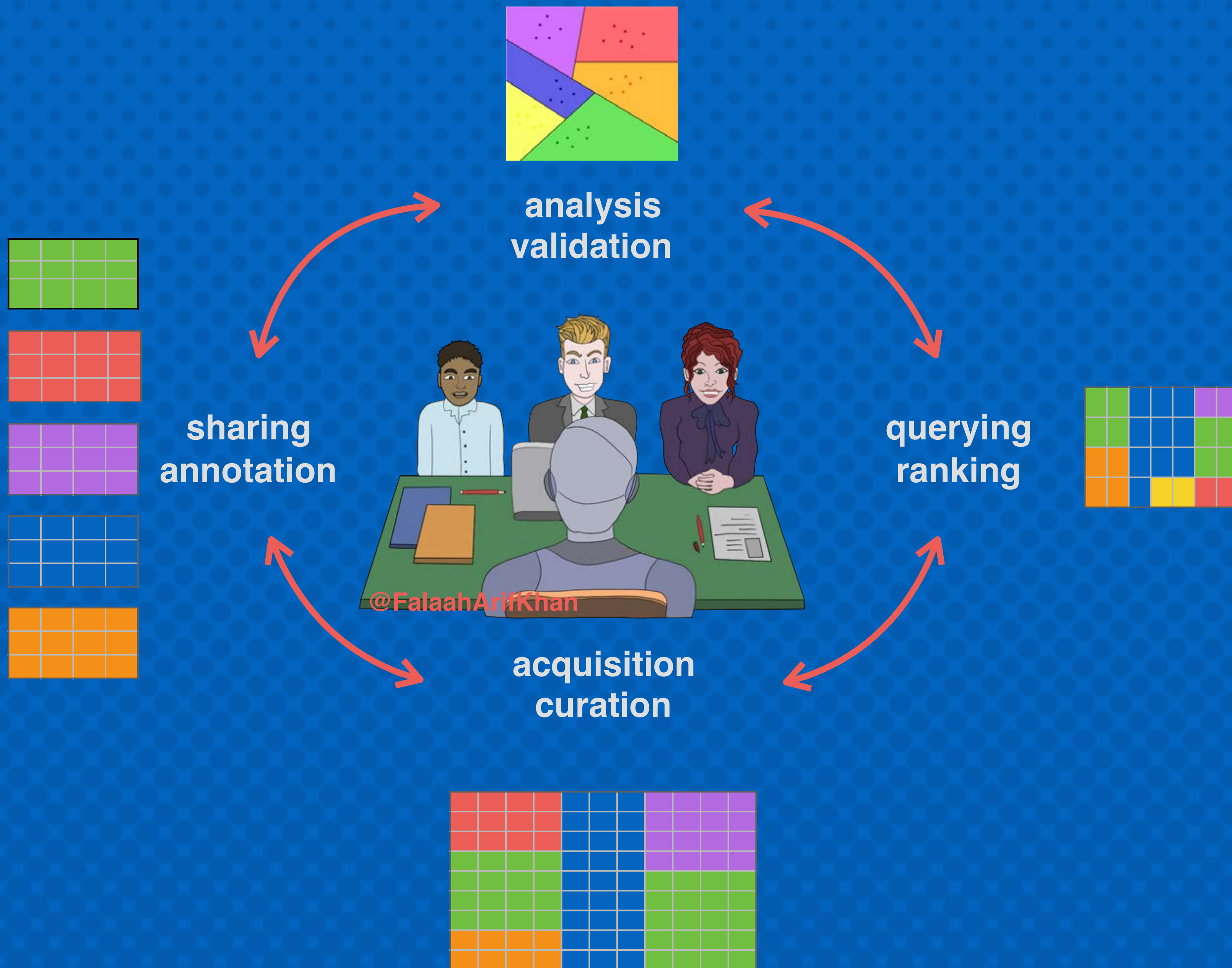
how are results used?



UID	sex	race	MarriageSta	DateOfBirth	age	juv_fel_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



Data lifecycle of an ADS

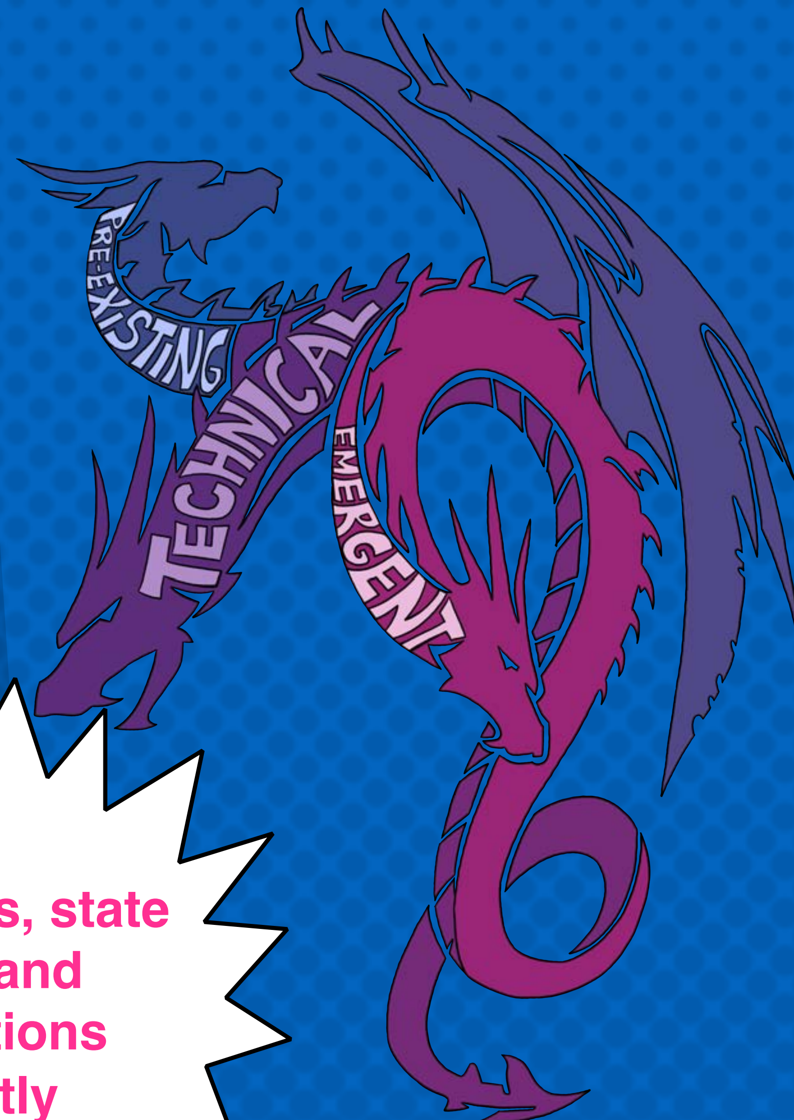


Bias in ADS, revisited

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

@FalaahArifKhan

module 3: data protection & privacy

Truth or dare

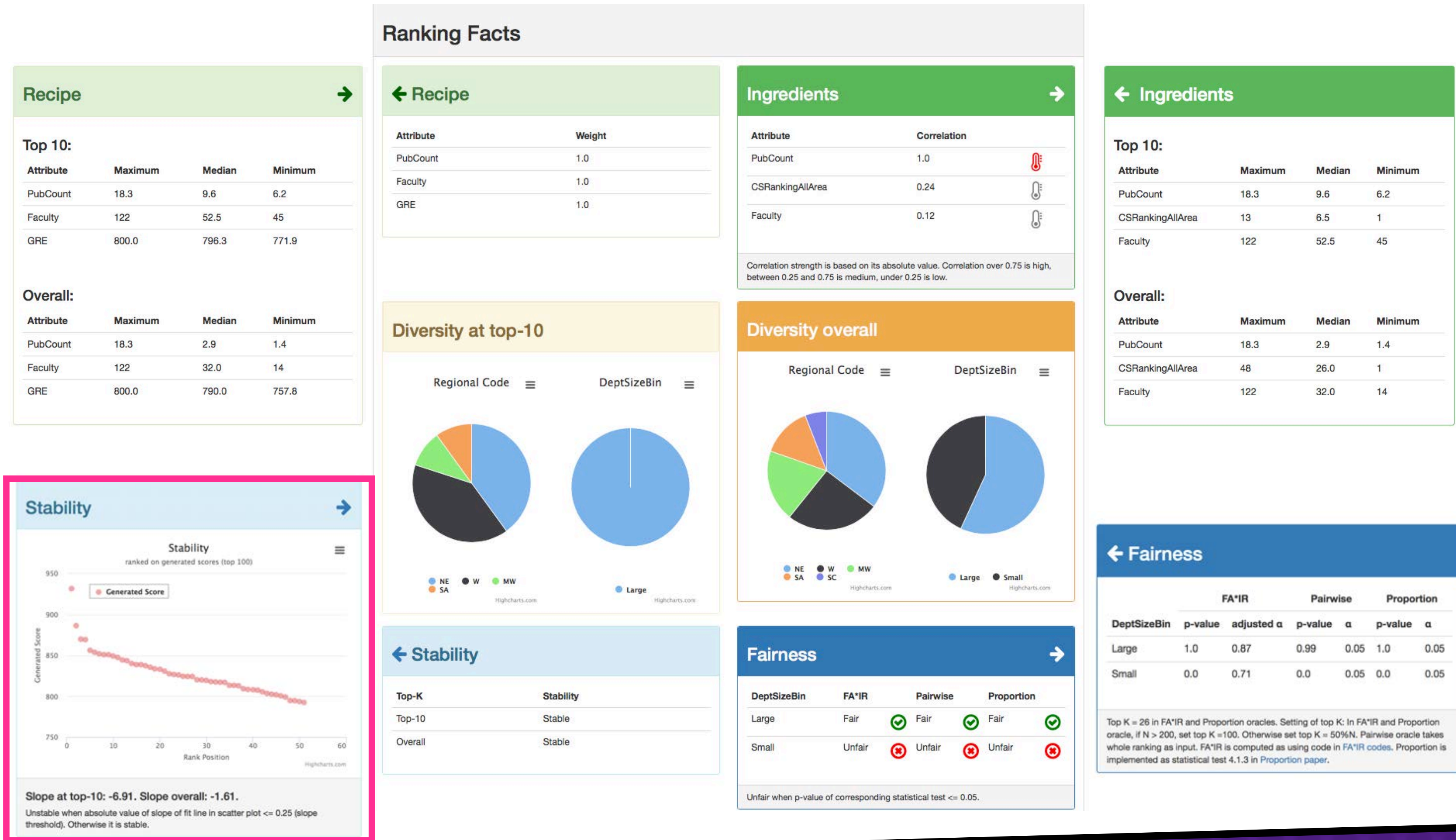
Did you go out drinking over the weekend?

protecting an individual
———
plausible deniability

learning about the population
———
noisy estimates

module 4: transparency & interpretability

Ranking Facts



[Yang, Stoyanovich, Asudeh, Howe, Jagadish, Miklau (2020)]

Stability in ranking

THE NEW YORKER

DEPT. OF EDUCATION FEBRUARY 14 & 21, 2011 ISSUE

THE ORDER OF THINGS

What college rankings really tell us.

By Malcolm Gladwell

1. Chevrolet Corvette 205
2. Lotus Evora 195
3. Porsche Cayman 195

1. Lotus Evora 205
2. Porsche Cayman 198
3. Chevrolet Corvette 192

1. Porsche Cayman 193
2. Chevrolet Corvette 186
3. Lotus Evora 182

Rankings are not benign. They enshrine very particular **ideologies**, and, at a time when American higher education is facing a crisis of accessibility and affordability, we have adopted **a de-facto standard** of college quality that is uninterested in both of those factors. And why? Because a group of magazine analysts in an office building in Washington, D.C., decided twenty years ago to **value selectivity over efficacy**, to **use proxies** that scarcely relate to what they're meant to be proxies for, and to **pretend that they can compare** a large, diverse, low-cost land-grant university in rural Pennsylvania with a small, expensive, private Jewish university on two campuses in Manhattan.

in summary

So what is RDS?

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

A technical course, with content drawn from:

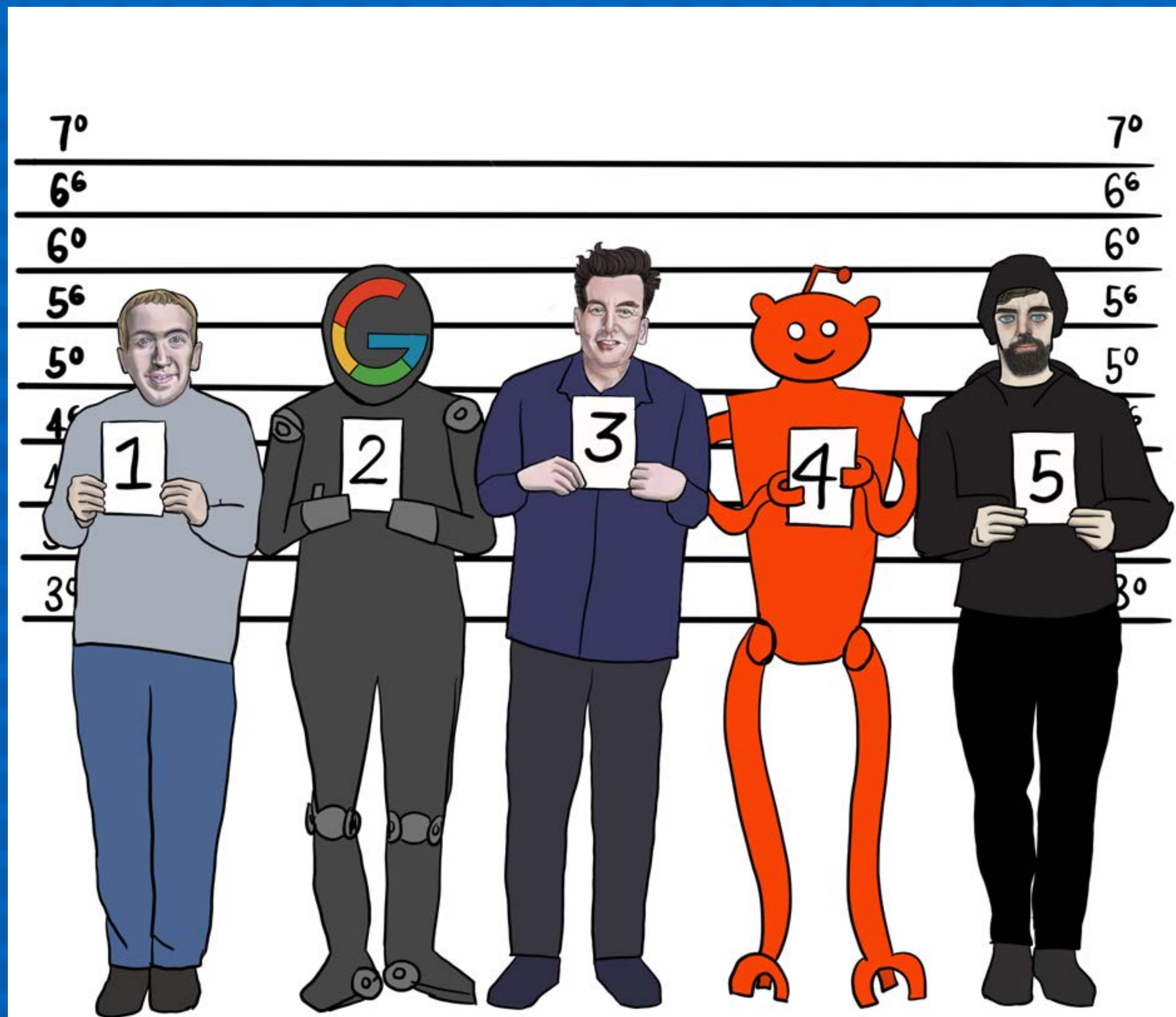
1. fairness, accountability and transparency
2. data engineering
3. security and privacy

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.

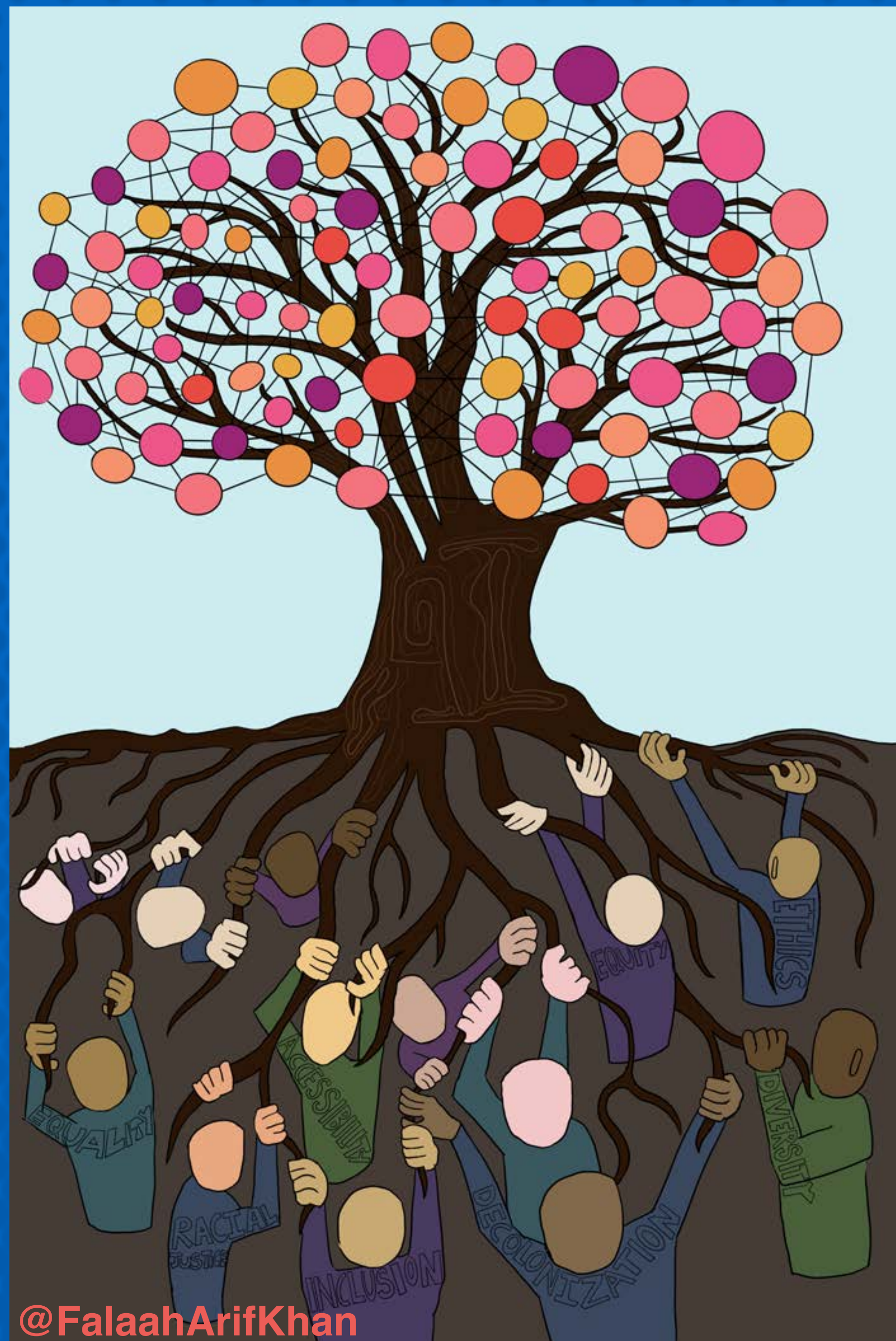
My perspective: a pragmatic engineer, **not** a technology skeptic.

We all are responsible

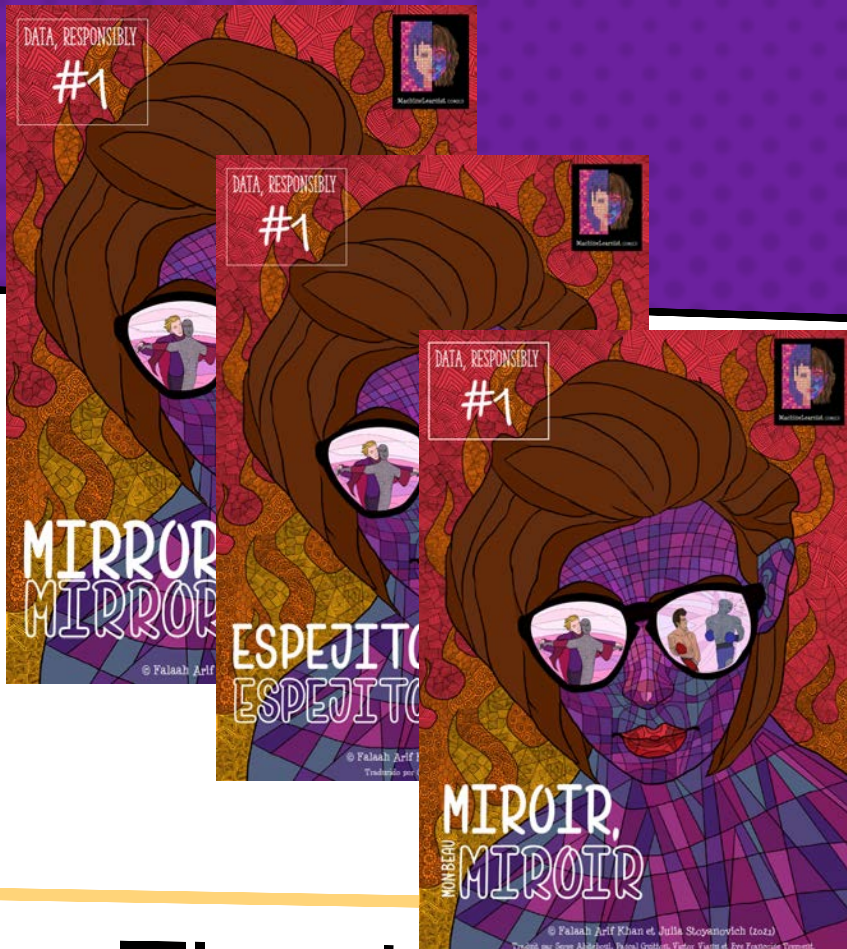


@FalaahArifKhan

Tech rooted in people



“*Mirror Mirror*”. Data, Responsibly Comics, Volume 1 (2020)



Thank you!

@stoyanoj

@FalaahArif Khan