Responsible Data Management

Julia Stoyanovich

Computer Science and Engineering
Center for Data Science
Center for Responsible AI
Visualization and Data Analytics Center
New York University

r/ai center for responsible ai

VIDA

NYU TANDON SCHOOL OF ENGINEERING

NYU Center for Data Science

r/ai center for responsible ai

Based on a recent *Comm. ACM* article

contributed articles

DOI:10.1145/3488717

Perspectives on the role and responsibility of the data-management research community in designing, developing, using, and overseeing automated decision systems.

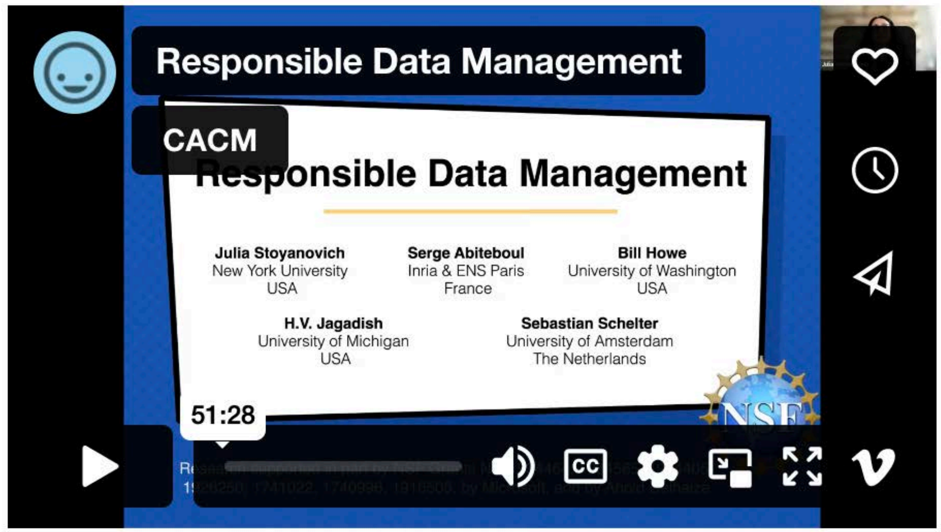
BY JULIA STOYANOVICH, SERGE ABITEBOUL, BILL HOWE, H.V. JAGADISH, AND SEBASTIAN SCHELTER

Responsible Data Management

INCORPORATING ETHICS AND legal compliance into data-driven algorithmic systems has been attracting significant attention from the computing research community, most notably under the umbrella of fair[8] and interpretable[16] machine learning. While important, much of this work has been limited in scope to the "last mile" of data analysis and has disregarded both the *system's design, development, and use life cycle* (What are we automating and why? Is the system working as intended? Are there any unforeseen consequences post-deployment?) and the *data life cycle* (Where did the data come from? How long is it valid and appropriate?). In this article, we argue two points. First, the decisions we make during data collection and preparation profoundly impact the robustness, fairness, and interpretability of the systems we build. Second, our responsibility for the operation of these systems does not stop when they are deployed.

Example: Automated hiring systems. To make our discussion concrete, consider the use of predictive analytics in hiring. Automated hiring systems are seeing ever broader use and are as varied as the hiring practices themselves, ranging from resume screeners that claim to identify promising applicants[a] to video and voice analysis tools that facilitate the interview process[b] and game-based assessments that promise to surface personality traits indicative of future success.[c] Bogen and Rieke[5] describe the hiring process from the employer's point of view as a series of decisions that forms a funnel, with stages corresponding to

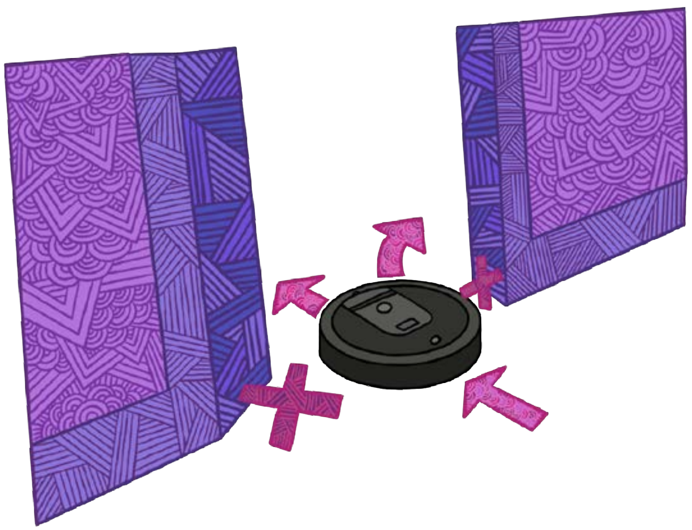
a https://www.crystalknows.com
b https://www.hirevue.com
c https://www.pymetrics.ai



https://cacm.acm.org/magazines/2022/6/261173-responsible-data-management

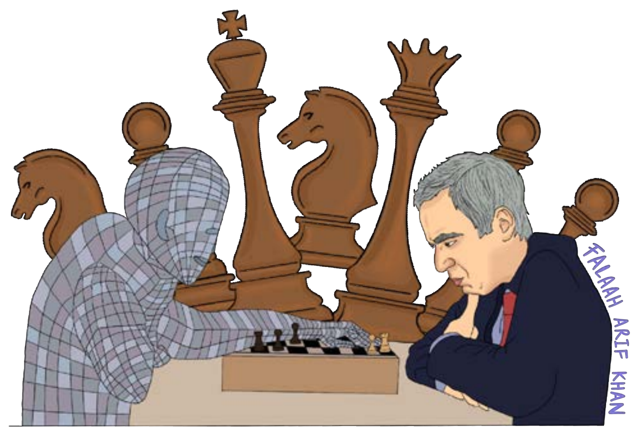
r/ai center for responsible ai

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

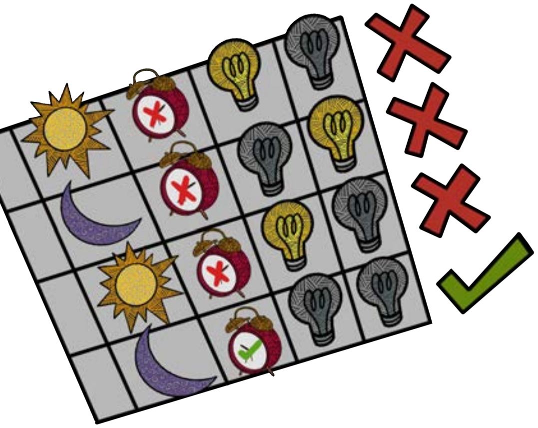
make our lives convenient

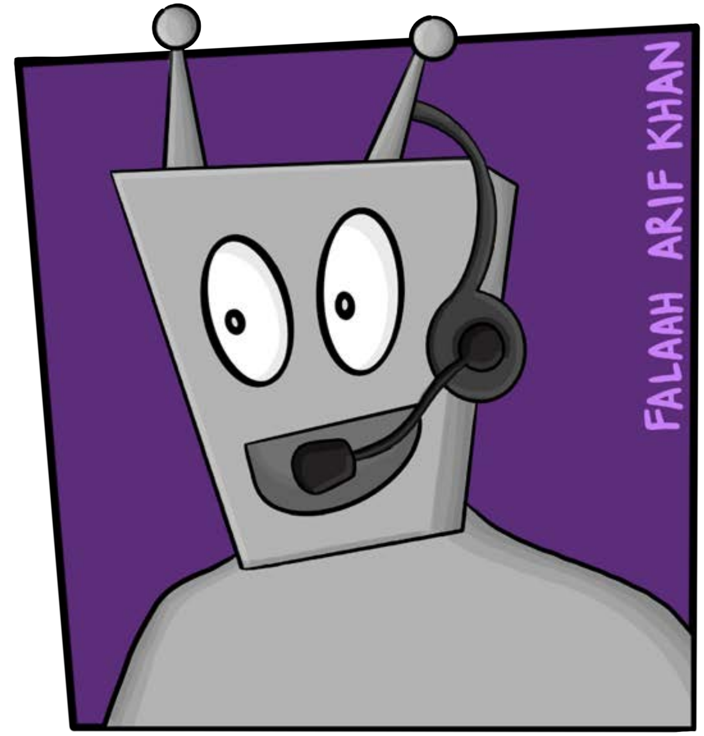
accelerate science

boost innovation

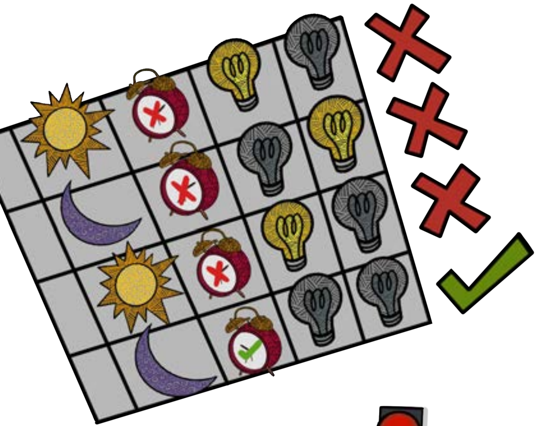
transform government

Machines make mistakes



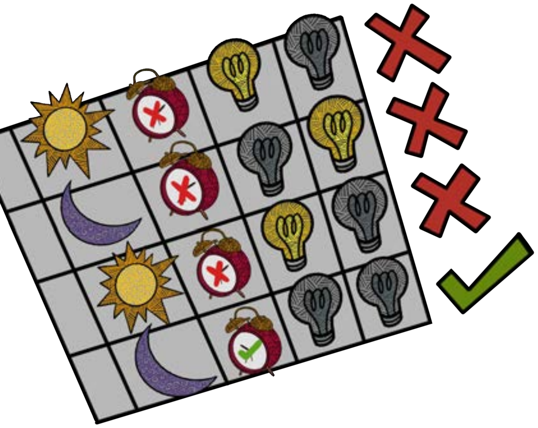


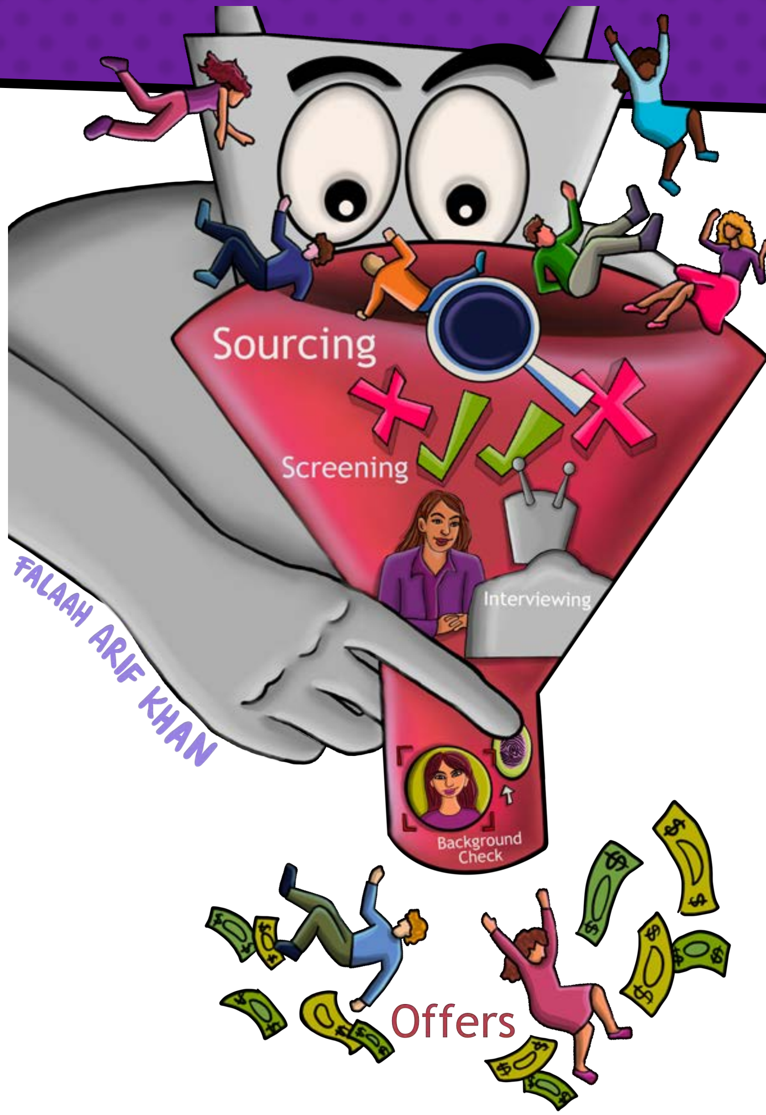
Mistakes lead to harms



center
for
responsible
ai

Harms can be cumulative





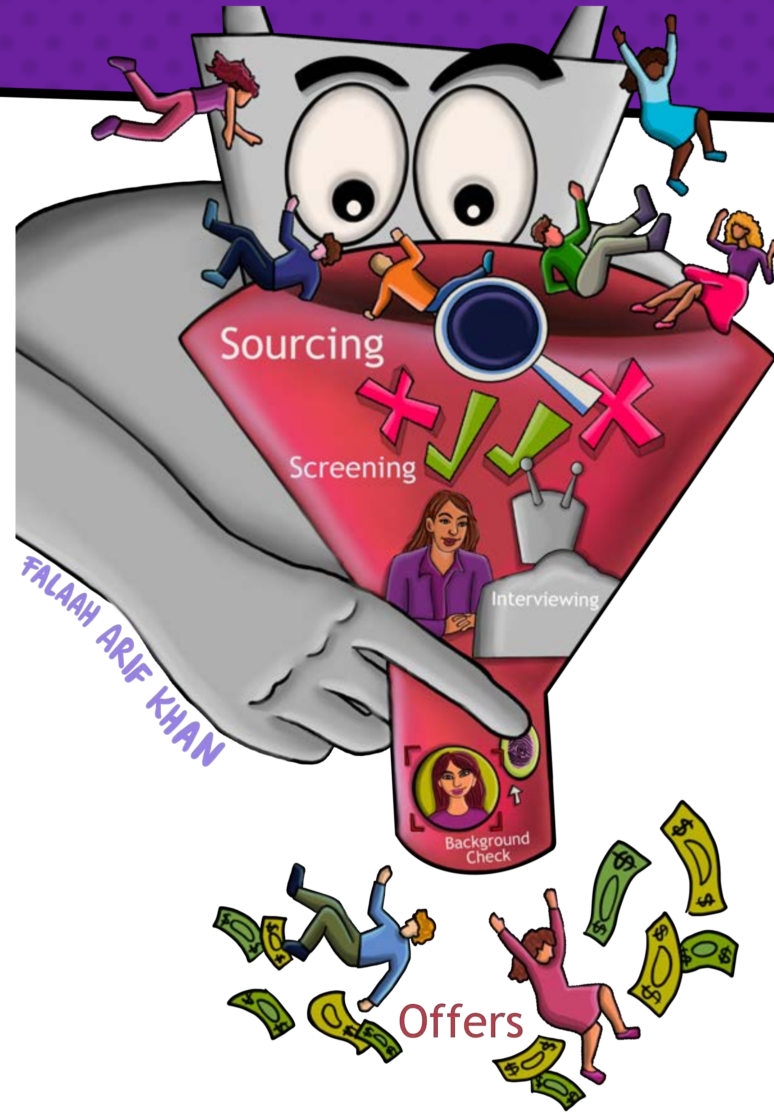
The promise of AI in hiring

Opportunity

efficiency for employers

efficiency for job seekers

improved workforce diversity



center for responsible ai

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.

Bias in algorithmic hiring

theguardian July 2015

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

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

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

Arbitrariness in algorithmic hiring

MIT Technology Review

ARTIFICIAL INTELLIGENCE

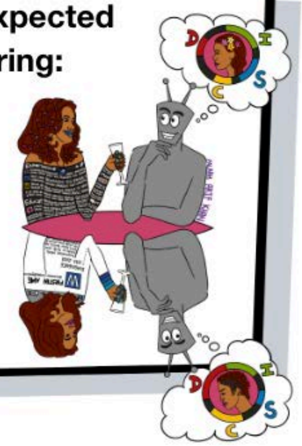
Podcast: Hired by an algorithm

Major companies are turning to AI to screen applicants and predict future job performance.

AAAI/ACM Conference on AI, Ethics, and Society (AIES 2022)

Resume Format, LinkedIn URLs and Other Unexpected Influences on AI Personality Prediction in Hiring:
Results of an Audit

Alene K. Rhea, Lauren D'Arinzo, Kelsey Markey, Hilke Schellmann, Mona Sloane, Paul Squires, Julia Stoyanovich
New York University, USA



center for responsible ai

New York City Local Law 144 of 2021

THE NEW YORK CITY COUNCIL
Corey Johnson, Speaker

December 11, 2021

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

Great! Now what?

all about that bias

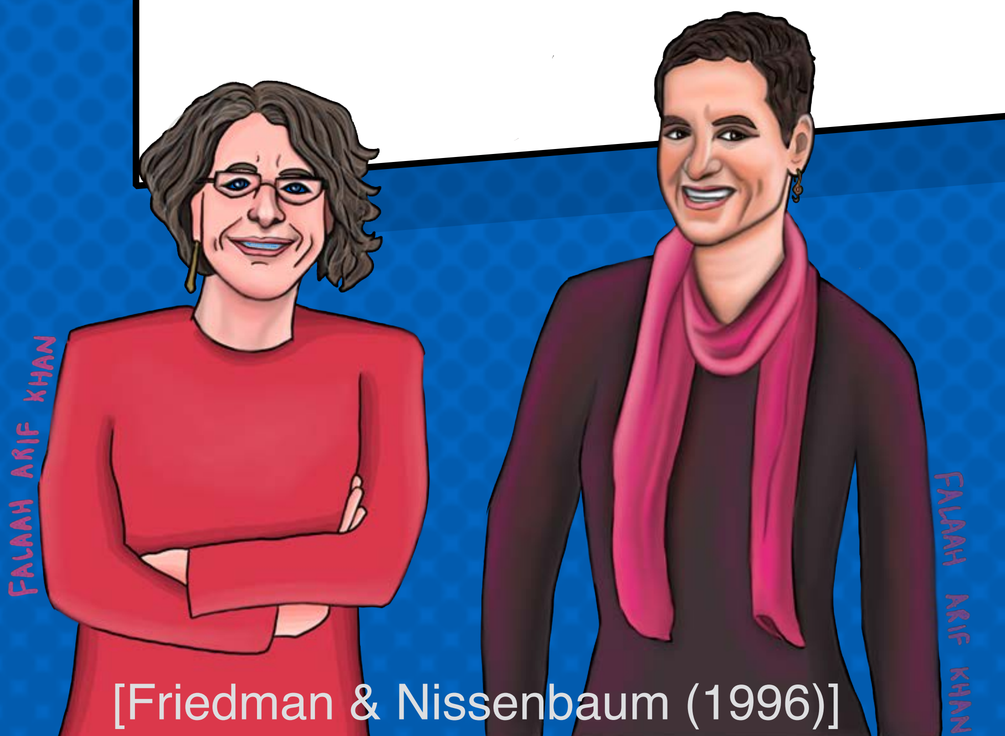
r/ai center for responsible ai

Bias in computer systems

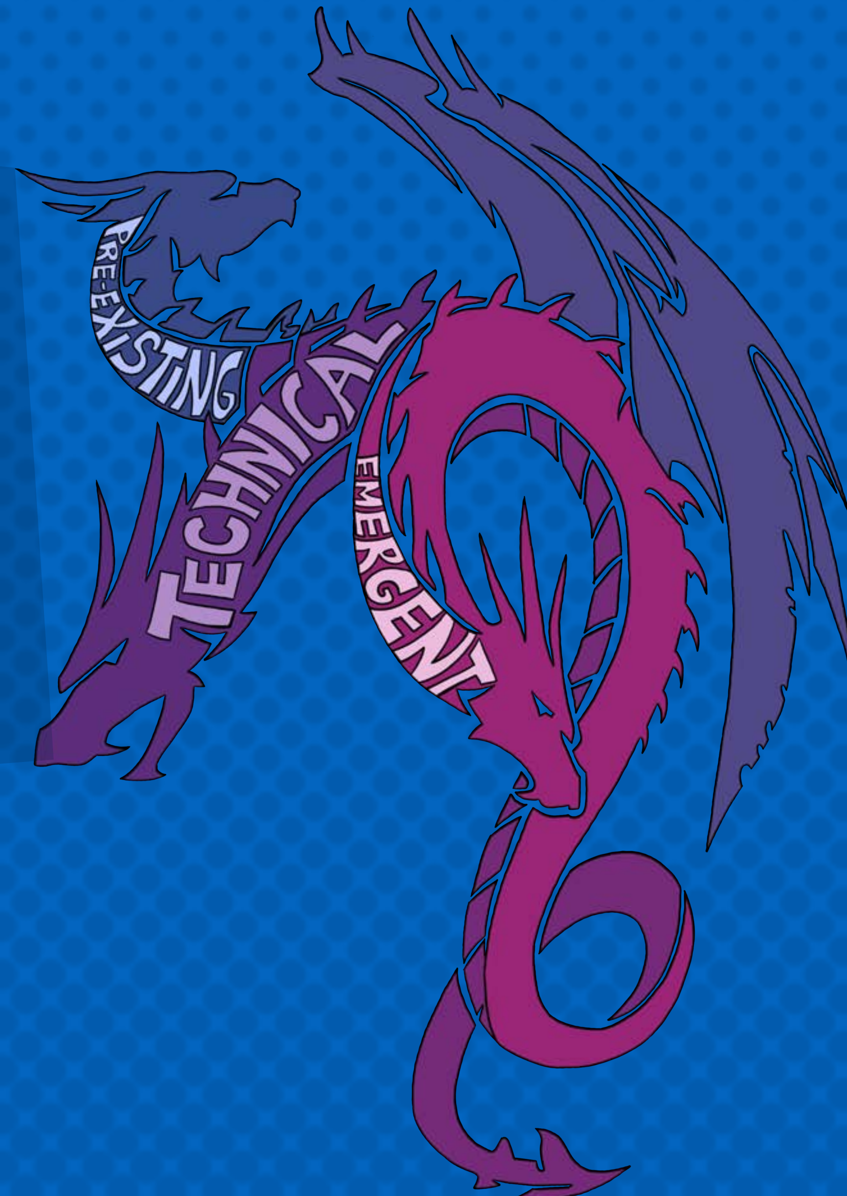
Pre-existing: exists independently of algorithm, has origins in society

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

Emergent: arises due to context of use



[Friedman & Nissenbaum (1996)]

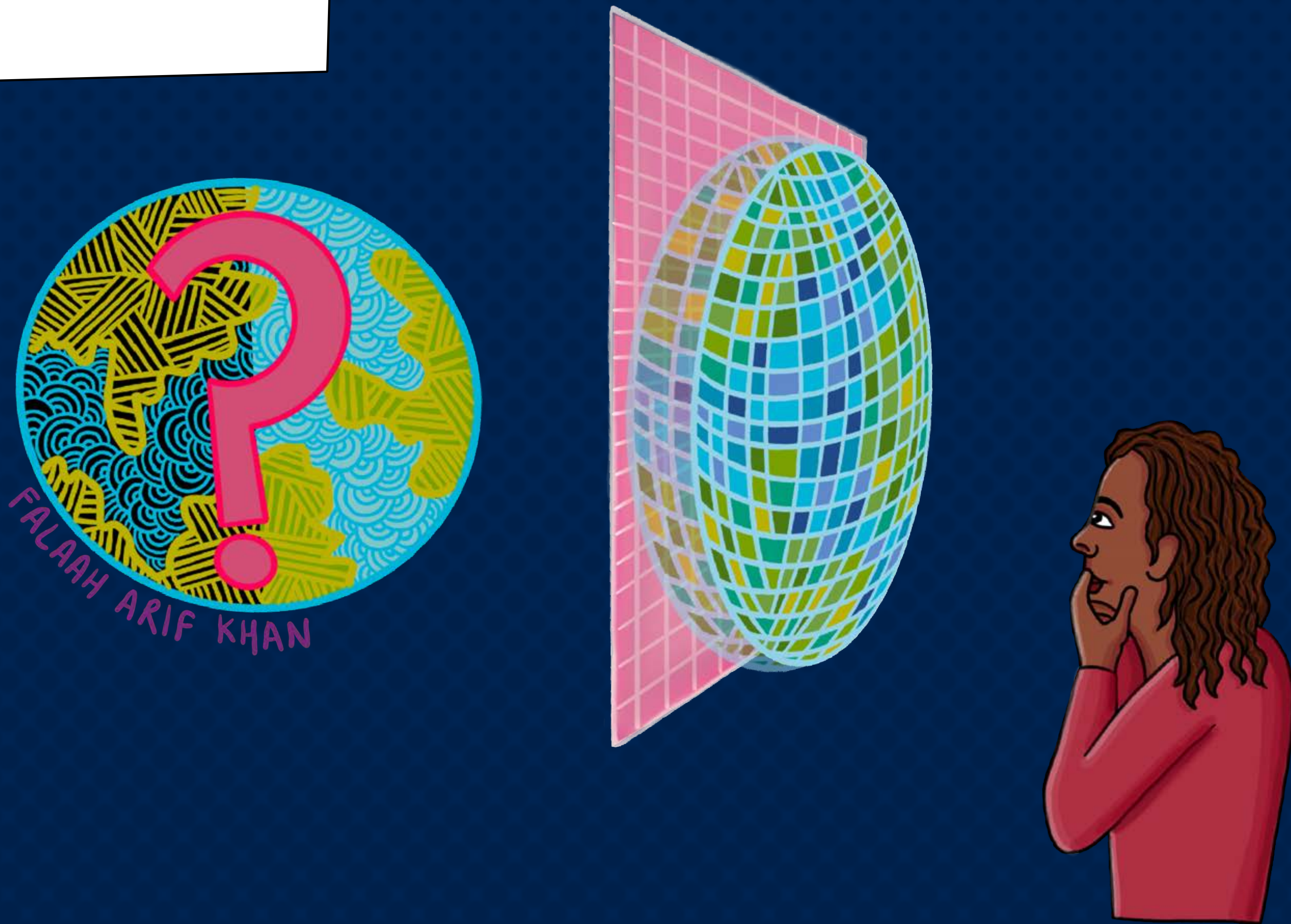


r/ai center for responsible ai

pre-existing bias

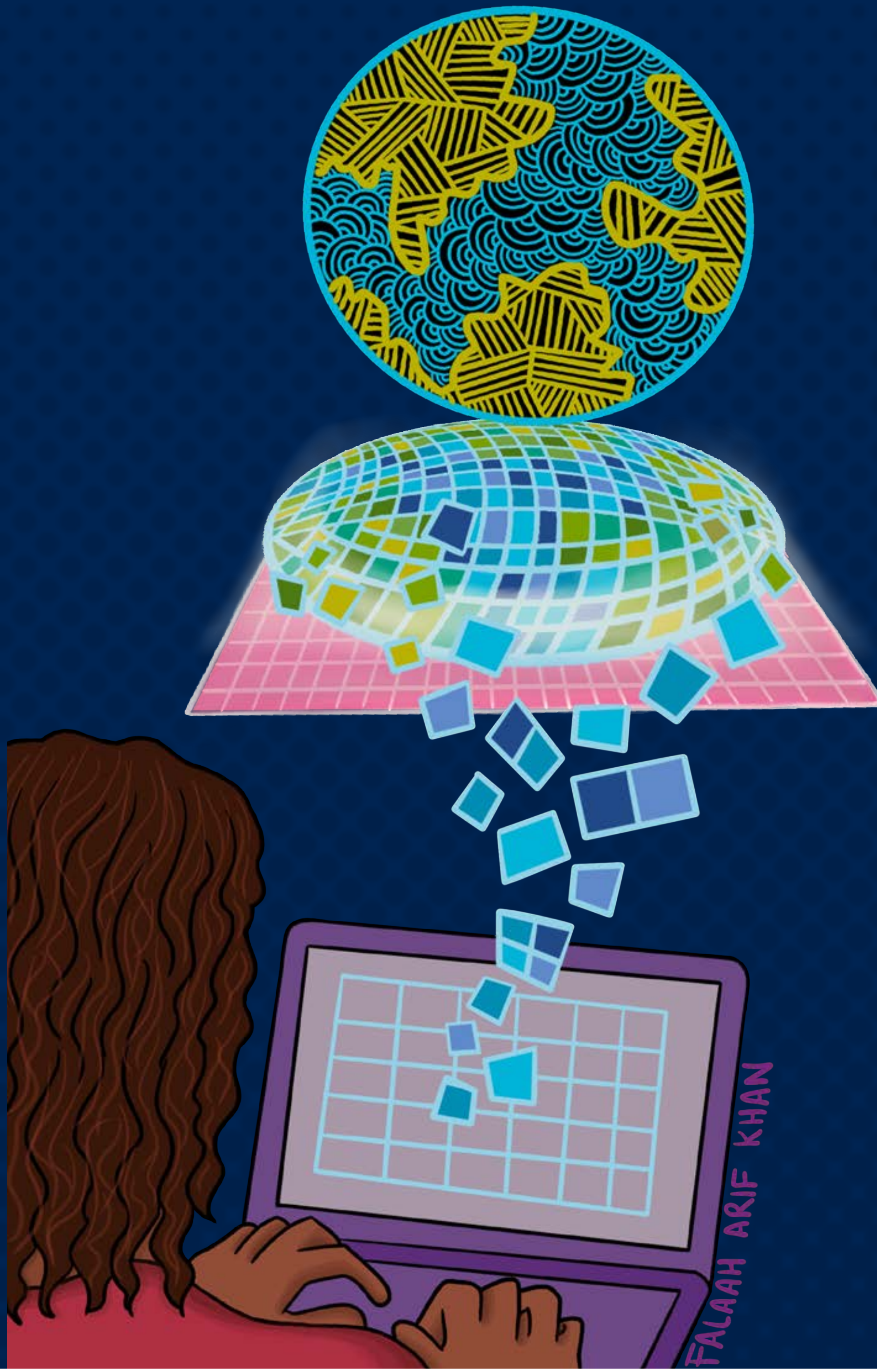
center for responsible ai

Pre-existing bias has origins in society



r/ai center for responsible ai

Pre-existing bias has origins in society

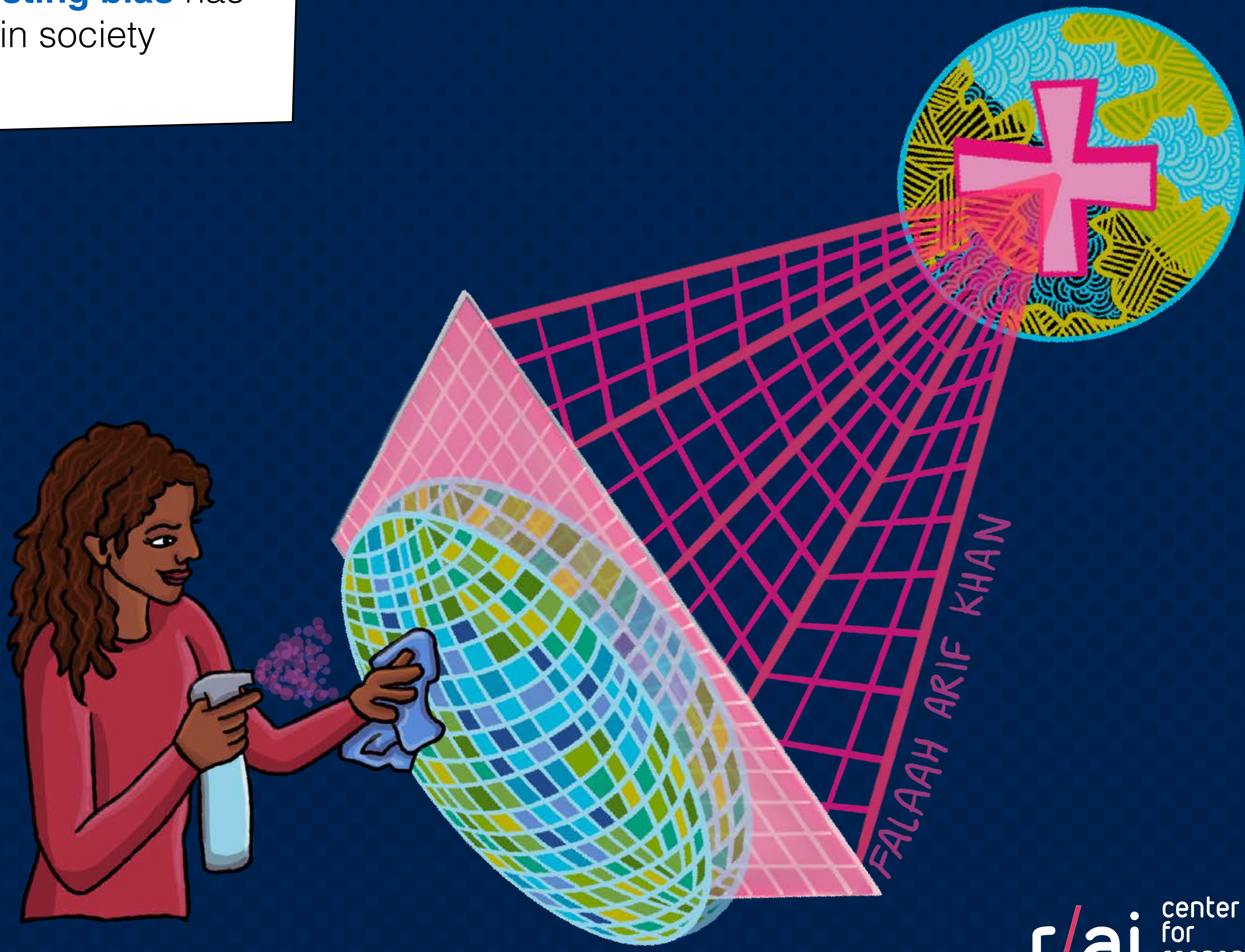


r/ai center for responsible ai

Pre-existing bias has origins in society

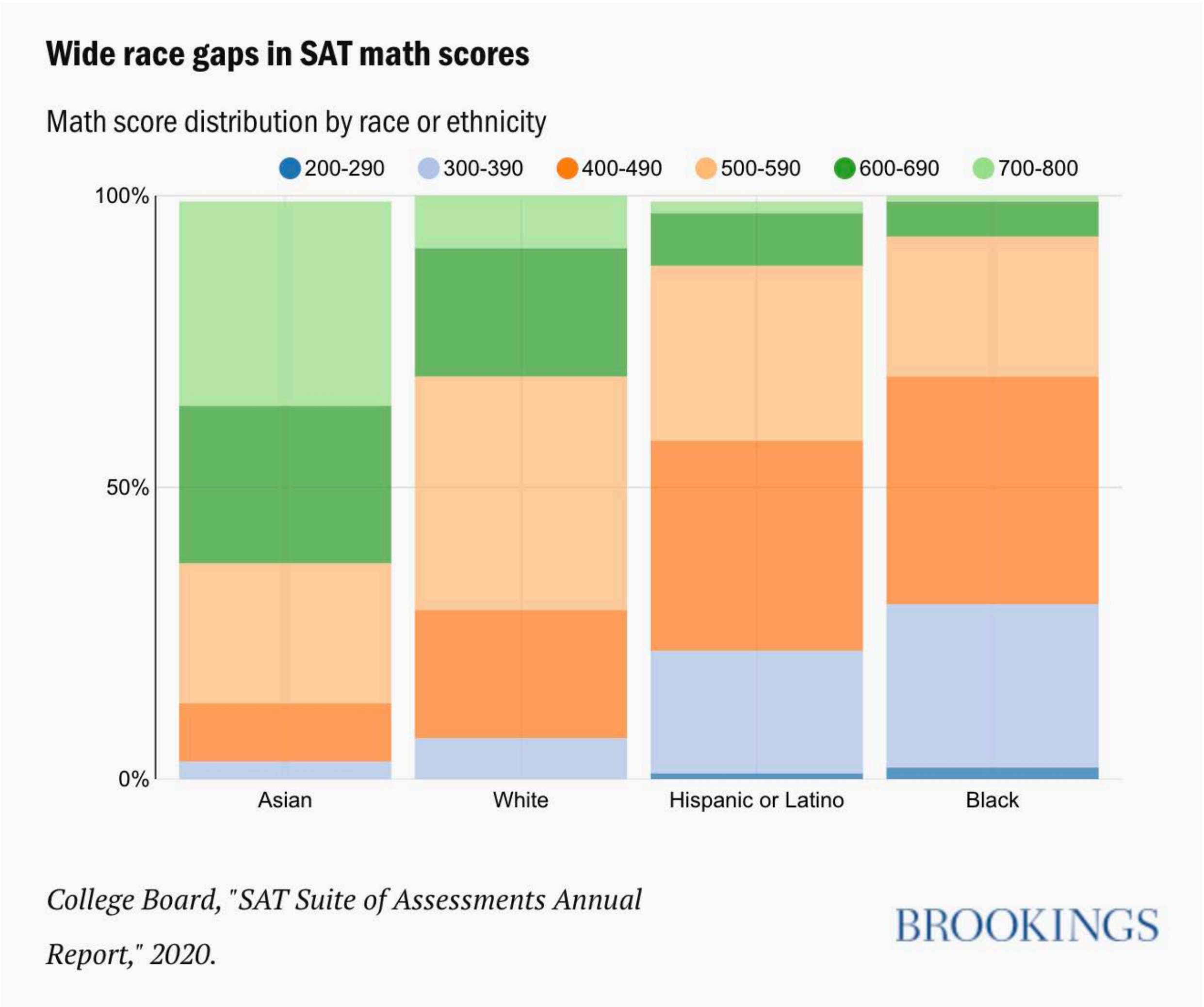
r/ai center for responsible ai

Pre-existing bias has origins in society



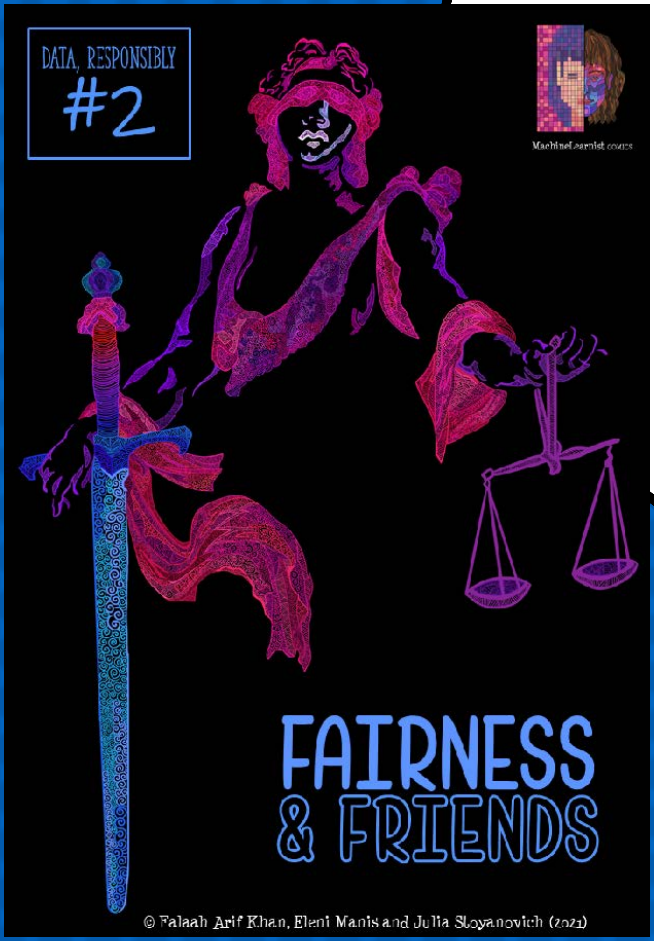
r/ai center for responsible ai

Example of pre-existing bias



https://www.brookings.edu/blog/up-front/2020/12/01/sat-math-scores-mirror-and-maintain-racial-inequity/

r/ai center for responsible ai



equality of opportunity

[Arif Khan, Manis, Stoyanovich (2022)]

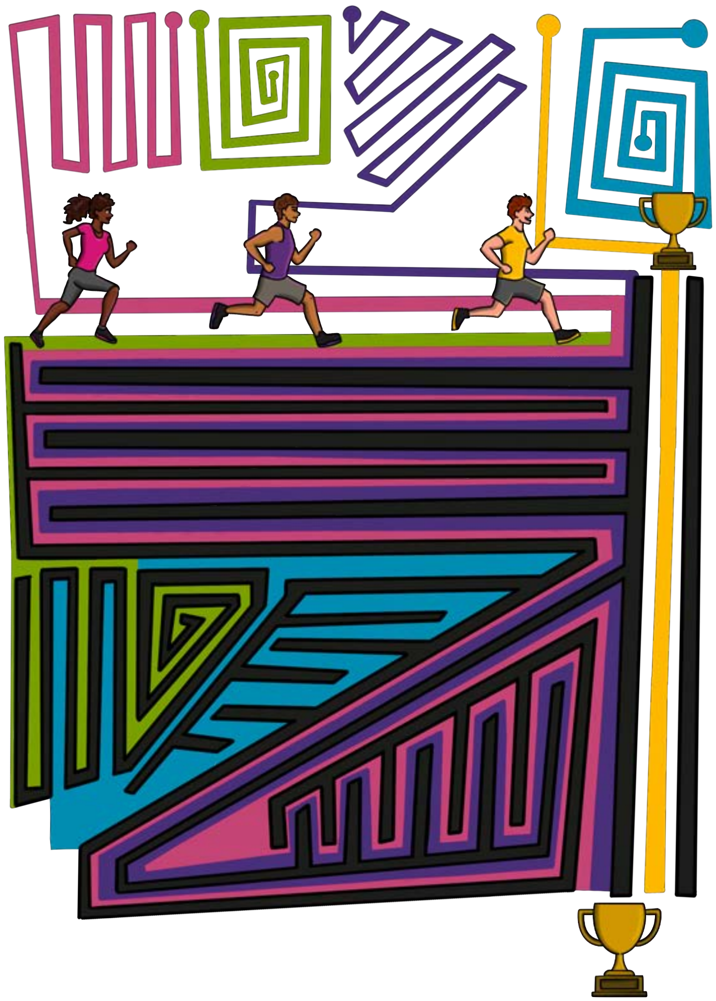
center for responsible ai

Principles of equality of opportunity

Fair contests: competitions should only judge people based on morally relevant "merit" (i.e., qualifications), not based on morally arbitrary factors (e.g., gender, race, socio-economic status)

Fair life chances: level the playing field over a lifetime

Domains of equality of opportunity



(1) Fairness at a specific decision point

distribution of social goods, like employment & loans

(2) Equality in developmental opportunity

access to opportunities that shape one's ability to compete for positions at a decision point

(3) Equality of opportunity over a lifetime

access to comparable opportunity sets

Fairness in Ranking, Part I: Score-Based Ranking

MEIKE ZEHLIKE, Humboldt University of Berlin, Max Planck Institute for Software Systems, and Zalando Research, Germany
KE YANG, New York University, NY, and University of Massachusetts, Amherst, MA, USA
JULIA STOYANOVICH, New York University, NY, USA

In the past few years, there has been much work on incorporating fairness requirements into algorithmic rankers, with contributions coming from the data management, algorithms, information retrieval, and recommender systems communities. In this survey, we give a systematic overview of this work, offering a broad perspective that connects formalizations and algorithmic approaches across sub-fields. An important contribution of our work is in developing a common narrative around the value frameworks that motivate specific fairness-enhancing interventions in ranking. This allows us to unify the presentation of mitigation objectives and of algorithmic techniques to help meet those objectives or identify trade-offs.

In this first part of this survey, we describe four classification frameworks for fairness-enhancing interventions, along which we relate the technical methods surveyed in this article, discuss evaluation datasets, and present technical work on fairness in score-based ranking. In the second part of this survey, we present methods that incorporate fairness in supervised learning, and also give representative examples of recent work on fairness in recommendation and matchmaking systems. We also discuss evaluation frameworks for fair score-based ranking and fair learning-to-rank, and draw a set of recommendations for the evaluation of fair ranking methods.

CCS Concepts: • **Information systems → Data management systems**; • **Social and professional topics → Computing/technology policy;**

Additional Key Words and Phrases: Fairness, ranking, set selection, responsible data science, survey

ACM Reference format:
Meike Zehlike, Ke Yang, and Julia Stoyanovich. 2022. Fairness in Ranking, Part I: Score-Based Ranking. *ACM Comput. Surv.* 55, 6, Article 118 (December 2022), 36 pages.
https://doi.org/10.1145/3533379

1 INTRODUCTION

The research community recognizes several important normative dimensions of information technology including privacy, transparency, and fairness. In this survey, we focus on fairness—a broad and inherently interdisciplinary topic of which the social and philosophical foundations are still unresolved [17].

This research was supported in part by NSF Awards No. 1934464, 1916505, and 1922658.
Authors' addresses: M. Zehlike, Humboldt University of Berlin, Max Planck Institute for Software Systems, and Zalando Research, Germany; email: meikezehlike@mpi-sws.org; K. Yang, New York University, NY, and University of Massachusetts, Amherst, MA, USA; email: ky630@nyu.edu; J. Stoyanovich, New York University, NY, USA; email: stoyanovich@nyu.edu.

0360-0300/2022/12-ART118 $15.00
https://doi.org/10.1145/3533379

Fairness in Ranking, Part II: Learning-to-Rank and Recommender Systems

MEIKE ZEHLIKE, Humboldt University of Berlin, Max Planck Institute for Software Systems, and Zalando Research, Germany
KE YANG, New York University, NY, and University of Massachusetts, Amherst, MA, USA
JULIA STOYANOVICH, New York University, NY, USA

In the past few years, there has been much work on incorporating fairness requirements into algorithmic rankers, with contributions coming from the data management, algorithms, information retrieval, and recommender systems communities. In this survey, we give a systematic overview of this work, offering a broad perspective that connects formalizations and algorithmic approaches across subfields. An important contribution of our work is in developing a common narrative around the value frameworks that motivate specific fairness-enhancing interventions in ranking. This allows us to unify the presentation of mitigation objectives and of algorithmic techniques to help meet those objectives or identify trade-offs.

In the first part of this survey, we describe four classification frameworks for fairness-enhancing interventions, along which we relate the technical methods surveyed in this article, discuss evaluation datasets, and present technical work on fairness in score-based ranking. In the second part of this survey, we present methods that incorporate fairness in supervised learning, and also give representative examples of recent work on fairness in recommendation and matchmaking systems. We also discuss evaluation frameworks for fair score-based ranking and fair learning-to-rank, and draw a set of recommendations for the evaluation of fair ranking methods.

CCS Concepts: • **Information systems → Data management systems**; • **Social and professional topics → Computing/technology policy;**

Additional Key Words and Phrases: Fairness, ranking, set selection, responsible data science, survey

ACM Reference format:
Meike Zehlike, Ke Yang, and Julia Stoyanovich. 2022. Fairness in Ranking, Part II: Learning-to-Rank and Recommender Systems. *ACM Comput. Surv.* 55, 6, Article 117 (December 2022), 41 pages.
https://doi.org/10.1145/3533380

1 INTRODUCTION

This is the second part of a survey on fairness in ranking. In the first part, we argued for the importance of a systematic overview of work on incorporating fairness requirements into algorithmic rankers. Which specific fairness requirements a decision maker will assert depends on the

This research was supported in part by NSF Awards No. 1934464, 1916505, and 1922658.
Authors' addresses: M. Zehlike, Humboldt University of Berlin, Max Planck Institute for Software Systems, and Zalando Research, Germany; email: meikezehlike@mpi-sws.org; K. Yang, New York University, NY, and University of Massachusetts, Amherst, MA, USA; email: ky630@nyu.edu; J. Stoyanovich, New York University, NY, USA; email: stoyanovich@nyu.edu.

0360-0300/2022/12-ART117 $15.00
https://doi.org/10.1145/3533380

[Zehlike, Yang, Stoyanovich (2023)]

Diverse balanced ranking

Goals

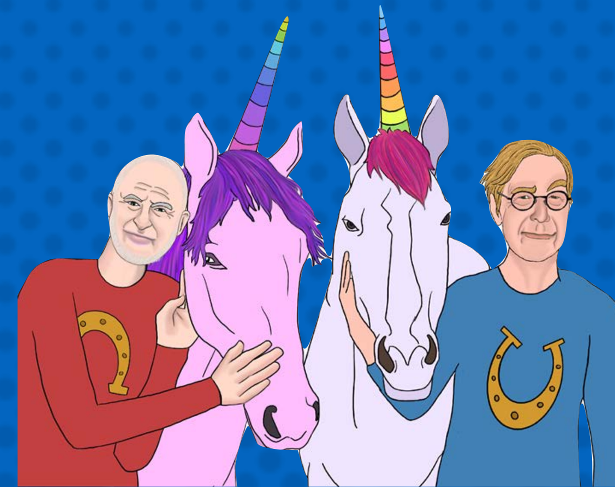
diversity: pick **k = 4** candidates, including 2 of each gender, and at least one per race

utility: maximize the total score of selected candidates

score = 373

	Male		Female	
White	A (99)	B (98)	C (96)	D (95)
Black	E (91)	F (91)	G (90)	H (89)
Asian	I (87)	J (87)	K (86)	L (83)

score = 372



Problem

picked the best White and male candidates (A, B) but did not pick the best Black (E, F), Asian (I, J), or female (C, D) candidates

Beliefs

scores are more informative within a group than across groups - **effort is relative to circumstance**

it is important to **reward effort**

[Yang, Gkatzelis, Stoyanovich (2019)]

r/ai center for responsible ai

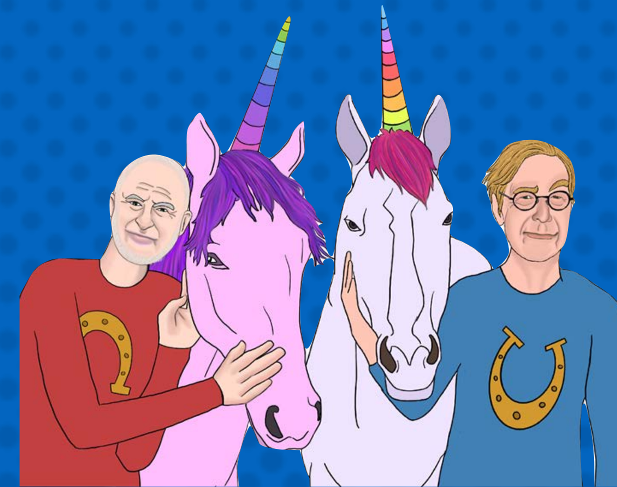
From beliefs to interventions

Fairness for female candidates **83 / 95 = 0.91**

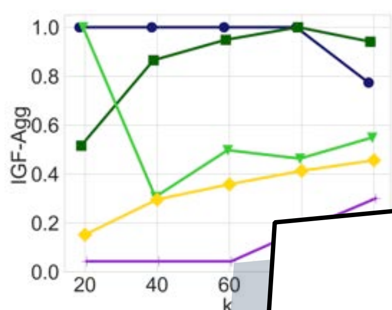
C	D	G	H	K	L
95	95	90	86	83	83

highest-scoring skipped

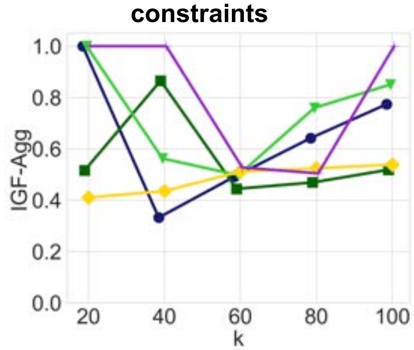
lowest-scoring selected



BEFORE: diversity constraints only



AFTER: diversity and fairness constraints



Beliefs

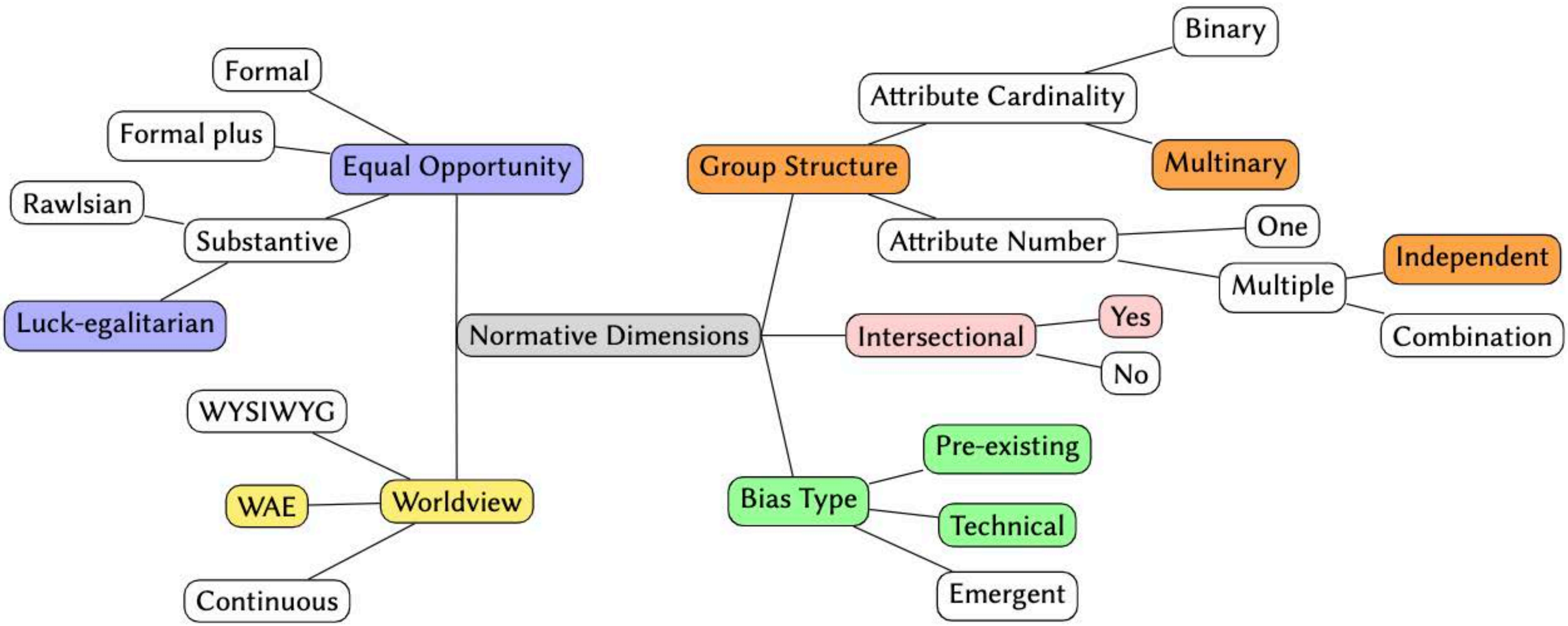
scores are more informative within a group than across groups - **effort is relative to circumstance**

it is important to **reward effort**

[Yang, Gkatzelis, Stoyanovich (2019)]

r/ai center for responsible ai

Normative mapping



Intersectional causal fairness

	gender	race	X	Y
B	m	w	6	12
C	m	**b**	5	9
D	**f**	w	6	8
E	m	w	4	7
F	**f**	**b**	3	6
K	**f**	a	5	5
L	m	**b**	1	3
O	**f**	w	1	1

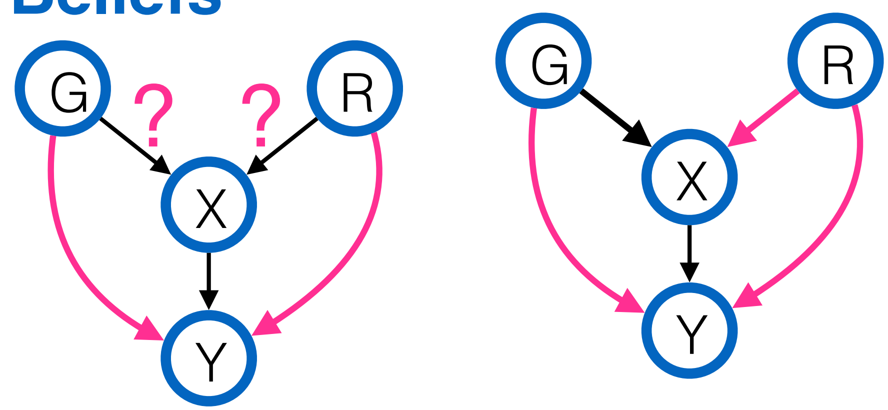
Goal

hire **k = 4** best-qualified candidates at a moving company

Problem

weight lifting ability is mapping to qualification score differently depending on gender

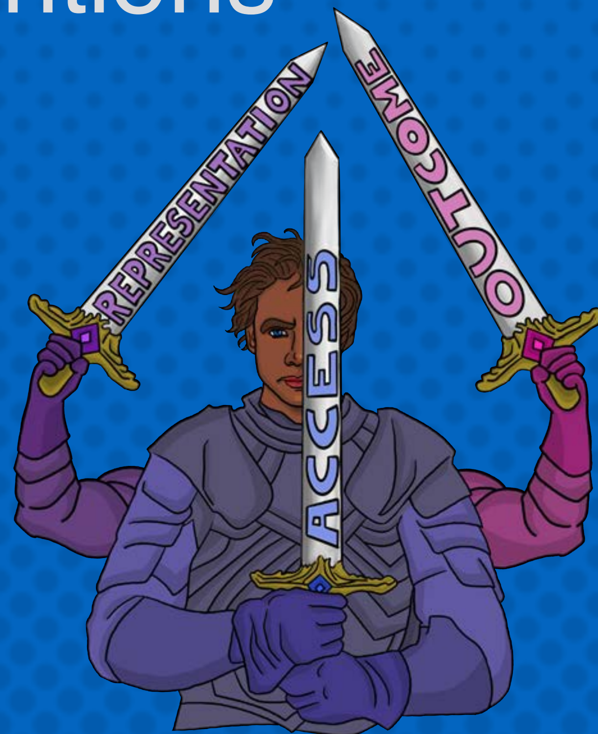
Beliefs



[Yang, Loftus, Stoyanovich (2021)]

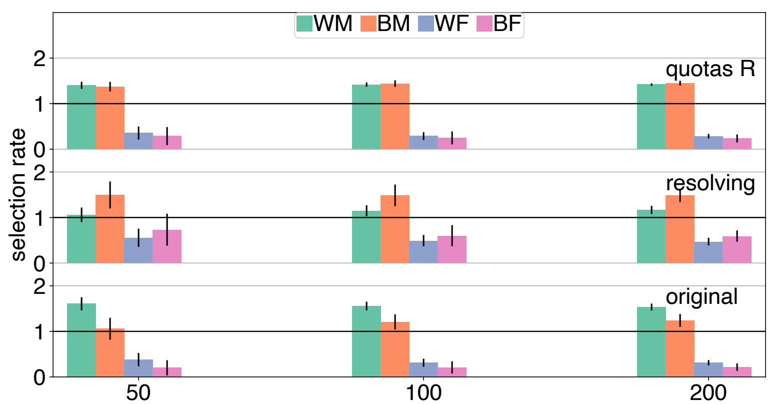
r/ai center for responsible ai

From beliefs to interventions



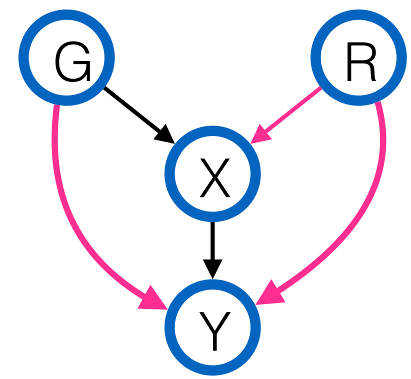
Idea: Compute **counterfactual scores**, treating each individual in the sample as though they had belonged to *one* intersectional group (e.g., Black women). Rank on those scores.

This process produces a **counterfactually fair ranking**.



Beliefs

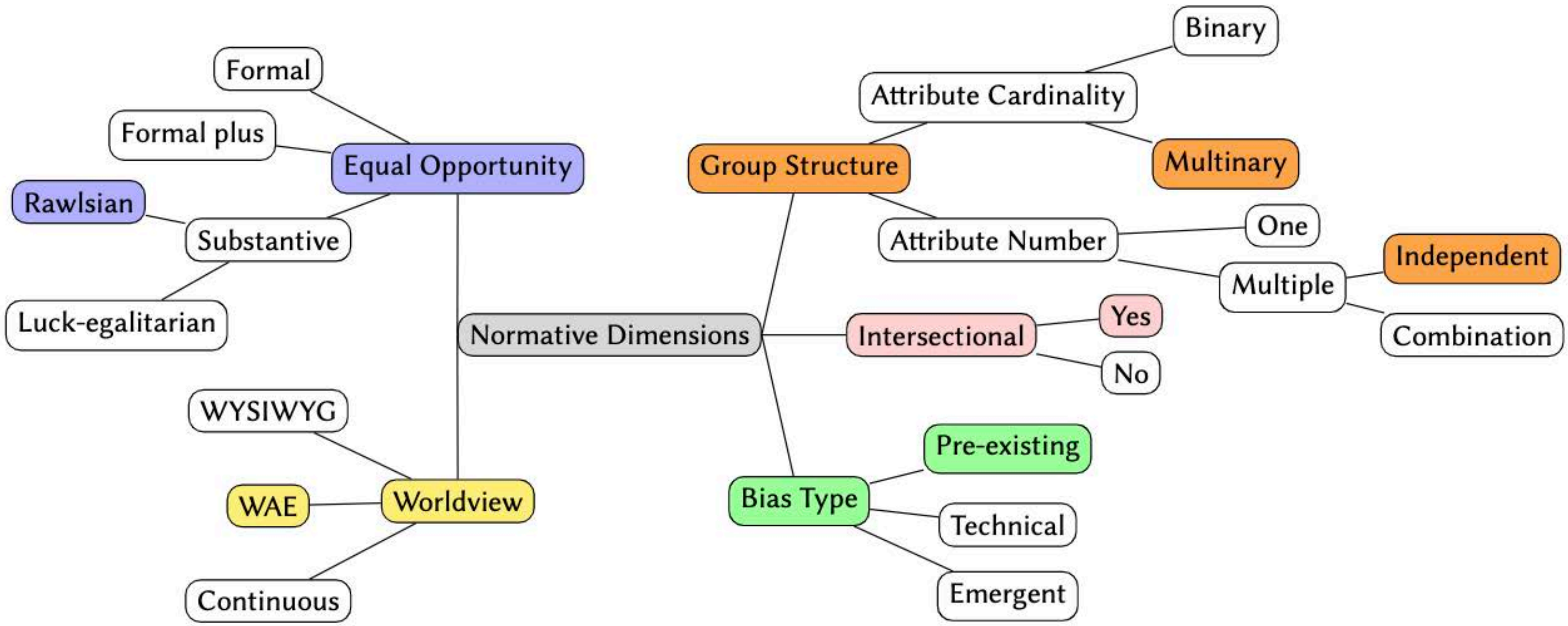
allow for resolving mediators



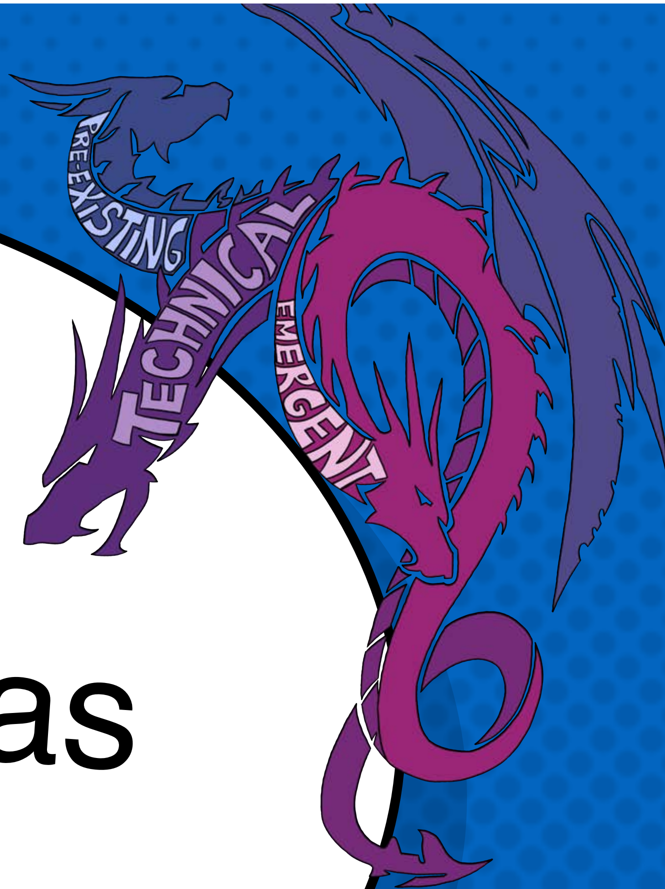
[Yang, Loftus, Stoyanovich (2021)]

r/ai center for responsible ai

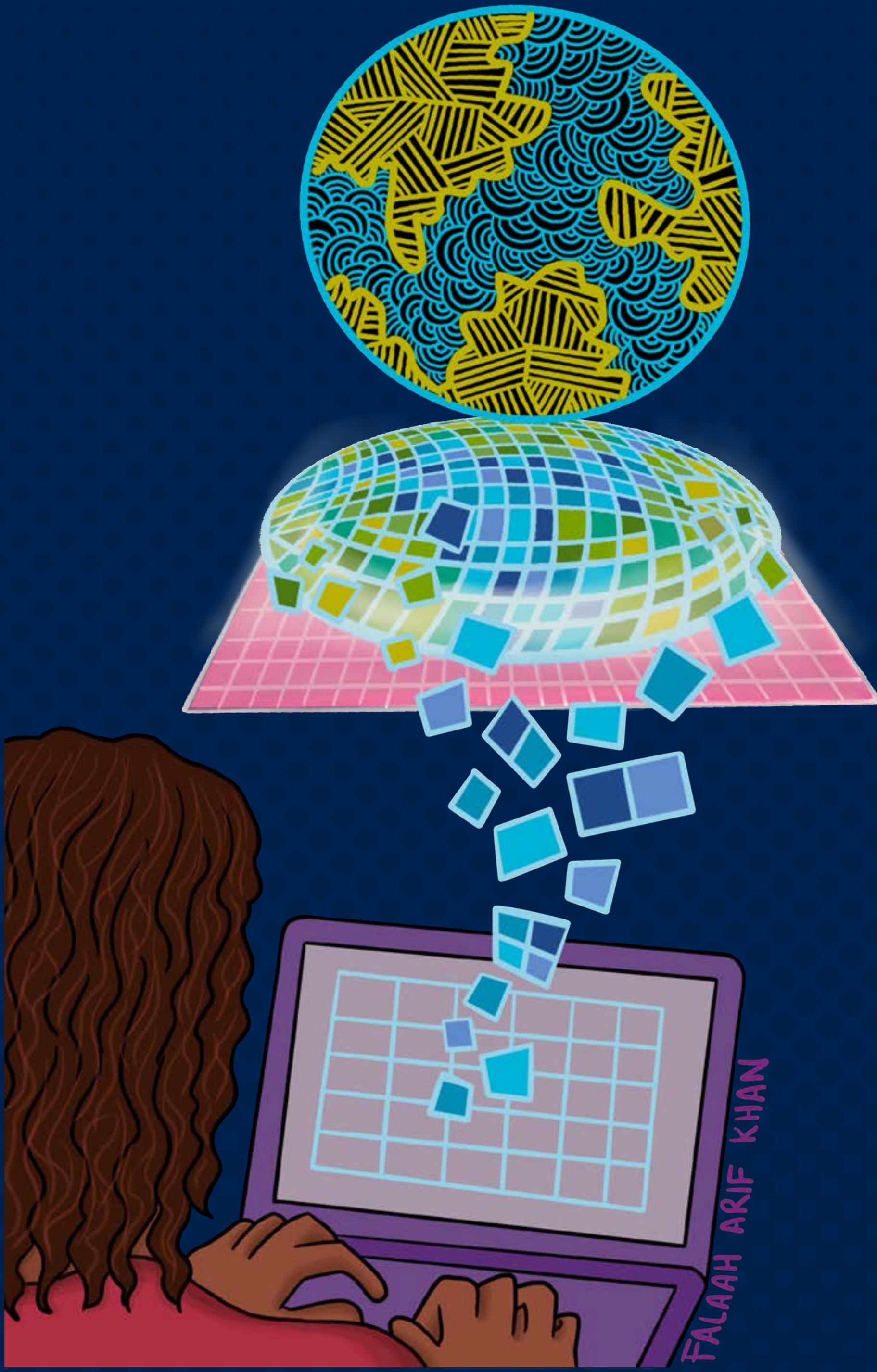
Normative mapping



technical bias



r/ai center for responsible ai

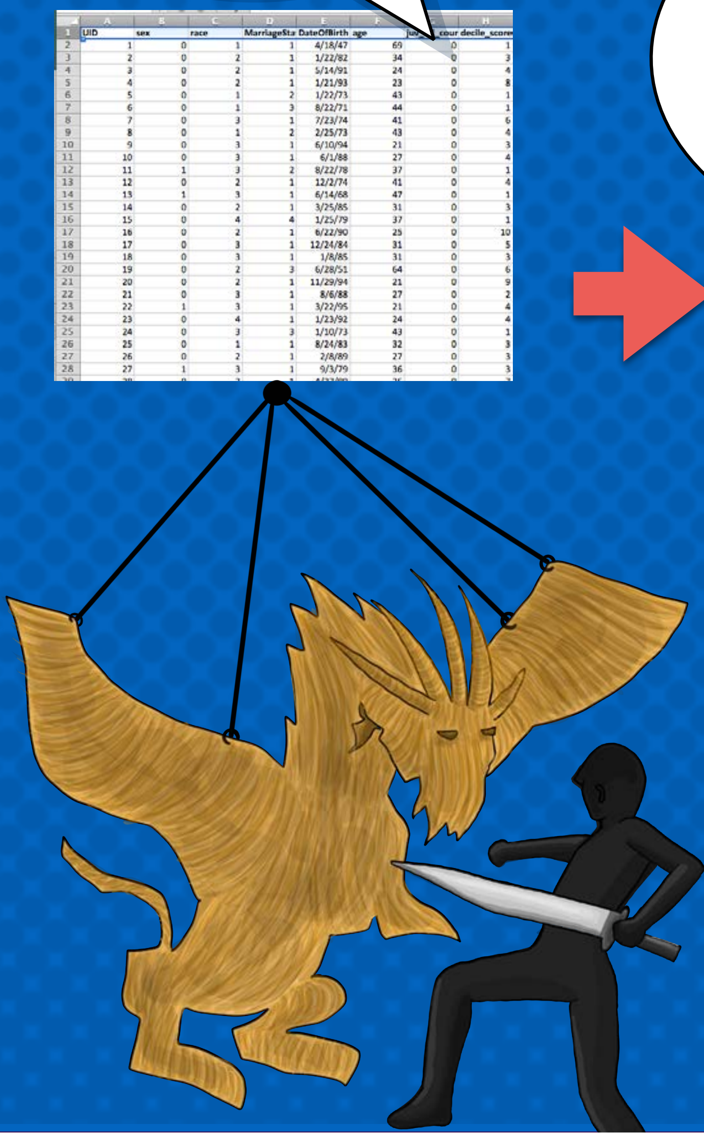


Technical bias may be introduced or exacerbated by the technical properties of an ADS

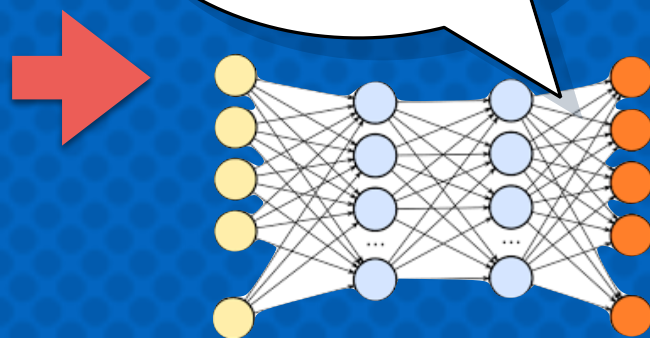
center for responsible ai

Fair-ML view

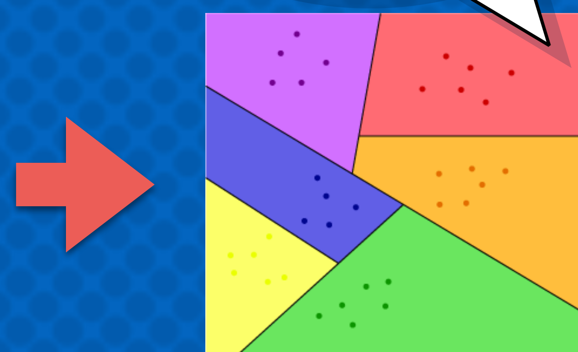
where did the data come from?



what happens inside the box?



how are results used?



center for responsible ai

Model development lifecycle

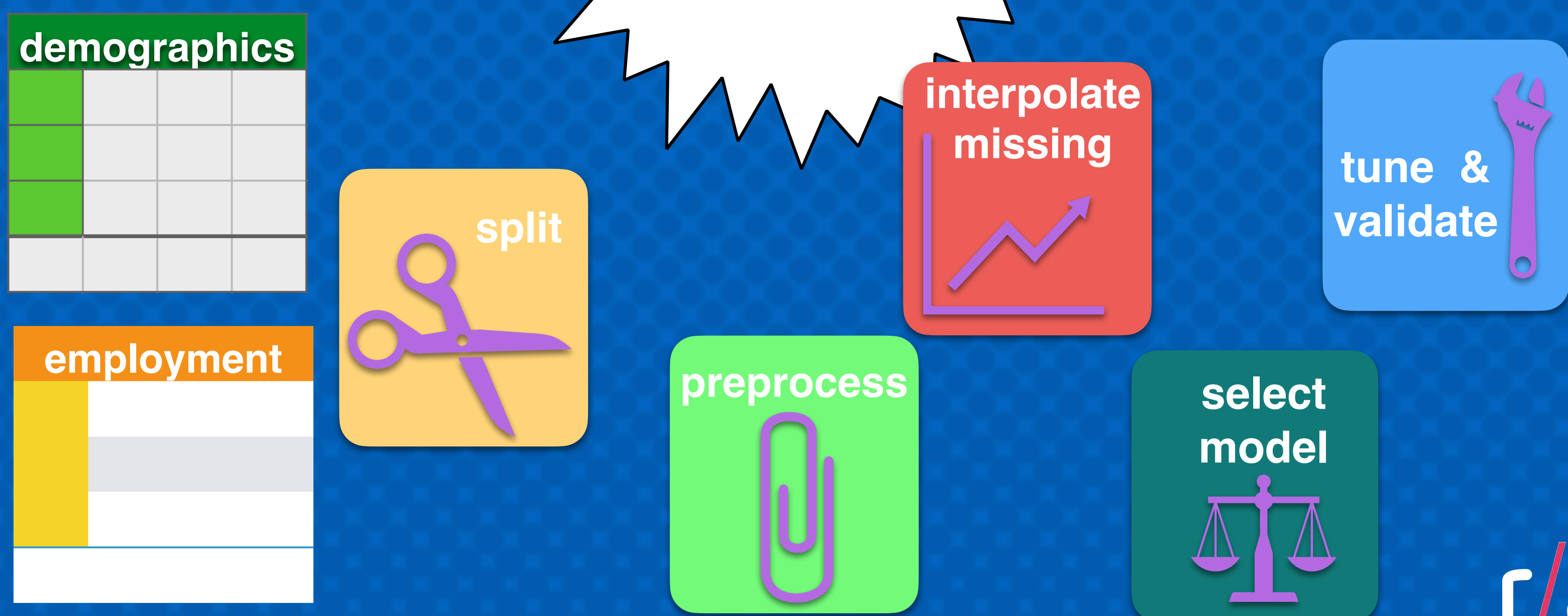
Goal

design a model to predict an appropriate level of compensation for job applicants

Problem

accuracy is lower for middle-aged women - **a fairness concern!**

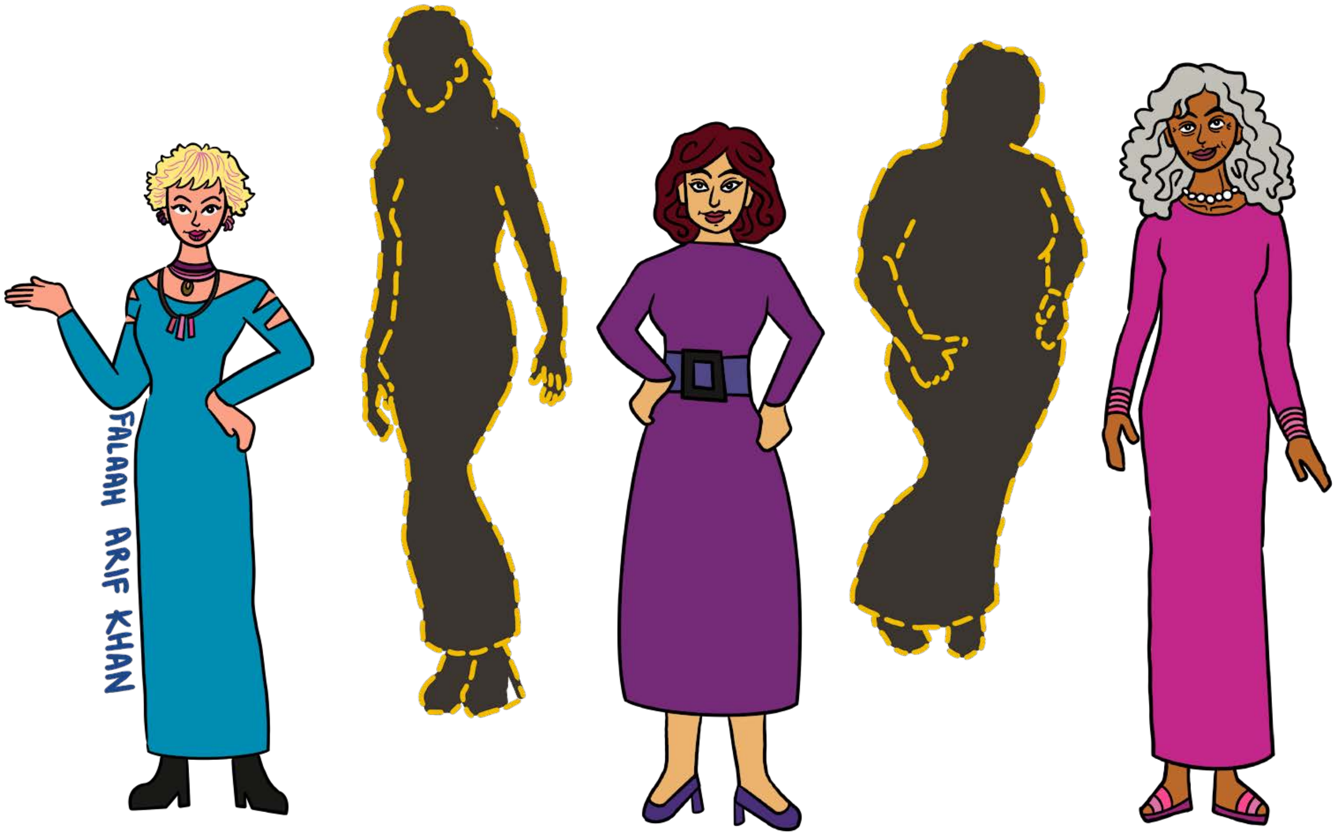
now what?



[Schelter, He, Khilnani, Stoyanovich (2020)]

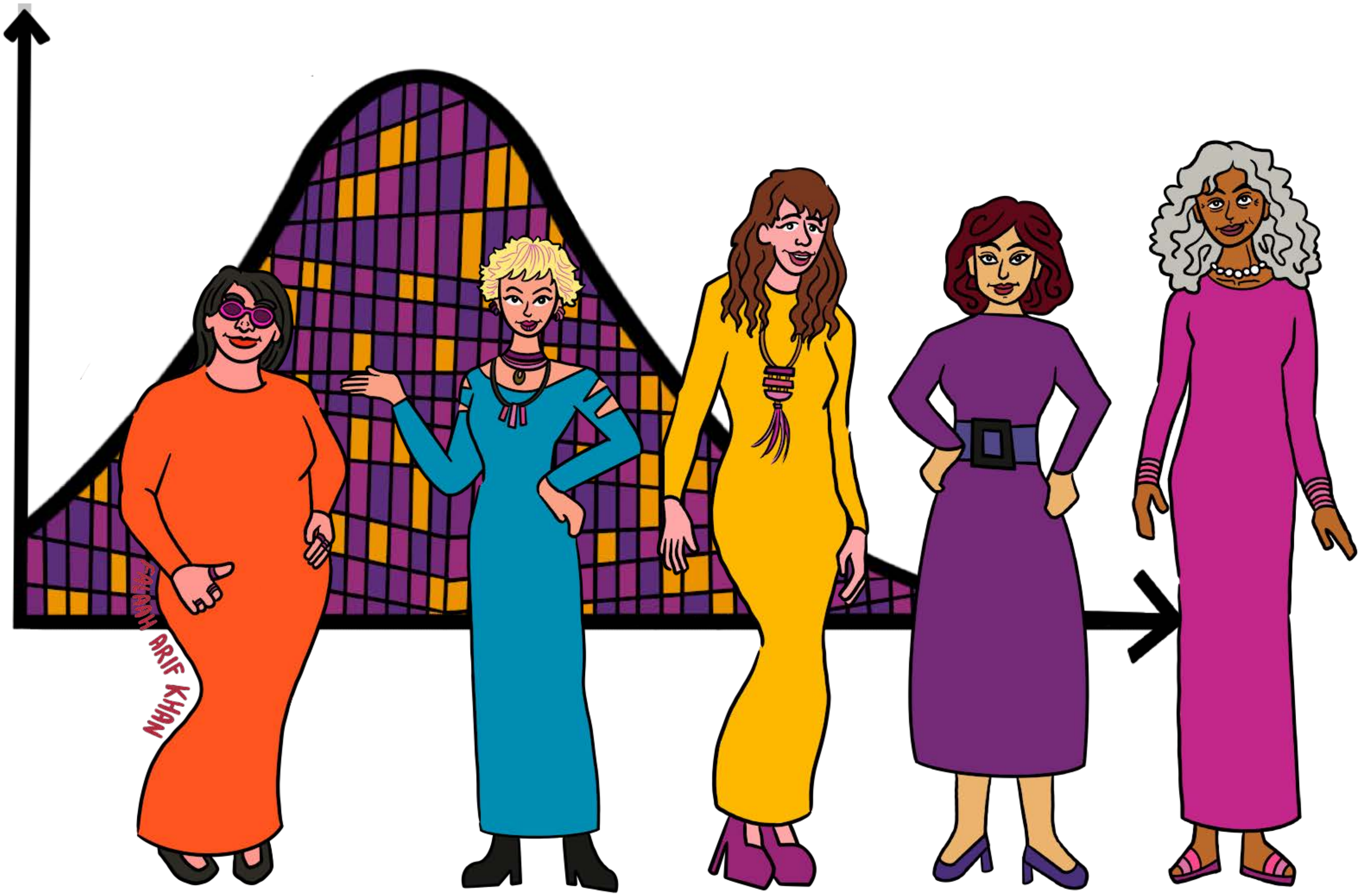
r/ai center for responsible ai

Missing values: Observed data



center for responsible ai

Missing values: Imputed distribution



r/ai center for responsible ai

Missing values: True distribution

center
for
responsible
ai

Missing value imputation

are values **missing at random** (e.g., *gender*, *age*, *years of experience*, *disability status* on job applications)?

are we ever interpolating **rare categories** (e.g., *Native American*)

are **all categories** represented (e.g., *non-binary gender*)?

Data filtering

"filtering" operations (like selection and join), **can arbitrarily change demographic group proportions**

select by zip code, country, years of C++ experience, others?

age_group	county
60	CountyA
60	CountyA
20	CountyA
60	CountyB
20	CountyB
20	CountyB

50% vs 50%

→

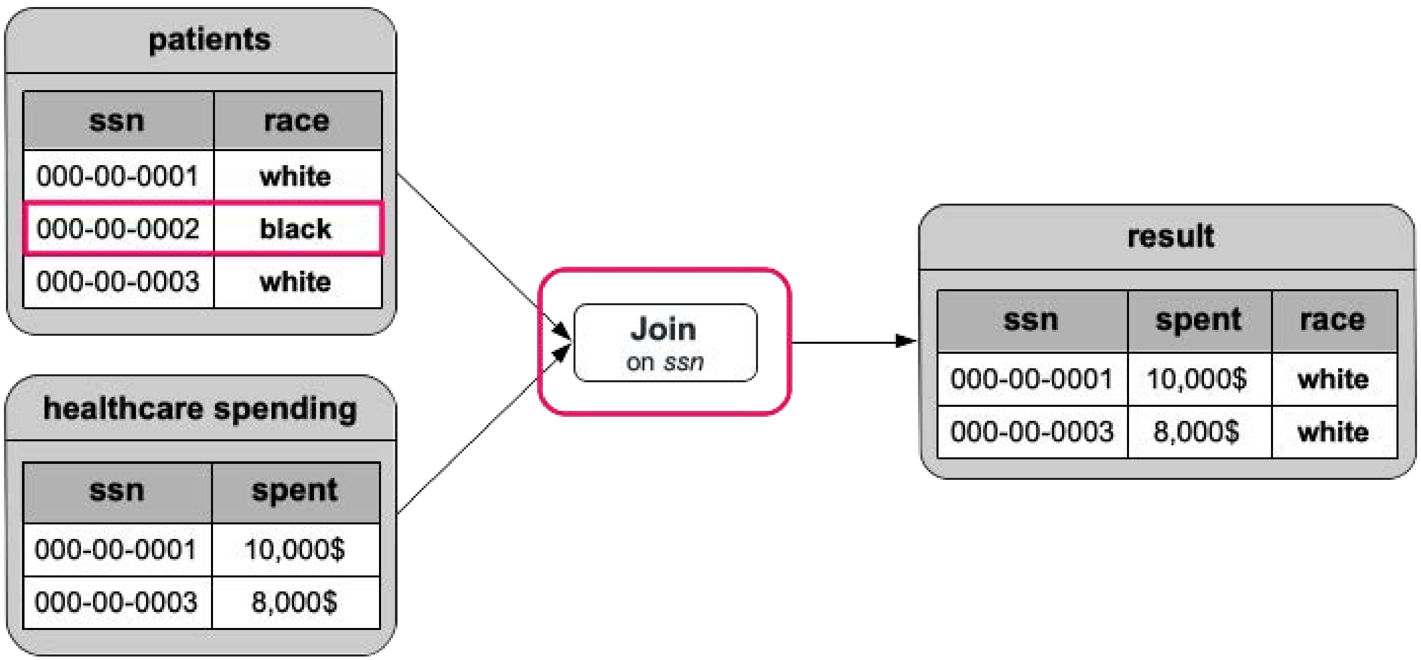
age_group	county
60	CountyA
60	CountyA
20	CountyA

66% vs 33%

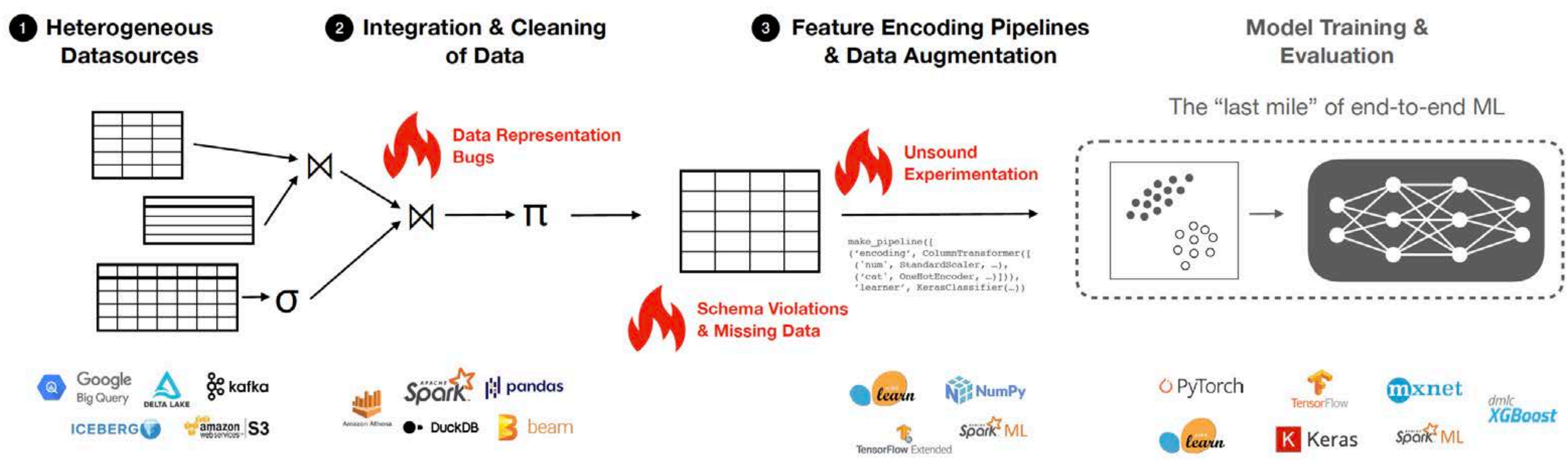
Data filtering

"filtering" operations (like selection and join), **can arbitrarily change demographic group proportions**

select by zip code, country, years of C++ experience, others?



ML pipelines in the wild



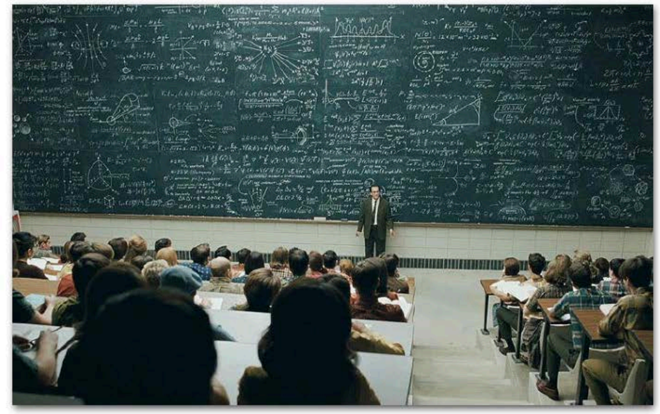
ML research vs. production

Research lab conditions

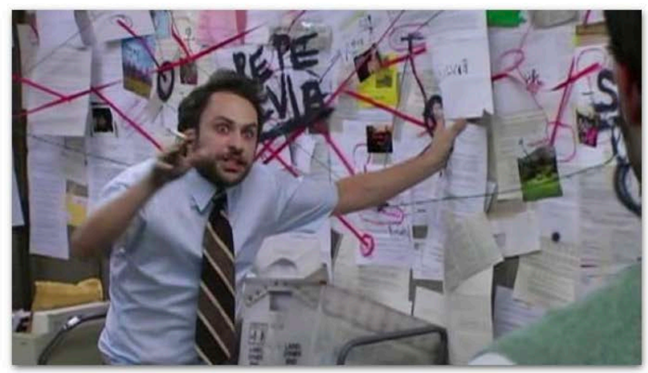
- mental model of working in a Jupiter notebook
- data is clean, static, well-understood, ML-ready
- developer has PhD in ML

Production conditions

- data produced continuously, never clean
- data originates from many sources, often not under developer's control
- model training is only one piece in a complex pipeline
- non-expert developers / operators / end-users
- **even experts can make mistakes!**



https://chrisguillebeau.com/files/2016/11/Mathboard.jpg



[Schelter et al. (2019)]

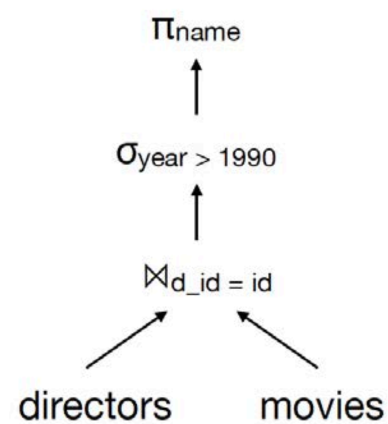
What makes inspection difficult?

Relational DBMS: explicit data model (relations), computations (queries) expressed declaratively in relational algebra

```
SELECT name
FROM directors
JOIN movies ON d_id = id
WHERE year > 1990
```

$$\pi_{name}\, \sigma_{year > 1990}\, (\text{directors} \bowtie_{d\_id = id} \text{movies})$$

Algebraic properties enable automatic inspection: identifying all input records that contributed to a query result (why-provenance)



ML pipelines: lack of unifying algebraic foundation for data preprocessing, different technologies "glued together"

[Schelter et al. (2019)]

center for responsible ai

The way forward

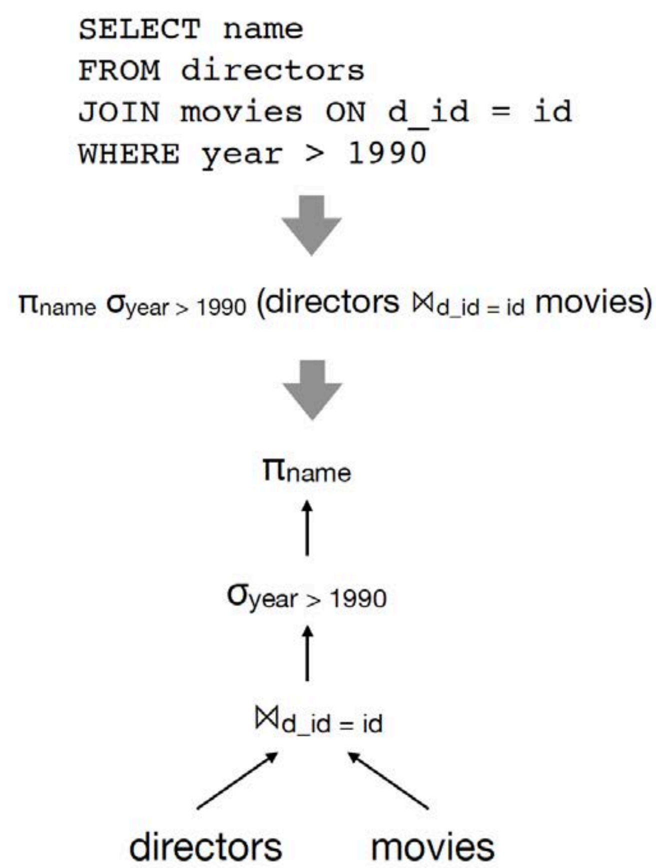
First approach: invent new holistic systems to regain control; would require rewriting all existing code

Second approach: manually annotating existing code; does not happen in practice

Our approach: retrofit inspection techniques into the existing data science landscape

Key observation: declarative specification of operations for preprocessing present in some popular ML libraries

- Pandas mostly applies relational operations
- Estimator / Transformer pipelines (scikit-learn / SparkML / Tensorflow Transform) offer nestable and composable way to declaratively specify feature transformations



[Grafberger, Stoyanovich, Schelter (2021)]

mlInspect: a data distribution debugger

Potential issues in preprocessing pipeline:

1. Join might change proportions of groups in data
2. Column 'age_group' projected out, but required for fairness
3. Selection might change proportions of groups in data
4. Imputation might change proportions of groups in data
5. 'race' as a feature might be illegal!
6. Embedding vectors may not be available for rare names!

Python script for preprocessing, written exclusively with native pandas and sklearn constructs

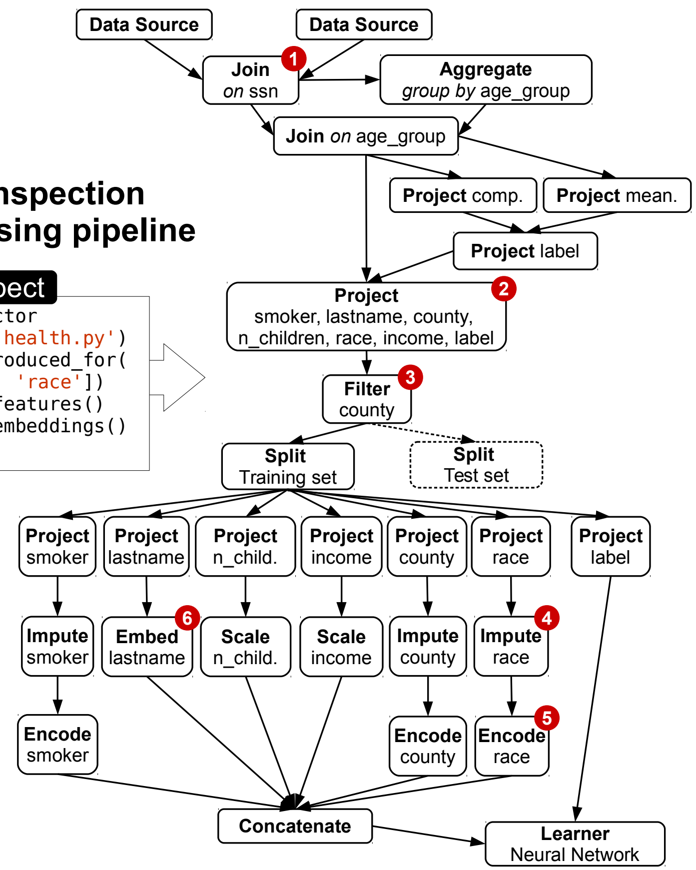
```
# load input data sources, join to single table
patients = pandas.read_csv(…)
histories = pandas.read_csv(…)
data = pandas.merge([patients, histories], on=['ssn'])

# compute mean complications per age group, append as column
complications = data.groupby('age_group')
 .agg(mean_complications=('complications','mean'))
data = data.merge(complications, on=['age_group'])

# Target variable: people with frequent complications
data['label'] = data['complications'] >
  1.2 * data['mean_complications']

# Project data to subset of attributes, filter by counties
data = data[['smoker', 'last_name', 'county',
             'num_children', 'race', 'income', 'label']]
data = data[data['county'].isin(counties_of_interest)]

# Define a nested feature encoding pipeline for the data
impute_and_encode = sklearn.Pipeline([
  (sklearn.SimpleImputer(strategy='most_frequent')),
  (sklearn.OneHotEncoder())])
featurisation = sklearn.ColumnTransformer(transformers=[
  (impute_and_encode, ['smoker', 'county', 'race']),
  (Word2VecTransformer(), 'last_name')
  (sklearn.StandardScaler(), ['num_children', 'income']])

# Define the training pipeline for the model
neural_net = sklearn.KerasClassifier(build_fn=create_model())
pipeline = sklearn.Pipeline([
  ('features', featurisation),
  ('learning_algorithm', neural_net)])

# Train-test split, model training and evaluation
train_data, test_data = train_test_split(data)
model = pipeline.fit(train_data, train_data.label)
print(model.score(test_data, test_data.label))
```

Declarative inspection of preprocessing pipeline

mlinspect

```
PipelineInspector
.on_pipeline('health.py')
 .no_bias_introduced_for(
 ['age_group', 'race'])
 .no_illegal_features()
 .no_missing_embeddings()
.verify()
```

Corresponding dataflow DAG for instrumentation, extracted by *mlinspect*



[Grafberger, Stoyanovich, Schelter (2021)]

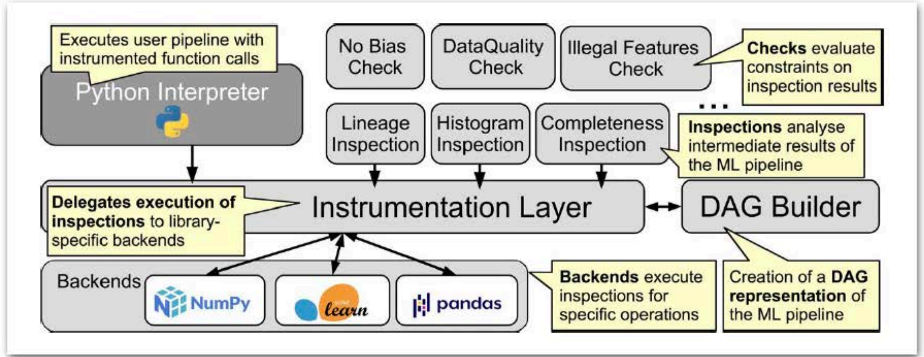
center for responsible ai

mlInspect: a data distribution debugger

mlInspect: library that instruments ML preprocessing code with custom inspections to analyze a single pipeline execution and detect potential issues

- works with “native” preprocessing pipelines (no annotation / manual instrumentation required) in pandas / sklearn / keras
- represents of preprocessing operations based on dataflow graph
- allows users to implement inspections as user-defined functions that are automatically applied to the inputs and outputs of operations

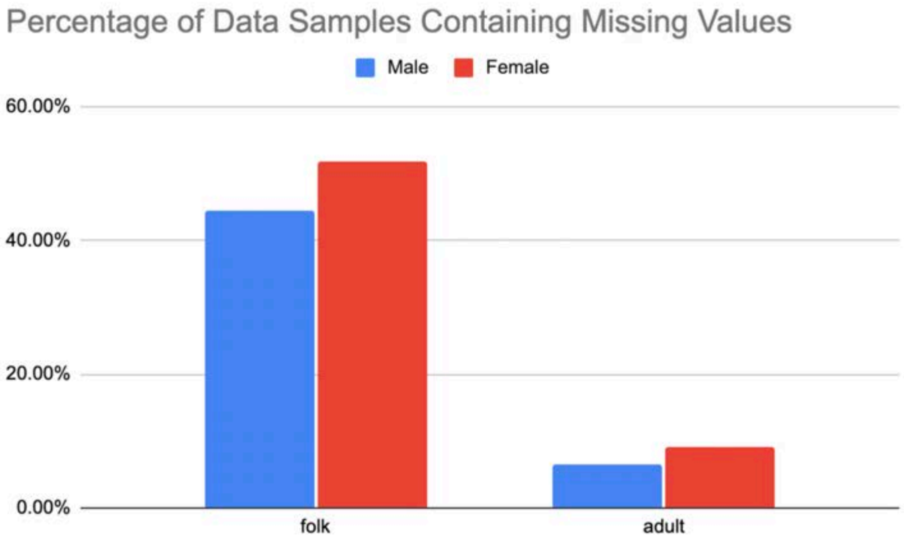
```
PipelineInspector
 .on_pipeline_from_py_file('healthcare.py')
  .expect_no_bias_introduced_for(['age_group', 'race'])
  .expect_no_use_of_illegal_features()
  .expect_no_missing_embeddings()
 .verify()
```



demo: https://surfdrive.surf.nl/files/index.php/s/ybriyzsdc6vcd2w 1:06-4:00
code: https://github.com/stefan-grafberger/mlinspect

Data quality and fairness

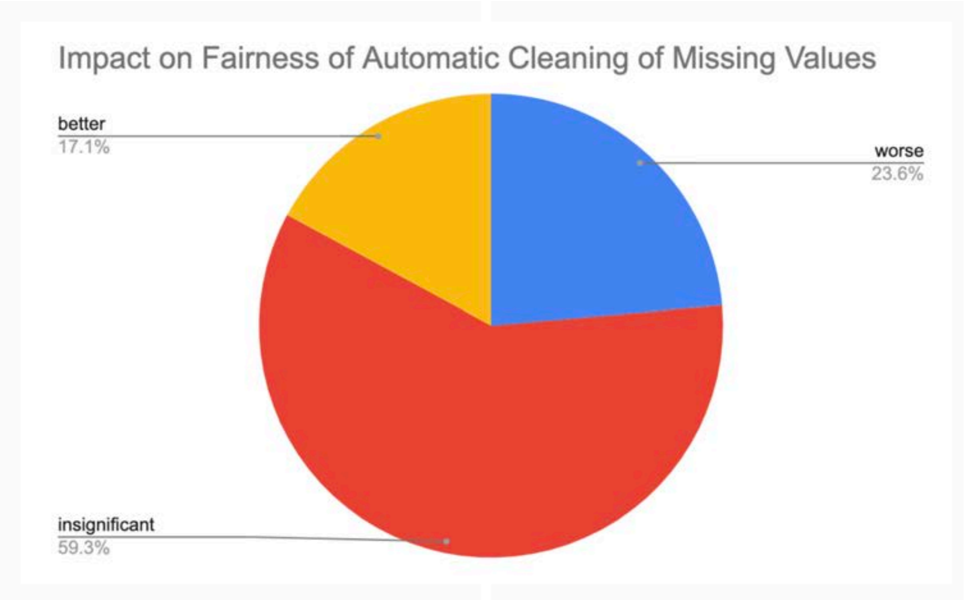
- poor-quality data can hurt ML model accuracy
- data from historically disadvantages groups may suffer from poorer quality
- systematic differences in data quality may hurt performance of predictors - a fairness concern
- **RQ1**: Does the incidence of data errors track demographic group membership in ML fairness datasets?



[Guha, Arif Khan, Stoyanovich, Schelter (2023)]

Data quality and fairness

- poor-quality data can hurt ML model accuracy
- data from historically disadvantages groups may suffer from poorer quality
- systematic differences in data quality may hurt performance of predictors - a fairness concern
- **RQ1**: Does the incidence of data errors track demographic group membership in ML fairness datasets?
- **RQ2**: Do common automated data cleaning techniques impact the fairness of ML models trained on the cleaned datasets?



[Guha, Arif Khan, Stoyanovich, Schelter (2023)]

Impact of automated data cleaning on fairness

Automated Data Cleaning Can Hurt Fairness in ML-based Decision Making

Shubha Guha
s.guha@uva.nl
University of Amsterdam

Falaah Arif Khan
fa2161@nyu.edu
New York University

Julia Stoyanovich
stoyanovich@nyu.edu
New York University

Sebastian Schelter
s.schelter@uva.nl
University of Amsterdam

ongoing work

model	auto-cleaning makes fairness worse	auto-cleaning makes fairness better	auto-cleaning makes fairness & accuracy better
xgboost	21.2% (45)	10.8% (23)	6.6% (14)
knn	24.5% (52)	13.7% (29)	11.8% (25)
log-reg	19.8% (42)	12.3% (26)	7.5% (16)

TABLE V
IMPACT OF AUTO-CLEANING ON ACCURACY AND FAIRNESS FOR DIFFERENT ML MODELS ON 212 CONFIGURATIONS IN TOTAL. WE LIST CASES WHERE FAIRNESS GETS WORSE, FAIRNESS GETS BETTER, AND WHERE BOTH FAIRNESS AND ACCURACY GET BETTER. AUTO-CLEANING IS MORE LIKELY TO WORSEN THAN TO IMPROVE FAIRNESS ACROSS ALL MODELS.

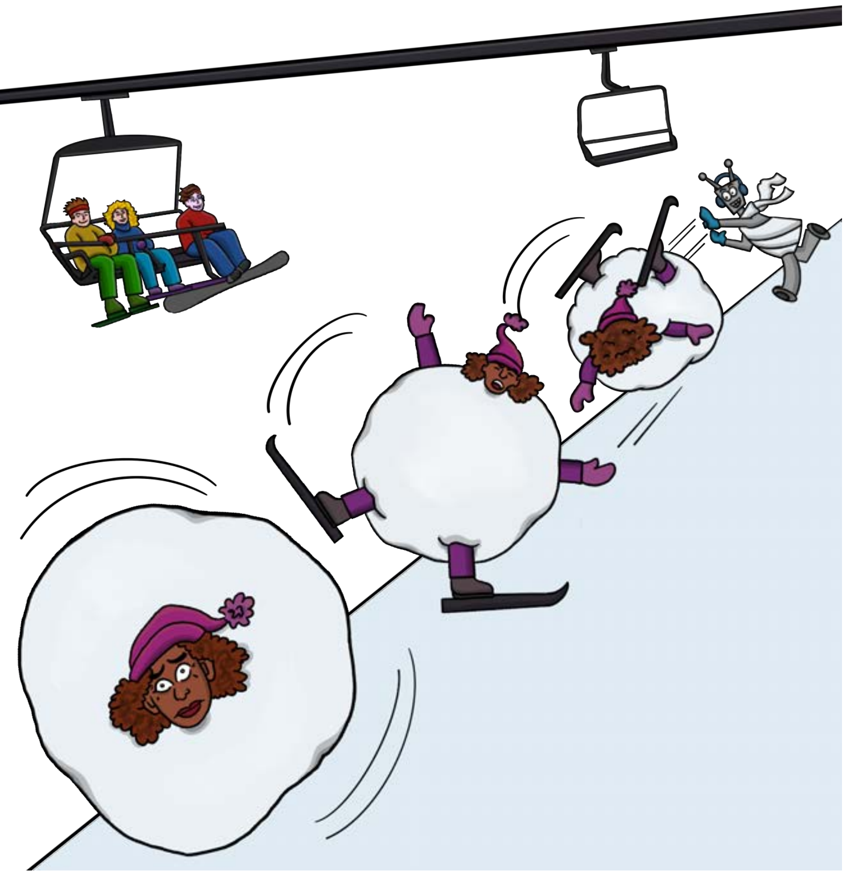
https://github.com/amsterdata/demodq

[Guha, Arif Khan, Stoyanovich, Schelter (2023)]

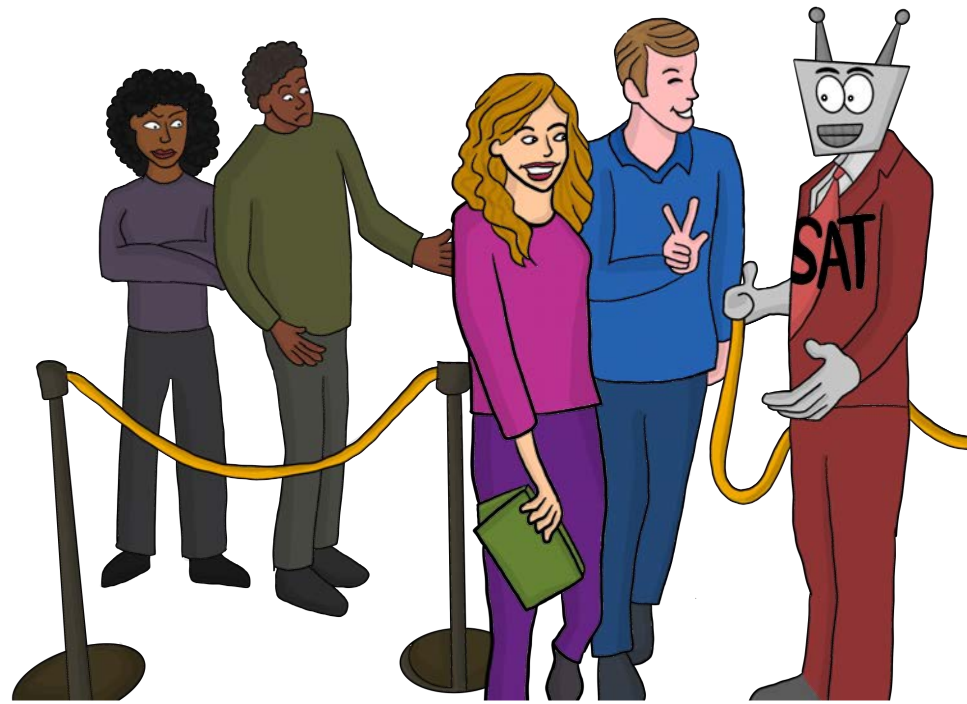
emergent bias

Example of emergent bias

snowball effect of privilege
and disadvantage



the circular problem of "merit"



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.

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

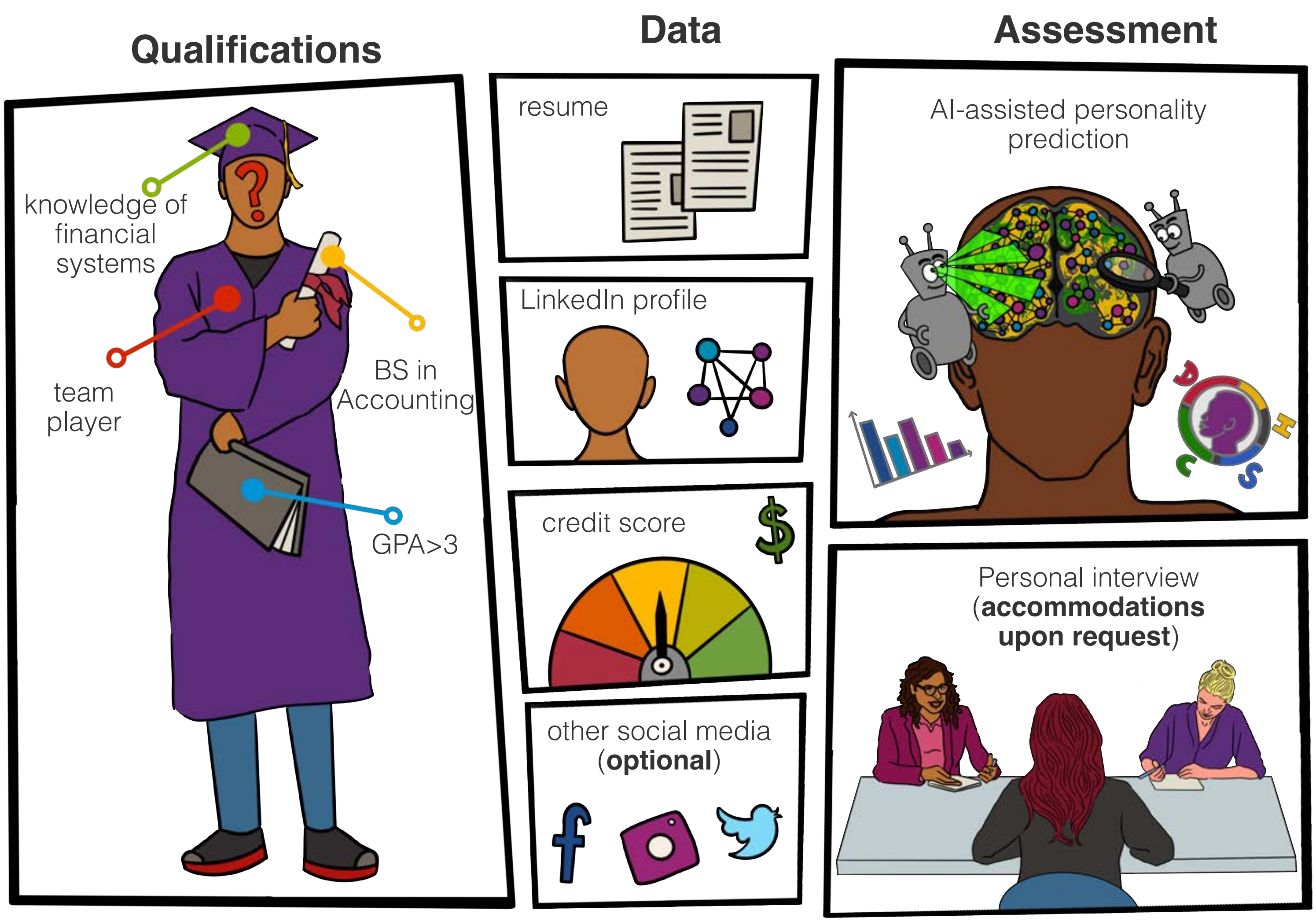
ACCOUNTANT

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.

Anatomy of a job posting label



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

r/ai center for responsible ai

wrapping up

Nuance, please!

Responsible Data Science @ NYU

FAIRNESS | DATA SCIENCE LIFECYCLE | DATA PROTECTION | TRANSPARENCY AND INTERPRETABILITY

Previous module:
FAIRNESS

WEEK 5

WEEK 6

WEEK 7

Next module:
DATA PROTECTION

Data Science Lifecycle

Lecture: Taming technical bias

Topics:

- Types of technical bias
- Data distibution debugging

Reading: See Responsibility in the Data Science Lifecycle

Lab: mlinspect

Responsible Data Science: Charting New Pedagogical Territory

NYU Center for Data Science Feb 17, 2020 · 4 min read

In response to the dearth of scholarship surrounding responsible data science (RDS), NYU CDS faculty are paving the way with a course dedicated to RDS and the publication of their pedagogical methodology.

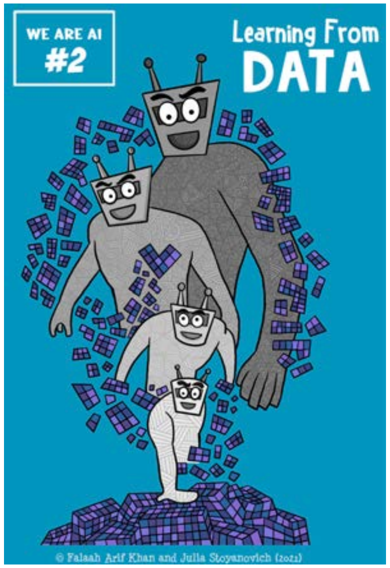
https://dataresponsibly.github.io/rds

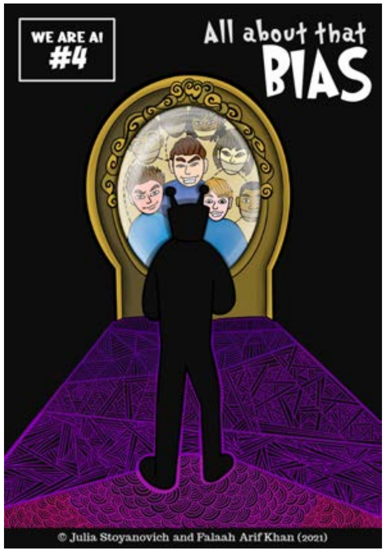
r/ai

https://dataresponsibly.github.io/we-are-ai/

r/ai center for responsible ai

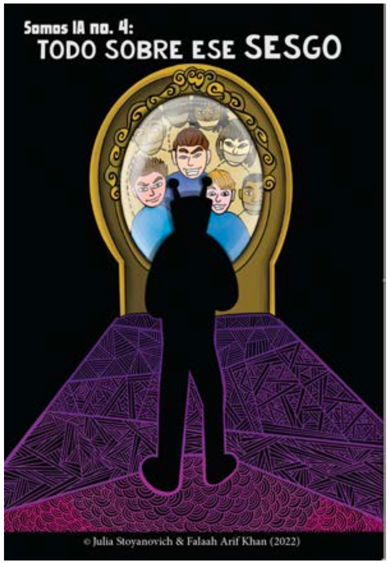
We are AI comics

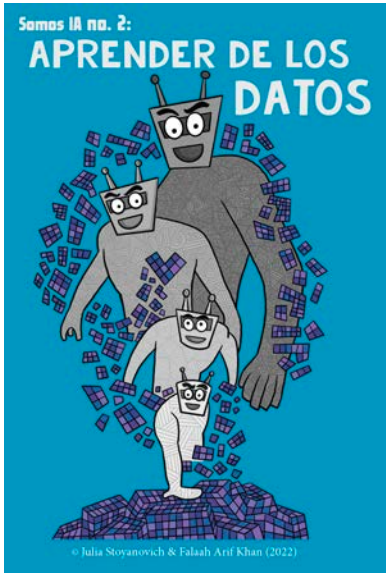


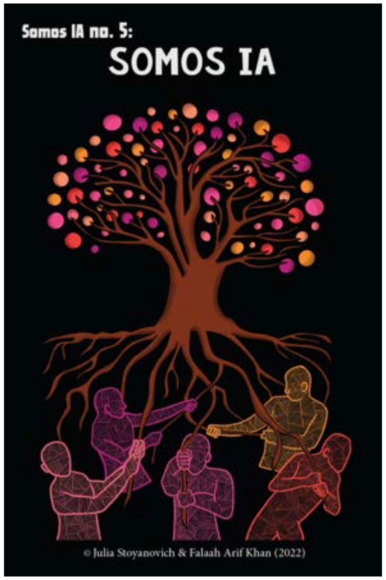


dataresponsibly.github.io/we-are-ai/comics

We are AI comics: in Spanish



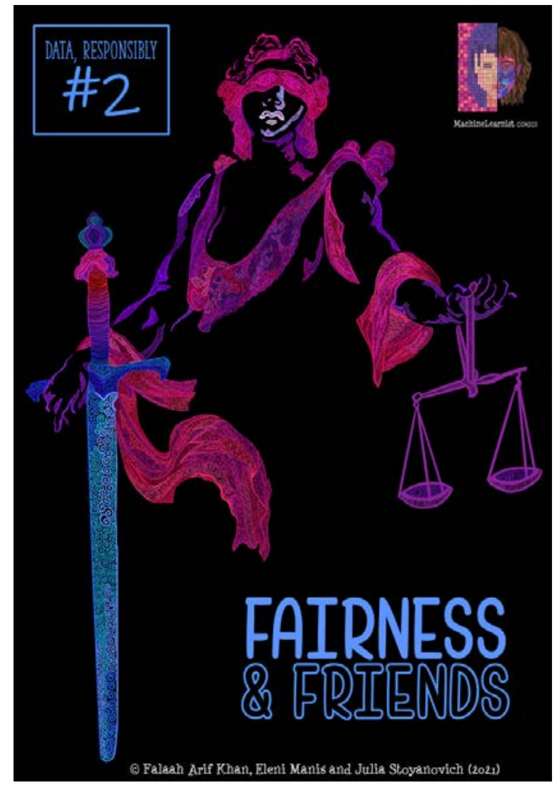




dataresponsibly.github.io/we-are-ai/comics

r/ai center for responsible ai

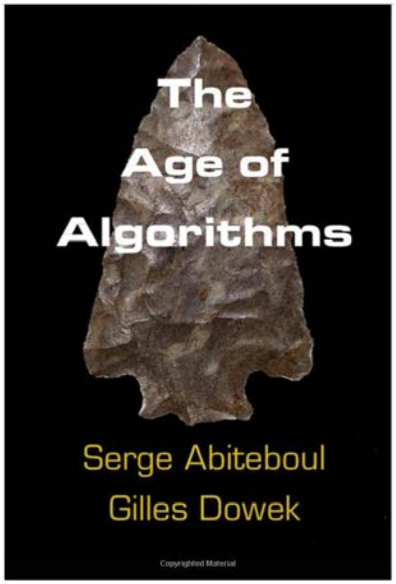
Scientific comics

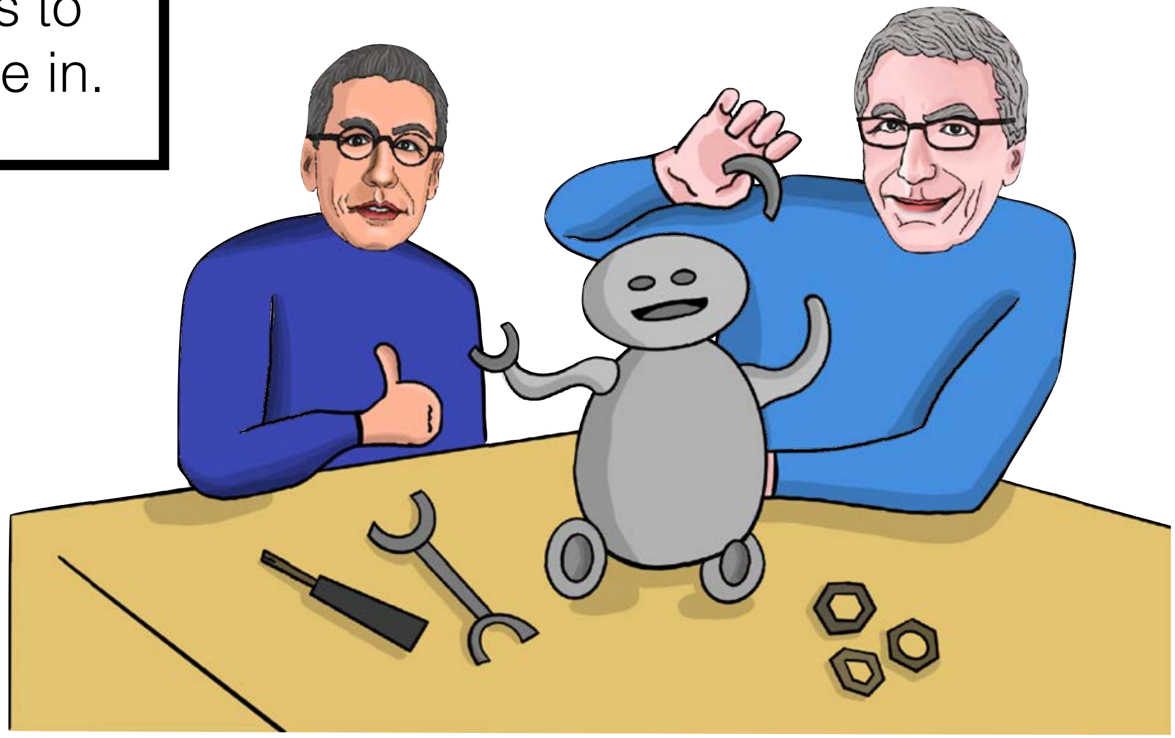


dataresponsibly.github.io/comics

AI is what *WE* make it!

Creations of the human spirit, **algorithms - and AI - are what we make them**. And they will be what we want them to be: it's up to us to choose the world we want to live in.





Thank you!

Julia Stoyanovich
New York University
USA

Serge Abiteboul
Inria & ENS Paris
France

Bill Howe
University of Washington
USA

H.V. Jagadish
University of Michigan
USA

Sebastian Schelter
University of Amsterdam
The Netherlands

Research supported in part by NSF Grants No. 1934464, 1934565, 1934405, 1926250, 1741022, 1740996, 1916505, by Microsoft, and by Ahold Delhaize