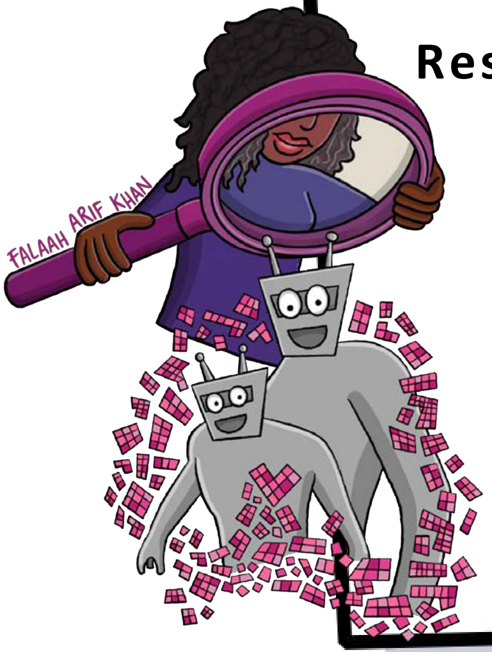
Introduction



Follow the data!

Responsible AI starts with responsible data management

Julia Stoyanovich
New York University
stoyanovich@nyu.edu

NYU

r/ai center for responsible ai

U.S. National Science Foundation

NYU Center for Responsible AI

r/ai center for responsible ai

How it started

October 2015

Le Monde

Plaidoyer pour une analyse « responsable » des données

Face aux risques d'atteinte à la vie privée, les chercheurs en informatique Serge Abiteboul et Julia Stoyanovich plaident pour une collecte et une analyse des données impartiales, transparentes et accessibles à tous.

Publié le 06 octobre 2015 à 15h41, modifié le 19 octobre 2015 à 16h16 | Lecture 5 min.

Serge Abiteboul and Julia Stoyanovich

NOVEMBER 20, 2015

DATA, RESPONSIBLY

November 2015

Big Data

(This blog post is an extended version of an October 12, 2015 Le Monde op-ed article (in French)) Our society is increasingly relying on algorithms in all aspects of its operation. We trust algorithms not only to help carry out routine tasks, such as accounting and automatic manufacturing, but also to make decisions on our [...]

Read more →

Association For Computing Machinery acm ACM SIGMOD Blog

r/ai center for responsible ai

How it started

July 2016

Data, Responsibly

Organizers

Serge Abiteboul (ENS – Cachan, FR)
Gerome Miklau (University of Massachusetts – Amherst, US)
Julia Stoyanovich (Drexel Univ. – Philadelphia, US)
Gerhard Weikum (MPI für Informatik – Saarbrücken, DE)

r/ai center for responsible ai

How it started

August 2017

Int. No. 1696

By Council Member Vacca

A Local Law to amend the administrative code of the city of New York, in relation to automated processing of data for the purposes of targeting services, penalties, or policing to persons

Be it enacted by the Council as follows:

Section 1. Section 23-502 of the administrative code of the city of New York is amended to add a new subdivision g to read as follows:

g. Each agency that uses, for the purposes of targeting services to persons, imposing penalties upon persons or policing, an algorithm or any other method of automated processing system of data shall:

1. Publish on such agency's website, the source code of such system; and

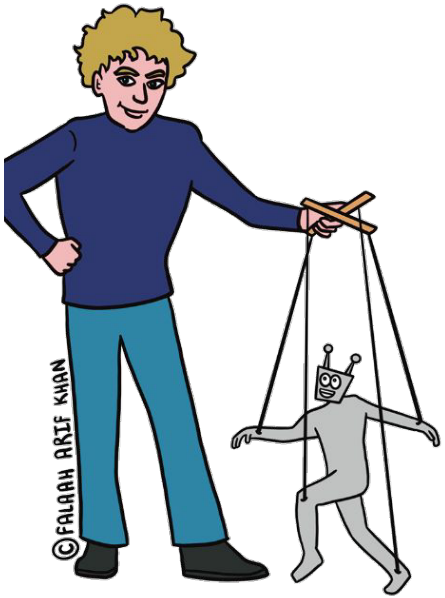
2. Permit a user to (i) submit data into such system for self-testing and (ii) receive the results of having such data processed by such system.

§ 2. This local law takes effect 120 days after it becomes law.

MAJ
LS# 10948
8/16/17 2:13 PM

October 2017

r/ai center for responsible ai



Policy

r/ai center for responsible ai

https://science.house.gov/2025/4/deepseek-a-deep-dive

Testimony of Dr. Julia Stoyanovich

Associate Professor of Computer Science & Engineering and of Data Science,
Director of the Center for Responsible AI at New York University

DeepSeek: A Deep Dive

Hearing of the Committee on Science, Space and Technology of the U.S. House of Representatives, Research and Technology Subcommittee

April 8, 2025

Note: ChatGPT 4o was used for stylistic purposes when drafting this testimony.

Thank you for the opportunity to testify today, on the topic of national security and technological implications of DeepSeek—a family of AI models developed in the People's Republic of China.

Launched on January 10, 2025, the DeepSeek AI assistant quickly rose to the top of the U.S. Apple App Store, as American consumers embraced it over competitors like *ChatGPT*.[1] The *DeepSeek-V3* and *DeepSeek-R1* models are now readily accessible to developers and researchers on Microsoft's Azure AI Foundry[2] and GitHub[3].

DeepSeek's large language models (LLMs) perform comparably to leading U.S.-based models while requiring significantly fewer resources—including hardware, power, and data annotation labor—to build.[4] And while LLM technology was already available to American consumers, developers, and researchers, DeepSeek's models introduced high-performing, cost-effective alternatives. Their release has acted as a catalyst for the U.S. AI industry—intensifying competition, prompting exploration of more efficient methods, and encouraging greater openness. By showing that advanced models can be built with relatively modest resources, DeepSeek has helped shift the U.S. AI landscape toward more accessible and collaborative innovation.

United Nations | Office for Disarmament Affairs

sipri

STOCKHOLM INTERNATIONAL PEACE RESEARCH INSTITUTE

SIPRI and UNODA launch joint initiative on responsible innovation in AI for peace and security

3 April 2023

This month SIPRI and the United Nations Office for Disarmament Affairs (UNODA) launched a three-year joint initiative on responsible innovation in artificial intelligence (AI) for peace and security. The initiative, which is funded by a decision of the Council of the European Union (Council Decision (CFSP) 2022/2269 of 18 November 2022), aims to support greater engagement from the civilian AI community in mitigating the risks that the misuse of civilian AI technology can pose to international peace and security.

Government of Canada | Gouvernement du Canada

CAN/ASC - EN 301 549:2024 - Accessibility requirements for ICT products and services (EN 301 549:2021, IDT)

CAN/ASC - EN 301 549:2024 - Accessibility requirements for ICT products and services (EN 301 549:2021, IDT)

Areas of focus

There are common areas where people with disabilities may experience barriers to accessibility in information and communication technology products and services. These include, but are not limited to:

- websites
- software
- electronic devices
- mobile apps

The technical committee on Accessibility Requirements for Information and Communication Technology Products and Services worked towards the adoption of the European harmonized standard "EN 301 549, Accessibility requirements for ICT products and services" in its entirety. The adoption of this standard as a National Standard of Canada in May 2024 represents a fundamental step in accessible Canadian procurement advancement.

This standard is a National Standard of Canada.

- This standard provides key requirements and best practices to help government departments and federally regulated entities as they continue their journey to improve accessibility in this priority area.
- The final standard was published on May 31, 2024.

Read the standard

r/ai center for responsible ai

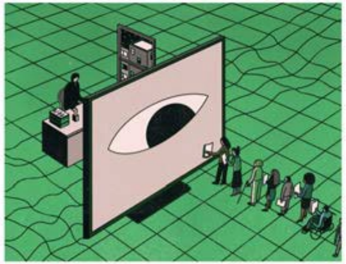
The New York Times

OPINION

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.

March 17, 2021



By Alexandra Reeve Givens, Hilke Schellmann and Julia Stoyanovich

Ms. Givens is the chief executive of the Center for Democracy & Technology. Ms. Schellman

THE WALL STREET JOURNAL.

BUSINESS | JOURNAL REPORTS: LEADERSHIP

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

As more companies use artificial intelligence in their hiring decisions, here's one way to make the system more transparent

By Julia Stoyanovich

Updated Sept. 22, 2021 11:00 am ET

Labels that explain a hiring process that uses AI could allow job seekers to opt out if they object to the

The New York Times

A Hiring Law Blazes a Path for A.I. Regulation

New York City's pioneering, focused approach sets rules on how companies use the technology in work force decisions.

r/ai center for responsible ai

NYU TANDON SCHOOL OF ENGINEERING

Statement of Julia Stoyanovich

Associate Professor of Computer Science & Engineering and of Data Science, Director of the Center for Responsible AI at New York University

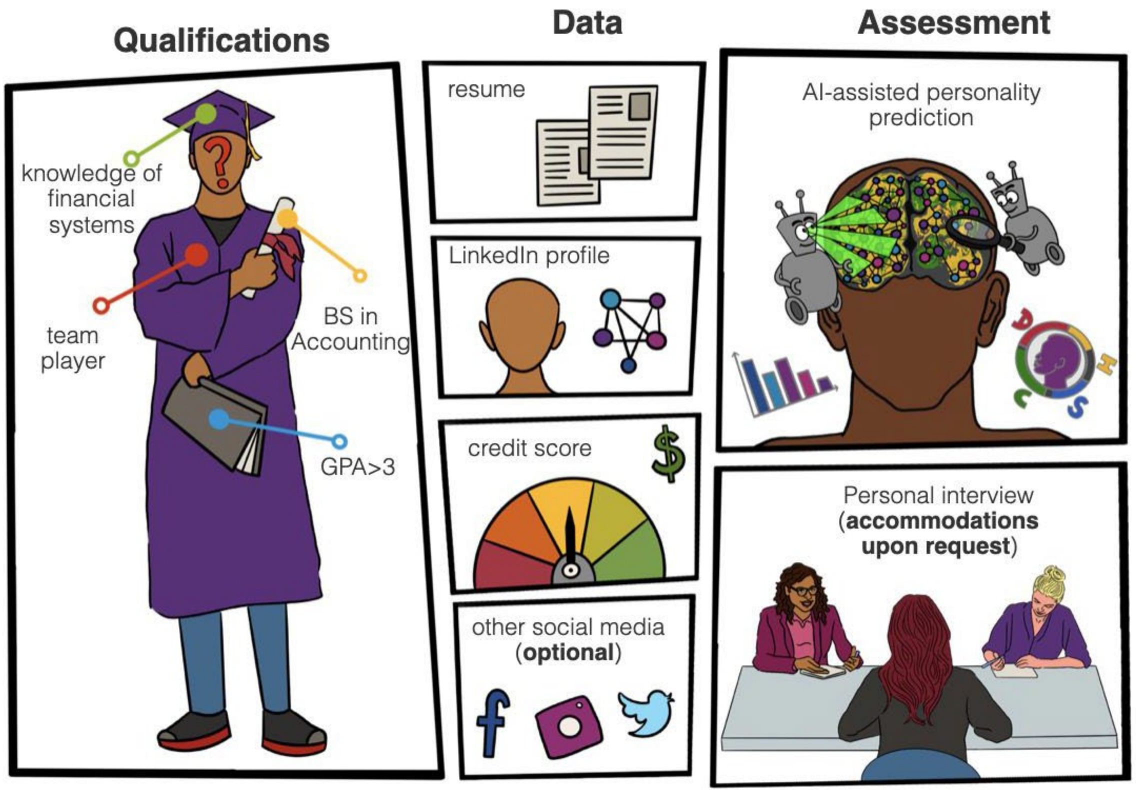
U.S. Senate AI Insight Forum: High Impact AI

November 1, 2023

Leader Schumer, Senators Rounds, Heinrich, and Young, thank you for inviting me to participate in this important forum! I am an associate professor of Computer Science & Engineering and of Data Science, and the founding Director of the Center for Responsible AI at New York University. My academic research is focused on AI and data engineering systems, and on how to incorporate legal requirements and ethical norms into the way these systems are designed, developed and used.[1] I teach responsible AI to students[2], practitioners in industry and government[3], and members of the public[4]. And I have been deeply involved in AI governance and regulation in New York City[5], New York State[6], and elsewhere, since 2017.

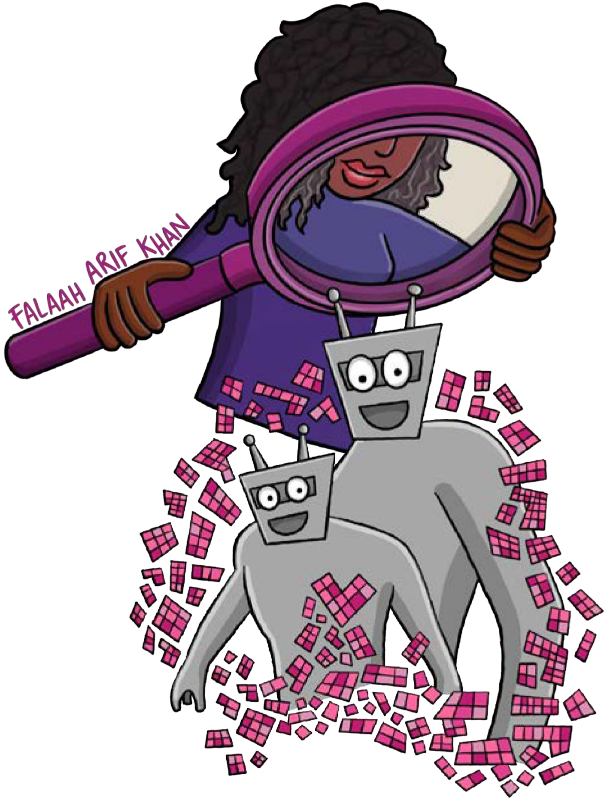
r/ai center for responsible ai

“Nutritional labels” for job postings



center for responsible ai

Responsible AI is about …

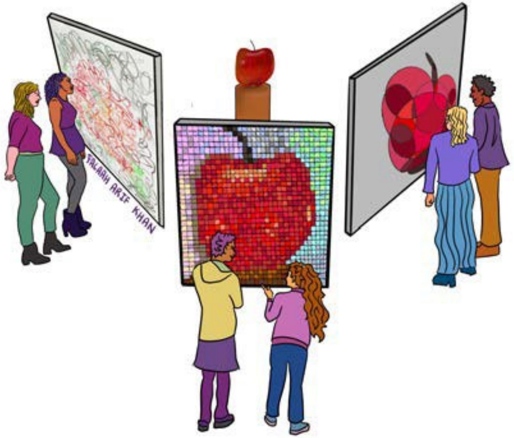


… exposing the knobs of responsibility to people

r/ai center for responsible ai

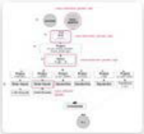
Research

Data-centric AI & responsible data management



Automated Data Cleaning Can Hurt Fairness in Machine Learning-based Decision Making
Shubha Guha, Falaah Arif Khan, Julia Stoyanovich, and Sebastian Schelter
In Proceedings of the 39th International Conference on Data Engineering, ICDE 2023

CITE PDF



Fairness-Aware Instrumentation of Preprocessing Pipelines for Machine Learning
Ke Yang, Biao Huang, Julia Stoyanovich, and Sebastian Schelter
In Proceedings of the Workshop on Human-In-the-Loop Data Analytics, HILDA at SIGMOD 2020

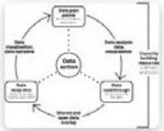
CITE PDF

Taming Technical Bias in Machine Learning Pipelines
Sebastian Schelter, and Julia Stoyanovich
IEEE Data Eng. Bull. 2020

CITE PDF

Responsible Data Management
Julia Stoyanovich, Serge Abiteboul, Bill Howe, H. V. Jagadish, and Sebastian Schelter
Communications of the ACM 2022

CITE PDF



Developing data capability with non-profit organisations using participatory methods
Anthony McCosker, Xiaofang Yao, Kath Albury, Alexia Maddox, Jane Farmer, and Julia Stoyanovich
Big Data & Society 2022

CITE PDF

NYU TANDON SCHOOL OF ENGINEERING

ФАКУЛЬТЕТ ПРИКЛАДНИХ НАУК

r/ai center for responsible ai

SIMONS FOUNDATION

NYU Global Programs

RAI for Ukraine

Responsible
AI Research for
Ukrainian
Scholars

Launched in June 2022

r/ai

bitly

http://r-ai.co/ukraine

Data-centric AI & responsible data management

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

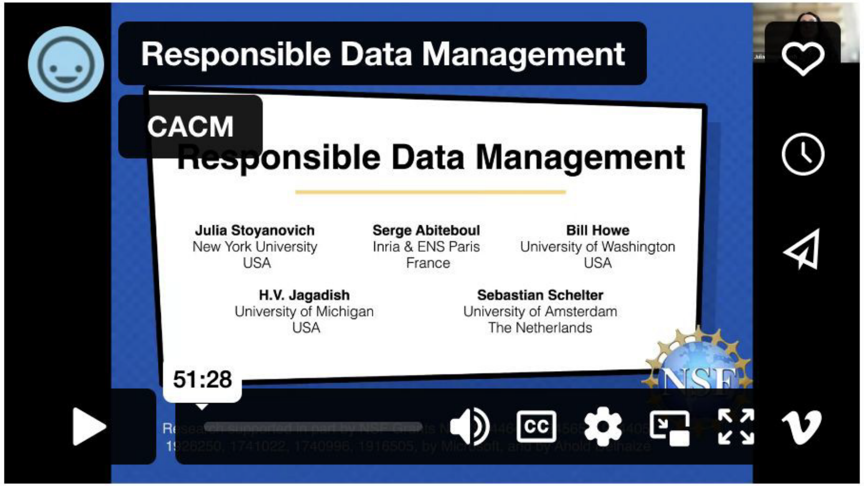
Responsible Data Management

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

64 COMMUNICATIONS OF THE ACM | JUNE 2022 | VOL. 65 | NO. 6



[Stoyanovich, Abiteboul, Howe, Jagadish, Schelter; *Comm. ACM 2022*]

center for responsible ai

Data-centric AI & responsible data management

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[a] and interpretable[b] 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

64 COMMUNICATIONS OF THE ACM | JUNE 2022 | VOL. 65 | NO. 6

arXiv:2504.11259v1 [cs.DB] 15 Apr 2025

The Cambridge Report on Database Research

Anastasia Ailamaki, Samuel Madden, Daniel Abadi, Gustavo Alonso, Sihem Amer-Yahia, Magdalena Balazinska, Philip A. Bernstein, Peter Boncz, Michael Cafarella, Surajit Chaudhuri, Susan Davidson, David DeWitt, Yanlei Diao, Xin Luna Dong, Michael Franklin, Juliana Freire, Johannes Gehrke, Alon Halevy, Joseph M. Hellerstein, Mark D. Hill, Stratos Idreos, Yannis Ioannidis, Christoph Koch, Donald Kossmann, Tim Kraska, Arun Kumar, Guoliang Li, Volker Markl, Renée Miller, C. Mohan, Thomas Neumann, Beng Chin Ooi, Fatma Ozcan, Aditya Parameswaran, Ippokratis Pandis, Jignesh M. Patel, Andrew Pavlo, Danica Porobic, Viktor Sanca, Michael Stonebraker, Julia Stoyanovich, Dan Suciu, Wang-Chiew Tan, Shiv Venkataraman, Matei Zaharia, and Stanley B. Zdonik

1 Introduction

On October 19–20, 2023, the authors of this report convened in Cambridge, MA, to discuss the state of the database research field[1], its recent accomplishments and ongoing challenges, and future directions for research and community engagement. This gathering continues a long-standing tradition in the database community, dating back to the late 1980s, in which researchers meet roughly every five years to produce a forward-looking report [1–9].

This report summarizes the key takeaways from our discussions. We begin with a retrospective on the community's academic, open-source, and commercial successes over the past five years. We then turn to future opportunities, with a focus on core data systems—particularly in the context of cloud computing and emerging hardware—as well as on the growing impact of data science, data governance, and generative AI.

This document is not intended as an exhaustive survey of all technical challenges or industry innovations in the field. Rather, it reflects the perspectives of senior community members on the most pressing challenges and promising opportunities ahead.

2 Evolution Over The Past Five Years

The past five years have continued to see important advances in the database and data systems landscape, particularly around new hardware, cloud-based data systems, and the continued adoption of statistical techniques, ML, and AI in both core data systems architecture and components.

The rise of Large Language Models (LLMs) has significantly shaped the collective consciousness of both computer science and society in recent years. While LLM-related technologies are still evolving and have yet to reach their full potential, they offer a promising solution to many complex data challenges, particularly those involving natural language and unstructured data. Already, LLMs have unlocked new possibilities for understanding human intentions and needs, paving the way for more intuitive, natural language-based querying and analysis interfaces. They have also demonstrated the capacity to comprehend data, including unstructured formats such as video and text, and to ground structured data in broader general knowledge. Additionally, LLMs are capable of synthesizing complex, multi-step data transformation programs. If fully realized, these technologies promise to revolutionize the ability of data systems to understand users, data, and programs. This has already prompted researchers to reconsider traditional database interactions, broadening the scope to incorporate unstructured data and natural language into conventional database systems. We explore these LLM-related opportunities in greater detail in Section 3.2 below.

2.1 Research Successes

In this section, we briefly review some of the key areas of progress in the community in the past few years.

2.1.1 Core Data Systems. In reaction to the low-level MapReduce-style tools of the previous decade of Big Data, database research and products make great strides toward usability and rich functionality of databases at massive scales. In particular, cloud-native architectures have matured significantly, and the industry has widely adopted the concept of disaggregated storage and compute, enabling a high degree of scalability and flexibility.

The hardware landscape continues to evolve rapidly to cater to resource-hungry AI, opening up new challenges and opportunities for data systems. The database community has made strides in leveraging improved hardware capabilities, such as NVMe SSDs, GPUs, DPUs, and specialized AI accelerators such as FPGAs and ASICs. For example, research on NVMe SSDs has led to the development of new storage engines that can fully utilize their high IOPS and low latency, often redesigning traditional data structures such as B-trees to minimize random accesses. Work on persistent memory has resulted in novel index structures that provide crash consistency without the overhead of traditional

[1] Broadly defined as the community of researchers publishing in ACM SIGMOD, VLDB, and related conferences, journals, and workshops.

1

[Stoyanovich, Abiteboul, Howe, Jagadish, Schelter; *Comm. ACM 2022*]

center for responsible ai

Grounding: Bias in computer systems

Pre-existing: independent of the technical system, has origins in society

Technical: introduced or exacerbated by the properties of the technical system

Emergent: arises due to the context of use

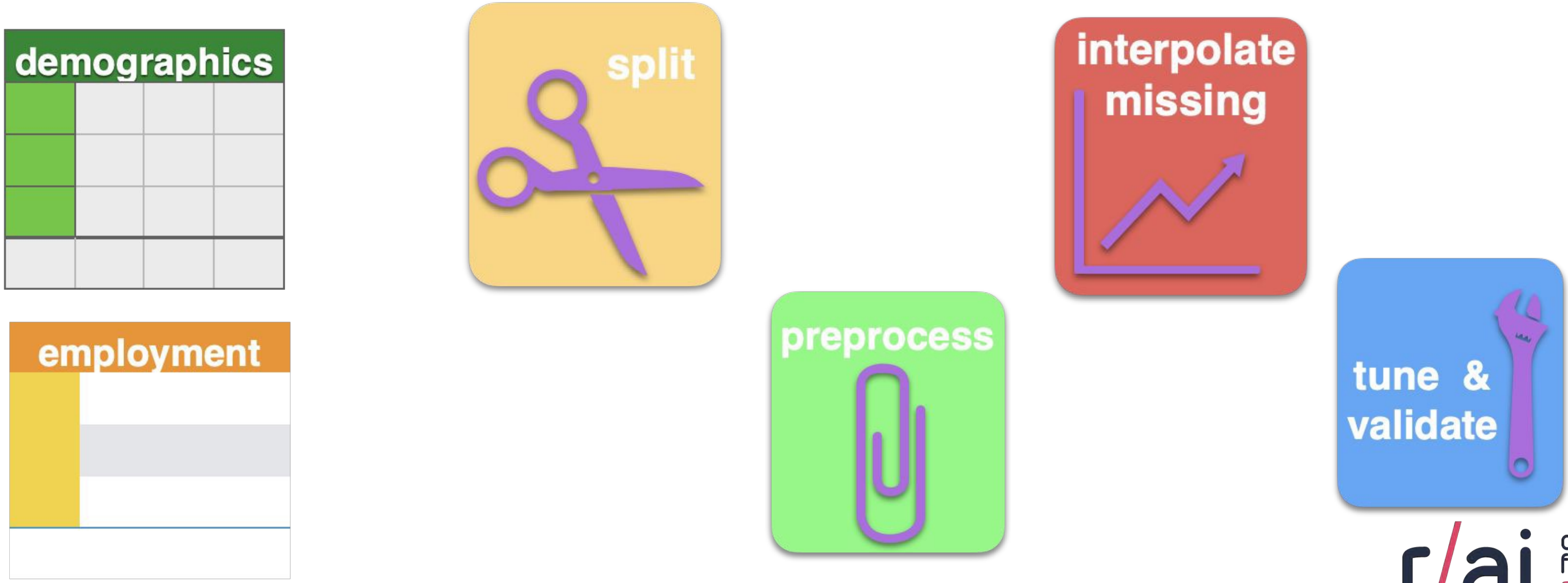
[Friedman & Nissenbaum, 1996]

r/ai center for responsible ai

Example: Taming technical bias

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

Problem: accuracy is lower for applicants who have more experience on the job - a **fairness concern**



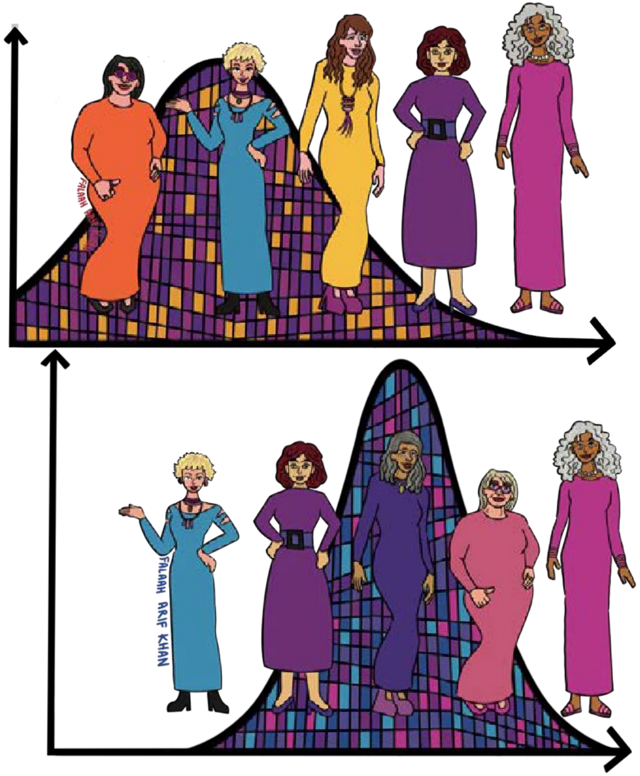
r/ai center for responsible ai

Missing values 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*)?



r/ai center for responsible ai

Data filtering

operations like **selection** and **join**, can arbitrarily change demographic group proportions

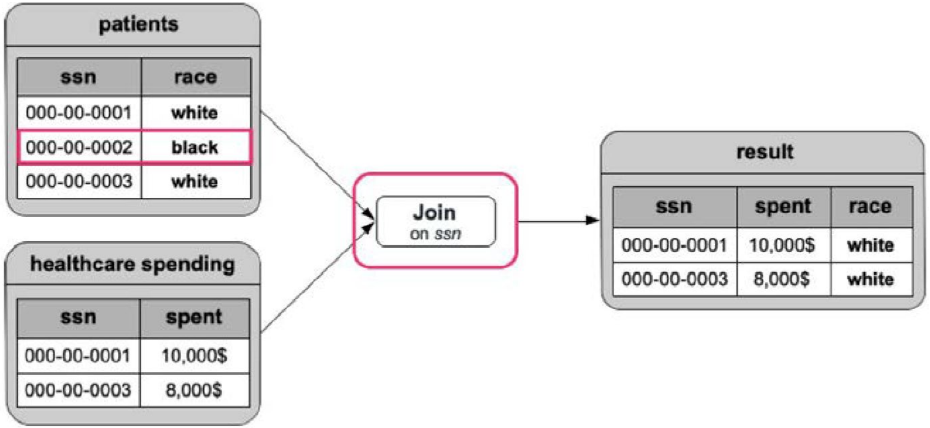
age_group	county
60	CountyA
60	CountyA
20	CountyA
60	CountyB
20	CountyB
20	CountyB

50% vs 50%

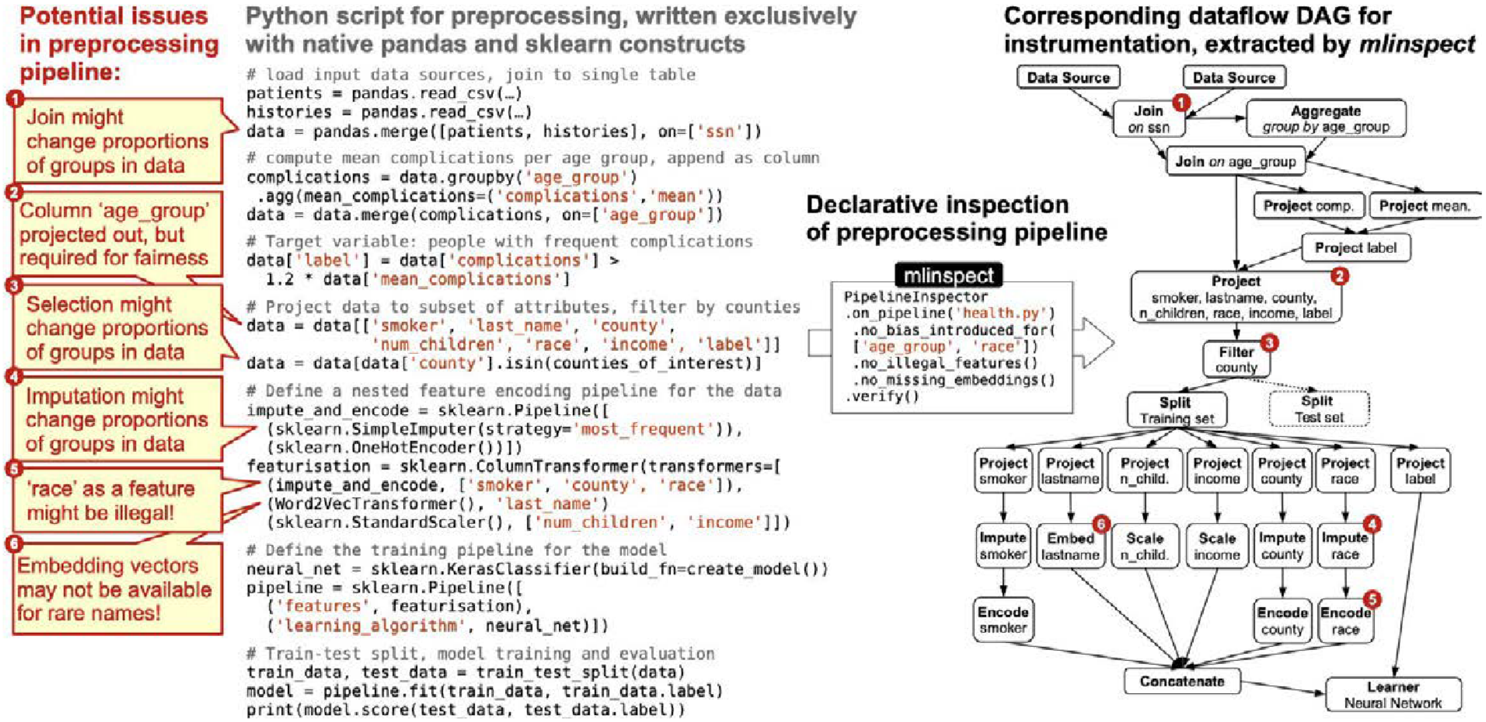
→

age_group	county
60	CountyA
60	CountyA
20	CountyA

66% vs 33%



Data distribution debugging

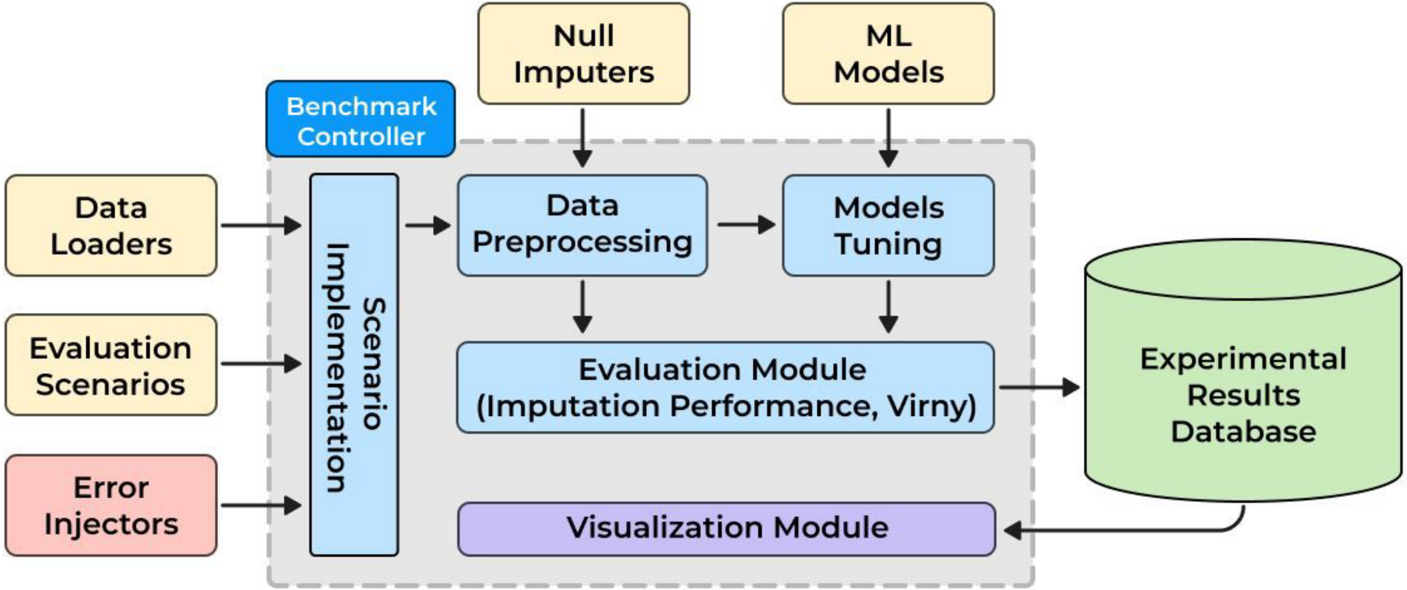


[Grafberger, Stoyanovich & Schelter; *CIDR 2021 & SIGMOD 2021*]

[Grafberger, Groth, Stoyanovich & Schelter; *VLDBJ 2022*]

center for responsible ai

Shades-of-NULL: A missing value imputation benchmark



[Arif Khan, Herasymuk, Protsiv, Stoyanovich; *arXiv* 2025]

r/ai center for responsible ai

Shades-of-NULL: Evaluation scenarios

	Train			Test		
Scenario	MCAR	MAR	MNAR	MCAR	MAR	MNAR
S1	✓			✓		
S2		✓			✓	
S3			✓			✓
S4	✓				✓	
S5	✓					✓
S6		✓		✓		
S7		✓				✓
S8			✓	✓		
S9			✓		✓	
S10	✓	✓	✓	✓	✓	✓

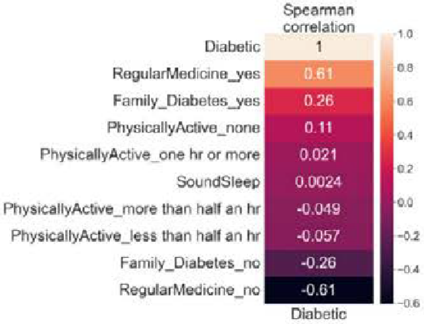
- **Missingness:** MCAR, MAR, MNAR
- Missingness as a form of **bias**
- **Mixed** missingness
- **Missingness** shift
- Socially-salient evaluation scenarios

[Arif Khan, Herasymuk, Protsiv, Stoyanovich; *arXiv* 2025]

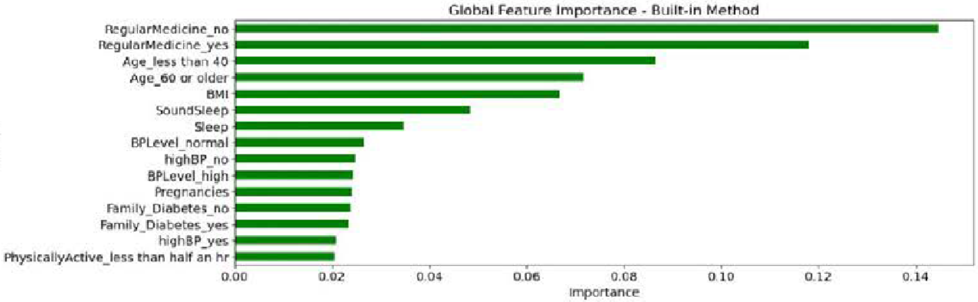
r/ai center for responsible ai

Shades-of-NULL: Missingness in context

Mechanism	Missing Column ($\mathcal{F}^m$)	Conditional Column (I)	Pr($\mathcal{F}^m$ | I is dis)	Pr($\mathcal{F}^m$ | I is priv)
MCAR	SoundSleep, Family_Diabetes, PhysicallyActive, RegularMedicine	N/A	0.3	0.3
MAR	Family_Diabetes, RegularMedicine	Sex	0.2 (female)	0.1 (male)
	PhysicallyActive, SoundSleep	Age	0.2 (≥ 40)	0.1 (< 40)
MNAR	Family_Diabetes	Family_Diabetes	0.25 (yes)	0.05 (no)
	RegularMedicine	RegularMedicine	0.2 (yes)	0.1 (no)
	PhysicallyActive	PhysicallyActive	0.25 (none, $< \frac{1}{2}$ hour)	0.05 ($> \frac{1}{2}$ hour, $>$ 1 hour)
	SoundSleep	SoundSleep	0.2 (< 5)	0.1 (≥ 5)



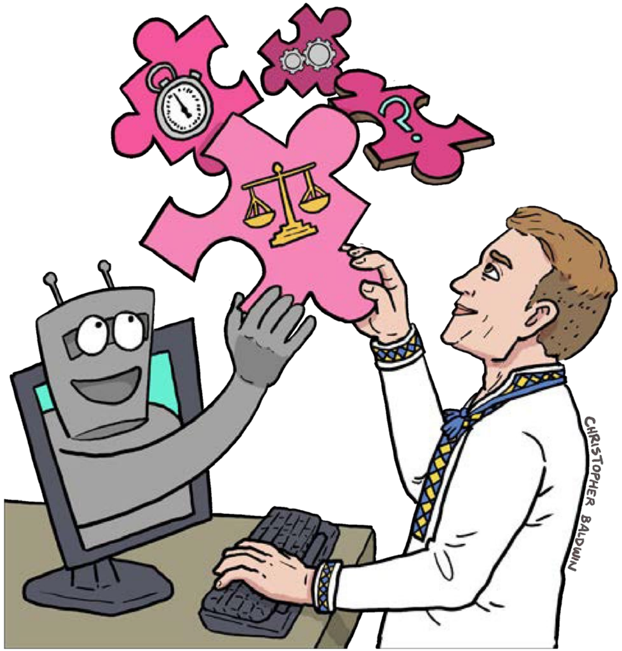
(a) Correlation with label



(b) Feature importance

[Arif Khan, Herasymuk, Protsiv, Stoyanovich; *arXiv* 2025]

Virny: Responsible model selection



Responsible Model Selection with Virny and VirnyView

Denys Herasymuk, Falaah Arif Khan, and Julia Stoyanovich

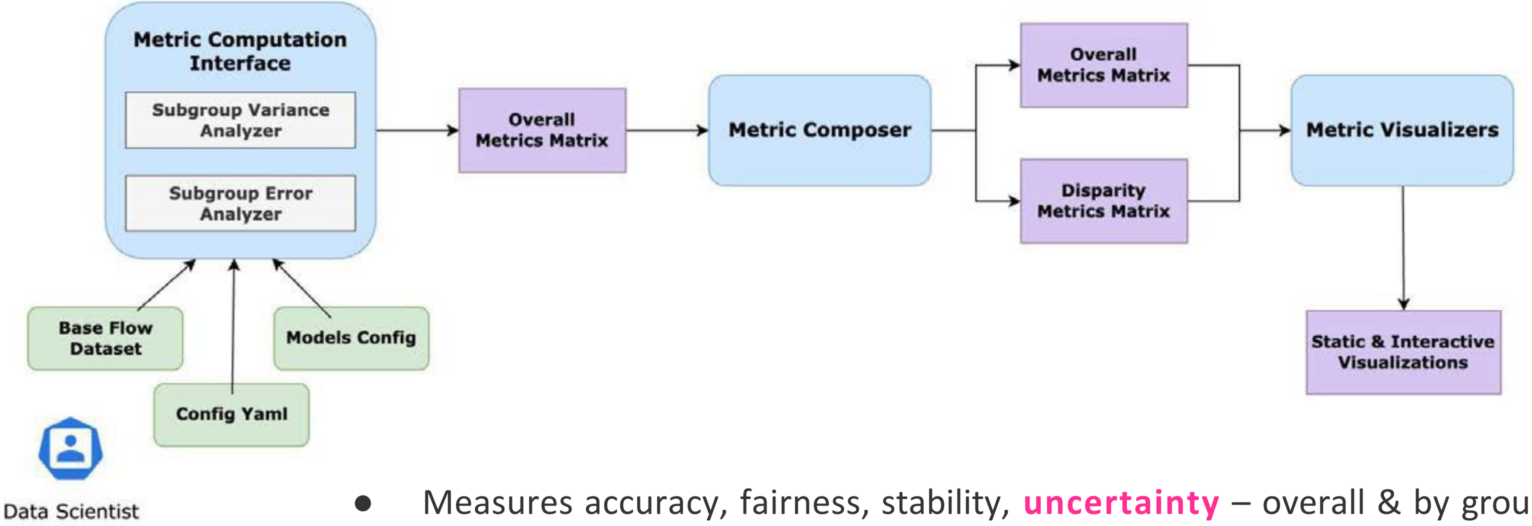
In Companion of the International Conference on Management of Data, SIGMOD/PODS, Santiago, Chile 2024

CITE PDF GITHUB

[Herasymuk, Arif Khan, Stoyanovich; *SIGMOD 2024*]

center for responsible ai

Virny: Responsible model selection

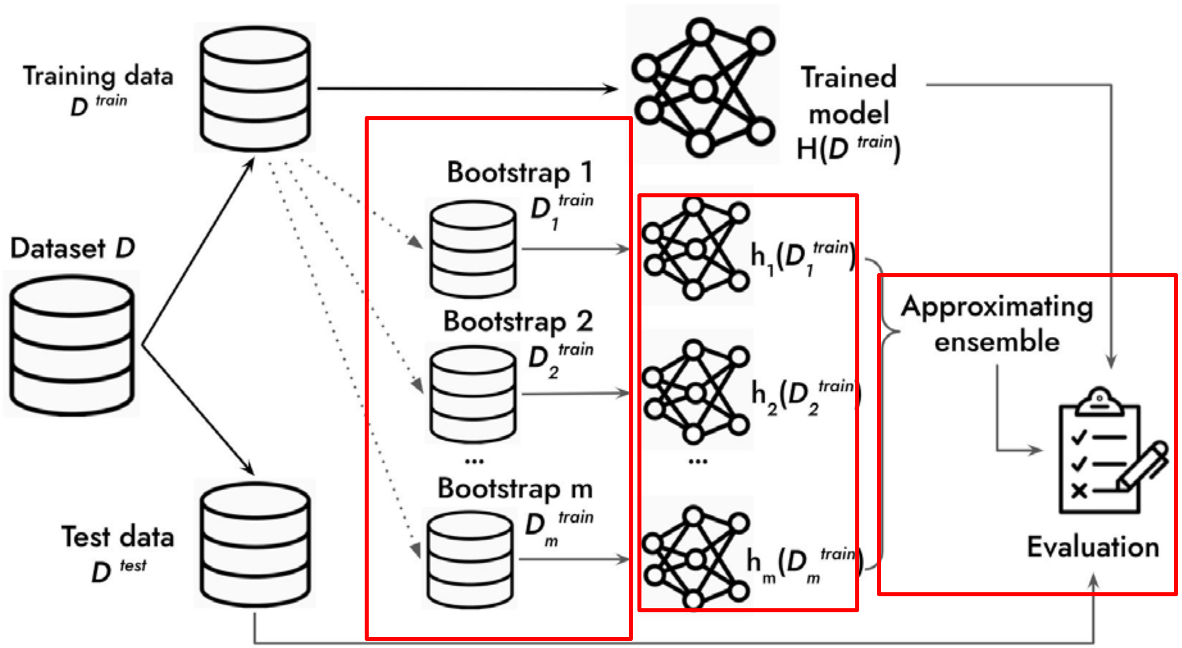


- Measures accuracy, fairness, stability, **uncertainty** – overall & by group
- Supports **multiple sensitive attributes** and their intersections
- Offers **diverse metric computation interfaces** for in-depth profiling of model performance

[Herasymuk, Arif Khan, Stoyanovich; *SIGMOD 2024*]

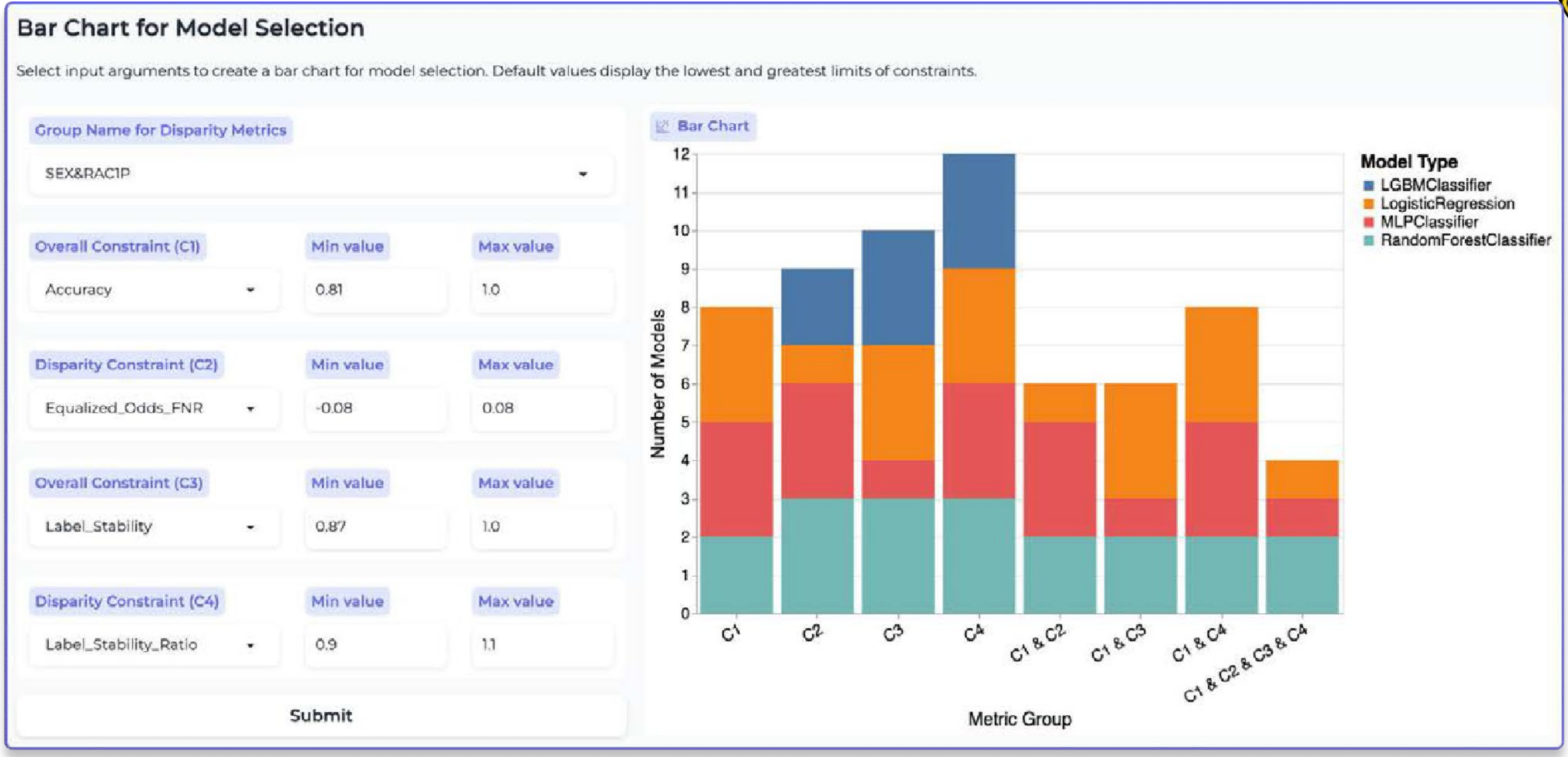
r/ai center for responsible ai

Virny: Measuring arbitrariness & uncertainty



[Herasymuk, Arif Khan, Stoyanovich; *SIGMOD 2024*]

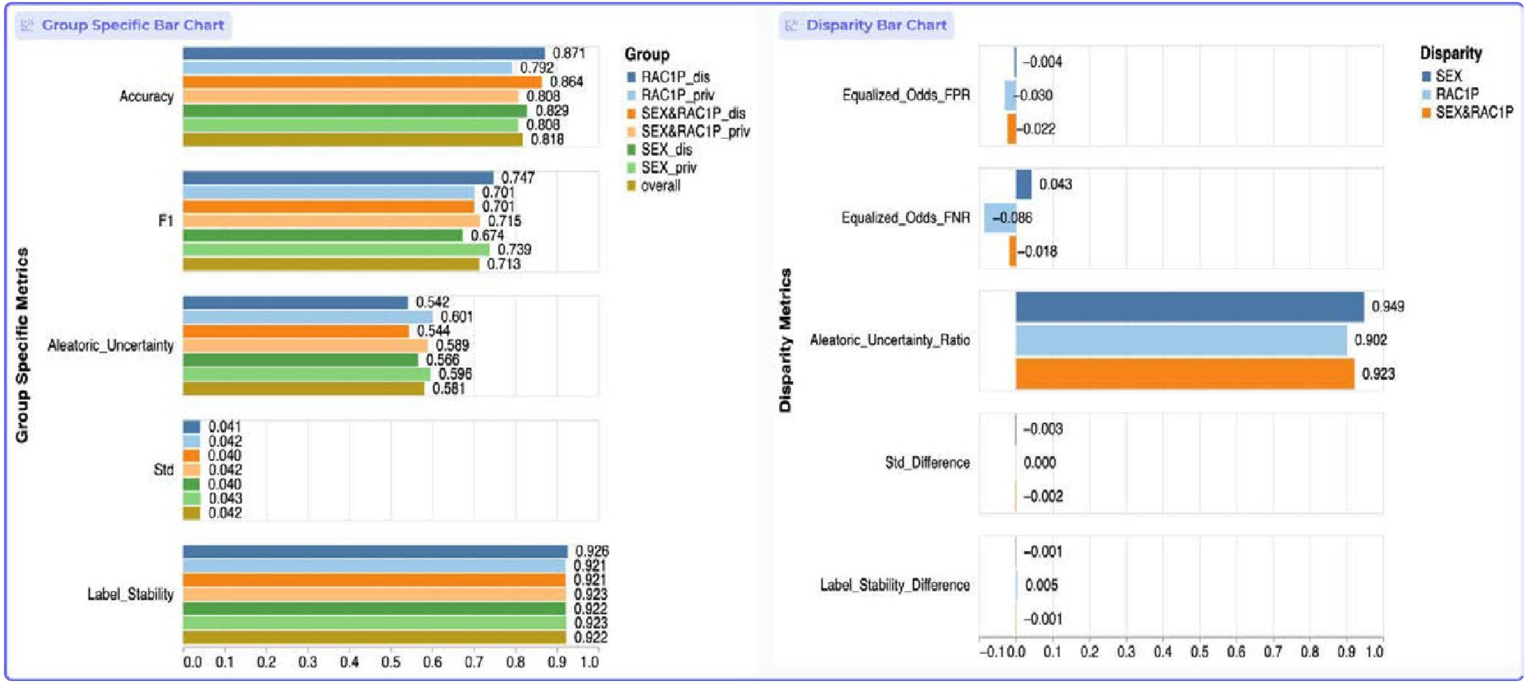
Virny: Model selection



[Herasymuk, Arif Khan, Stoyanovich; *SIGMOD 2024*]

r/ai center for responsible ai

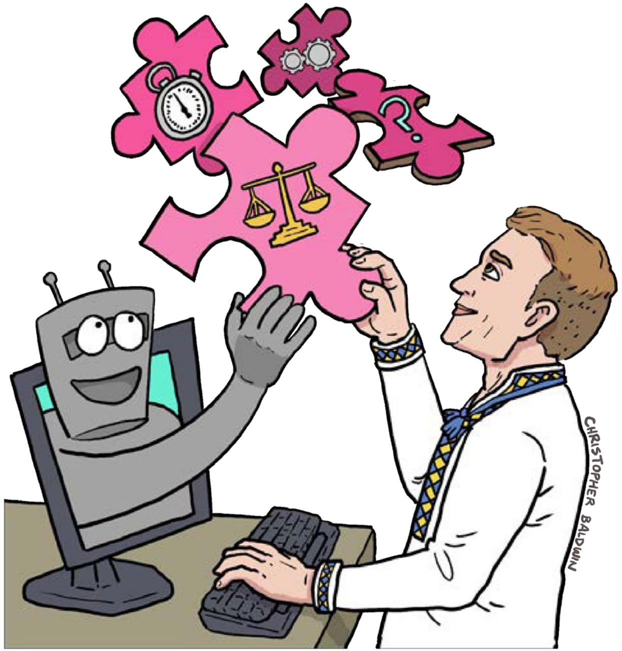
Virny: Model nutritional label



[Herasymuk, Arif Khan, Stoyanovich, *SIGMOD 2024*]

r/ai center for responsible ai

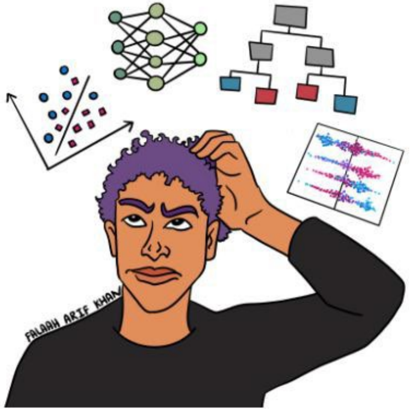
Virny: ongoing work



- **Speed up** responsible model selection - **VirnyFlow**
- Integrate multiple of model performance objectives into **hyperparameter optimization** - **VirnyFlow**
- **Assess** the impact of **ML lifecycle stages** on different aspects of model performance - **Shades-of-NULL**

r/ai center for responsible ai

Transparency & explainability



It's Just Not That Simple: An Empirical Study of the Accuracy-Explainability Trade-off in Machine Learning for Public Policy
Andrew Bell, Ian Solano-Kamaiko, Oded Nov, and Julia Stoyanovich
In Proceedings of the 5th Annual ACM Conference on Fairness, Accountability, and Transparency, FAccT 2022
CITE PDF

Introducing contextual transparency for automated decision systems
Mona Sloane, Ian Solano-Kamaiko, Jun Yuan, Aritra Dasgupta, and Julia Stoyanovich
Nature Machine Intelligence 2023
CITE PDF

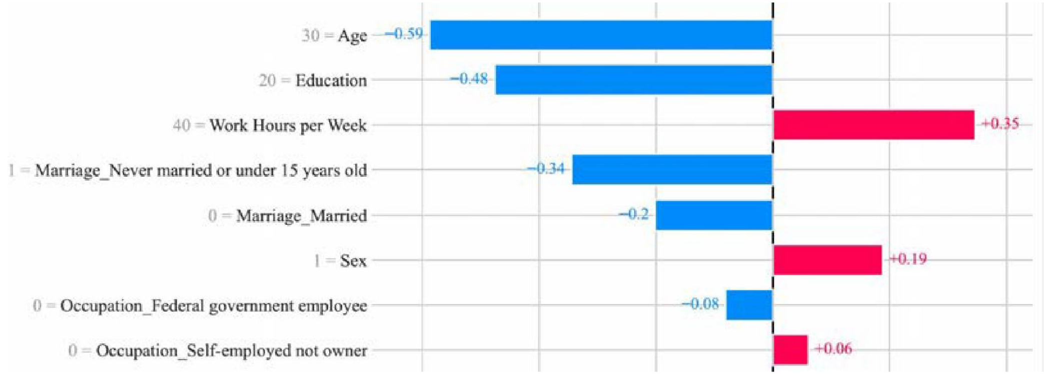
Think About the Stakeholders First! Towards an Algorithmic Transparency Playbook for Regulatory Compliance
Andrew Bell, Oded Nov, and Julia Stoyanovich
Data & Policy 2023
CITE PDF

The Imperative of Interpretable Machines
Julia Stoyanovich, Jay J. Van Bavel, and Tessa V. West
Nature Machine Intelligence 2020
CITE

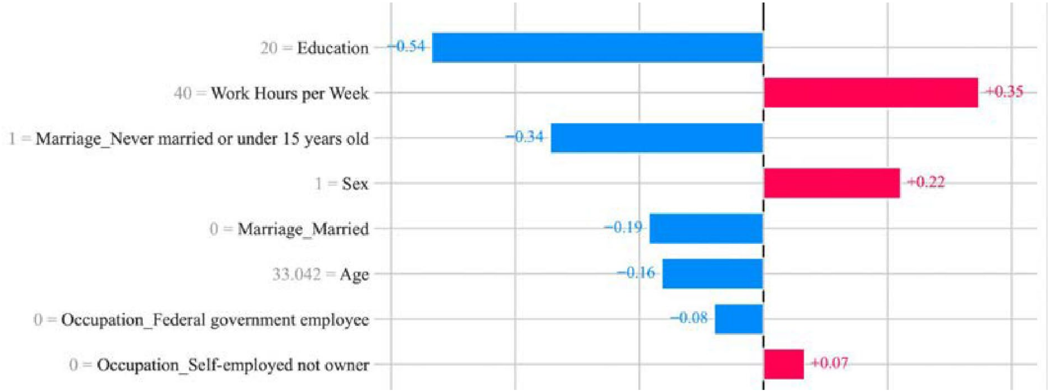
Nutritional Labels for Data and Models
Julia Stoyanovich, and Bill Howe
IEEE Data Eng. Bull. 2019
CITE PDF

r/ai center for responsible ai

SHAP attack!



continuous age value: most important feature (-0.59)



bucketized age (12 equi-width intervals) value: low importance (-0.16)

[Hwang, Bell, Fonseca, Pliatsika, Stoyanovich, Whang; *FAccT 2025*]

r/ai center for responsible ai

ShaRP: Explaining ranked outcomes

name	gpa	sat	essay	f	g
Bob	4	5	5	4.6	5
Cal	4	5	5	4.6	5
Dia	5	4	4	4.4	4
Eli	4	5	3	4.2	3
Fay	5	4	3	4.2	3
Kat	5	4	2	4.0	2
Leo	4	4	3	3.8	3
Osi	3	3	3	3.0	3

(a)

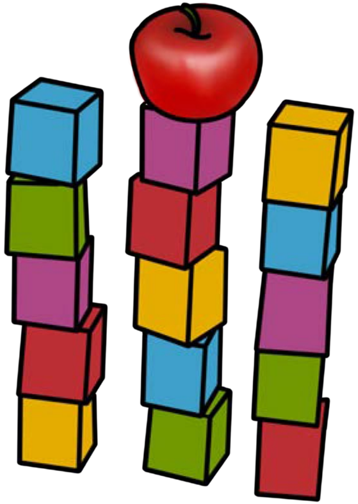
$r_{\mathcal{D},f}$
Bob
Cal
Dia
Eli
Fay
Kat
Leo
Osi

(b)

$r_{\mathcal{D},g}$
Bob
Cal
Dia
Eli
Fay
Leo
Osi
Kat

(c)

Figure 1: (a) Dataset $\mathcal{D}$ of college applicants, scored on *gpa*, *sat*, and *essay*. (b) Ranking $r_{\mathcal{D},f}$ of $\mathcal{D}$ on $f = 0.4 \times gpa + 0.4 \times sat + 0.2 \times essay$; the highlighted top-4 candidates will be interviewed and potentially admitted. (c) Ranking $r_{\mathcal{D},g}$ on $g = 1.0 \times essay$; the top-4 coincides with that of $r_{\mathcal{D},f}$, signifying that *essay* has the highest importance for f, despite carrying the lowest weight in the scoring function.



Example: CSRankings

CSRankings: Computer Science Rankings

CSRankings is a metrics-based ranking of top computer science institutions around the world. **Click on a triangle** (▶) to expand areas or institutions. **Click on a name** to go to a faculty member's home page. **Click on a chart icon** (the 📊 after a name or institution) to see the distribution of their publication areas as a bar chart. **Click on a Google Scholar icon** to see publications, and **click on the DBLP logo** to go to a DBLP entry. *Applying to grad school? Read this first.* For info on grad stipends, check out CSStipendRankings.org. **Do you find CSrankings useful? Sponsor CSrankings on GitHub.**

Rank institutions in USA by publications from 2014 to 2024

$$f = \sqrt[27]{(AC_{AI}^{5}+1)(AC_{Sys}^{12}+1)(AC_{Th}^{3}+1)(AC_{Int}^{7}+1)}$$

All Areas [off | on]

AI [off | on]

- ▶ Artificial intelligence ☑
- ▶ Computer vision ☑
- ▶ Machine learning ☑
- ▶ Natural language processing ☑
- ▶ The Web & information retrieval ☑

Systems [off | on]

- ▶ Computer architecture ☑
- ▶ Computer networks ☑
- ▶ Computer security ☑
- ▶ Databases ☑
- ▶ Design automation ☑
- ▶ Embedded & real-time systems ☑
- ▶ High-performance computing ☑
- ▶ Mobile computing ☑
- ▶ Measurement & perf. analysis ☑
- ▶ Operating systems ☑
- ▶ Programming languages ☑
- ▶ Software engineering ☑

Theory [off | on]

- ▶ Algorithms & complexity ☑
- ▶ Cryptography ☑
- ▶ Logic & verification ☑

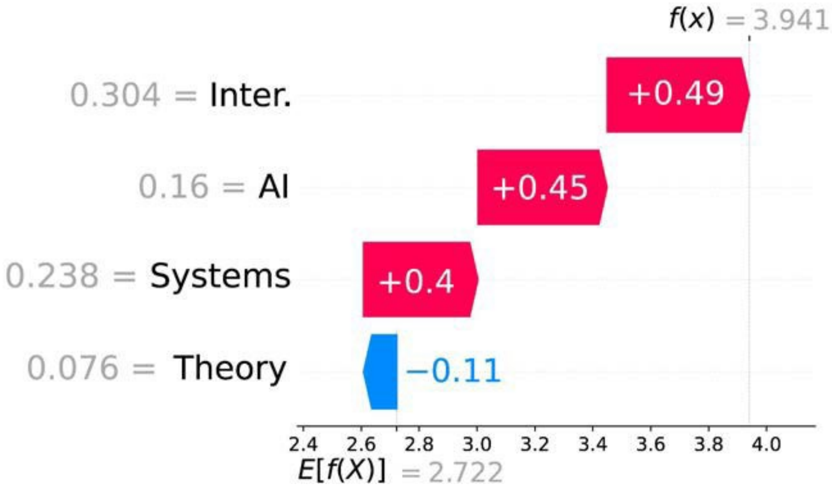
Interdisciplinary Areas [off | on]

- ▶ Comp. bio & bioinformatics ☑
- ▶ Computer graphics ☑
- ▶ Computer science education ☑
- ▶ Economics & computation ☑
- ▶ Human-computer interaction ☑
- ▶ Robotics ☑
- ▶ Visualization ☑

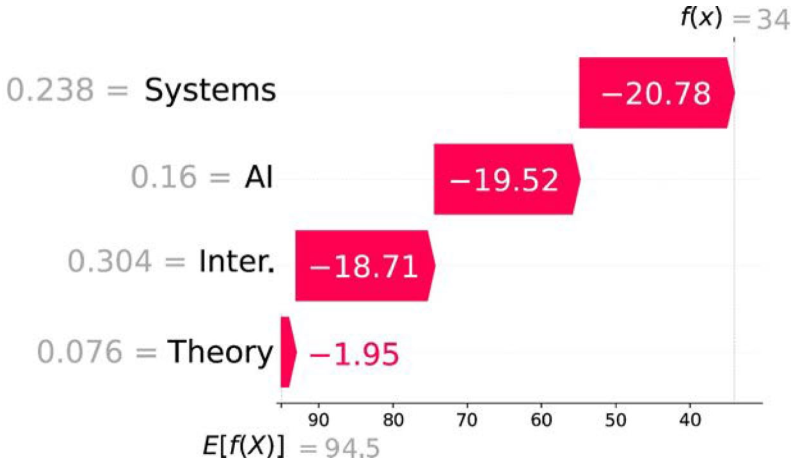
https://csrankings.org/

Feature contributions to score vs. rank

Texas A&M, **score** quantity of interest: *Interdisciplinary* is most important



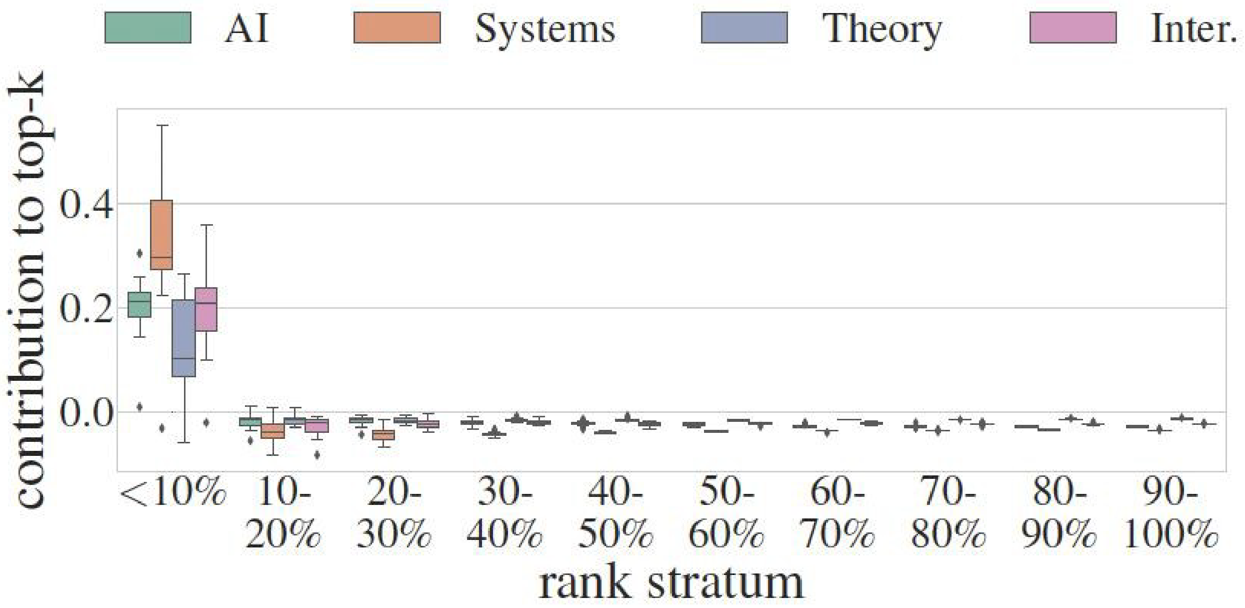
Texas A&M, **rank** quantity of interest: *Systems* is most important



[Pliatsika, Fonseca, Akhynko, Shevchenko, Stoyanovich; *arXiv 2024*]

r/ai center for responsible ai

Feature contributions to the top-20



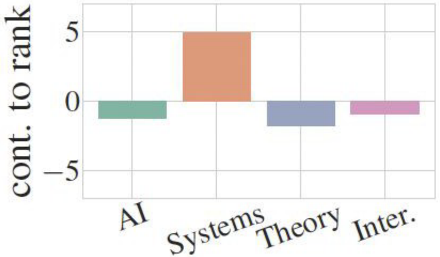
(a) Feature contribution to the top-k QoI, for k = 10%. Systems is the most important feature, followed by Interdisciplinary and AI.

[Pliatsika, Fonseca, Akhynko, Shevchenko, Stoyanovich; *arXiv 2024*]

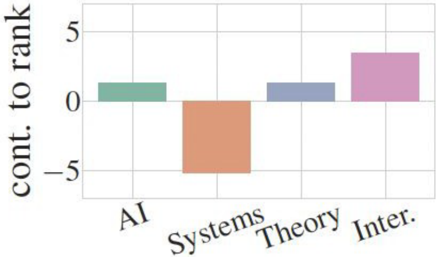
Feature contributions: Pairwise outcomes

Institution	AI	Systems	Theory	Inter.	Rank
Georgia Tech	28.5	7.8	6.9	10.2	5
Stanford	36.7	5.4	13.3	11.5	6
UMich	30.4	9.0	9.3	5.9	7

(b) Feature values and rank of three highly ranked departments: Georgia Tech, Stanford, and UMich.



(c) Pairwise method using the rank QoI: Georgia Tech ranks higher than Stanford because of its relative strength in Systems.

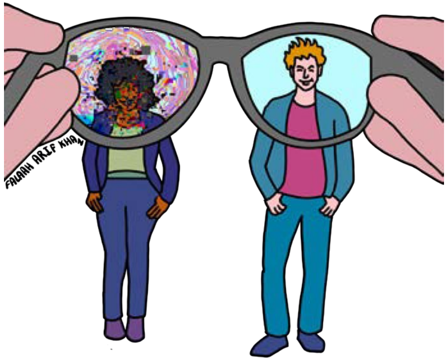


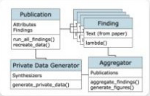
(d) Pairwise method using the rank QoI: Stanford ranks higher than UMich despite Stanford's relative weakness in Systems.

[Pliatsika, Fonseca, Akhynko, Shevchenko, Stoyanovich; *arXiv 2024*]

r/ai center for responsible ai

Privacy & data protection

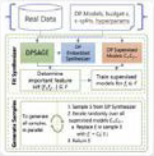




Epistemic Parity: Reproducibility as an Evaluation Metric for Differential Privacy
Lucas Rosenblatt, Bernease Herman, Anastasia Holovenko, Wonkwon Lee, Joshua R. Loftus, Elizabeth Mckinnie, Taras Rumezhak, Andrii Stadnik, Bill Howe, and Julia Stoyanovich
Proc. VLDB Endow. 2023
CITE PDF GITHUB

The Many Facets of Data Equity
H.V. Jagadish, Julia Stoyanovich, and Bill Howe
ACM Journal of Data and Information Quality 2023
CITE PDF

Personal Data for Personal Use: Vision or Reality?
Xin Luna Dong, Bo Li, Julia Stoyanovich, Anthony Kum Hoe Tung, Gerhard Weikum, Alon Y. Halevy, and Wang-Chiew Tan
In Companion of the 2023 International Conference on Management of Data, SIGMOD/PODS 2023, Seattle, WA, USA, June 18-23, 2023 2023
CITE PDF

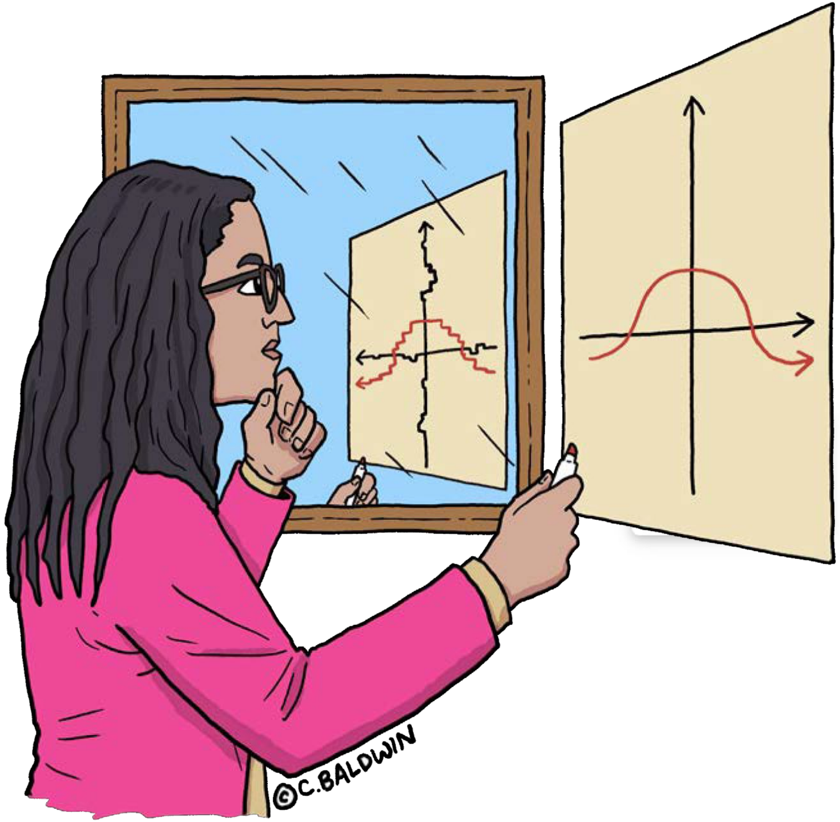


Spending Privacy Budget Fairly and Wisely
Lucas Rosenblatt, Joshua Allen, and Julia Stoyanovich
Theory and Practice of Differential Privacy (@ICML) 2022
CITE PDF

Transparency, Fairness, Data Protection, Neutrality: Data Management Challenges in the Face of New Regulation
Serge Abiteboul, and Julia Stoyanovich
ACM Journal of Data and Information Quality 2019
CITE PDF

r/ai center for responsible ai

Epistemic parity



Epistemic Parity: Reproducibility as an Evaluation Metric for Differential Privacy

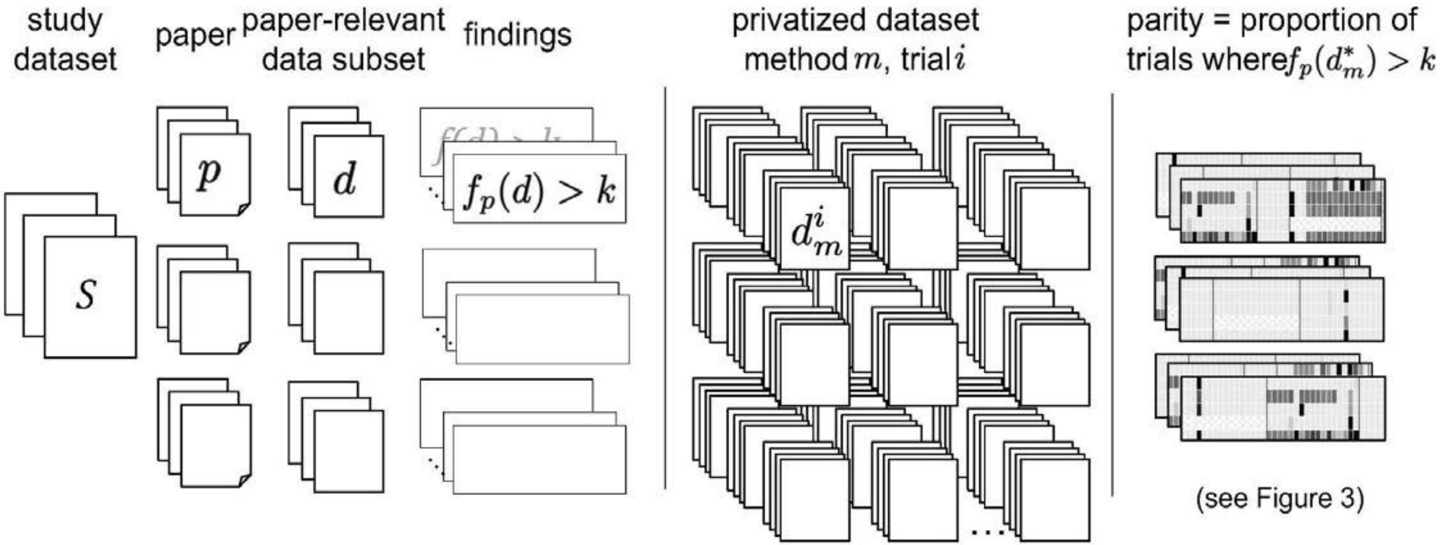
Lucas Rosenblatt, Bernease Herman, Anastasia Holovenko, Wonkwon Lee, Joshua R. Loftus, Elizabeth Mckinnie, Taras Rumezhak, Andrii Stadnik, Bill Howe, and Julia Stoyanovich

Proc. VLDB Endow. 2023

CITE PDF GITHUB

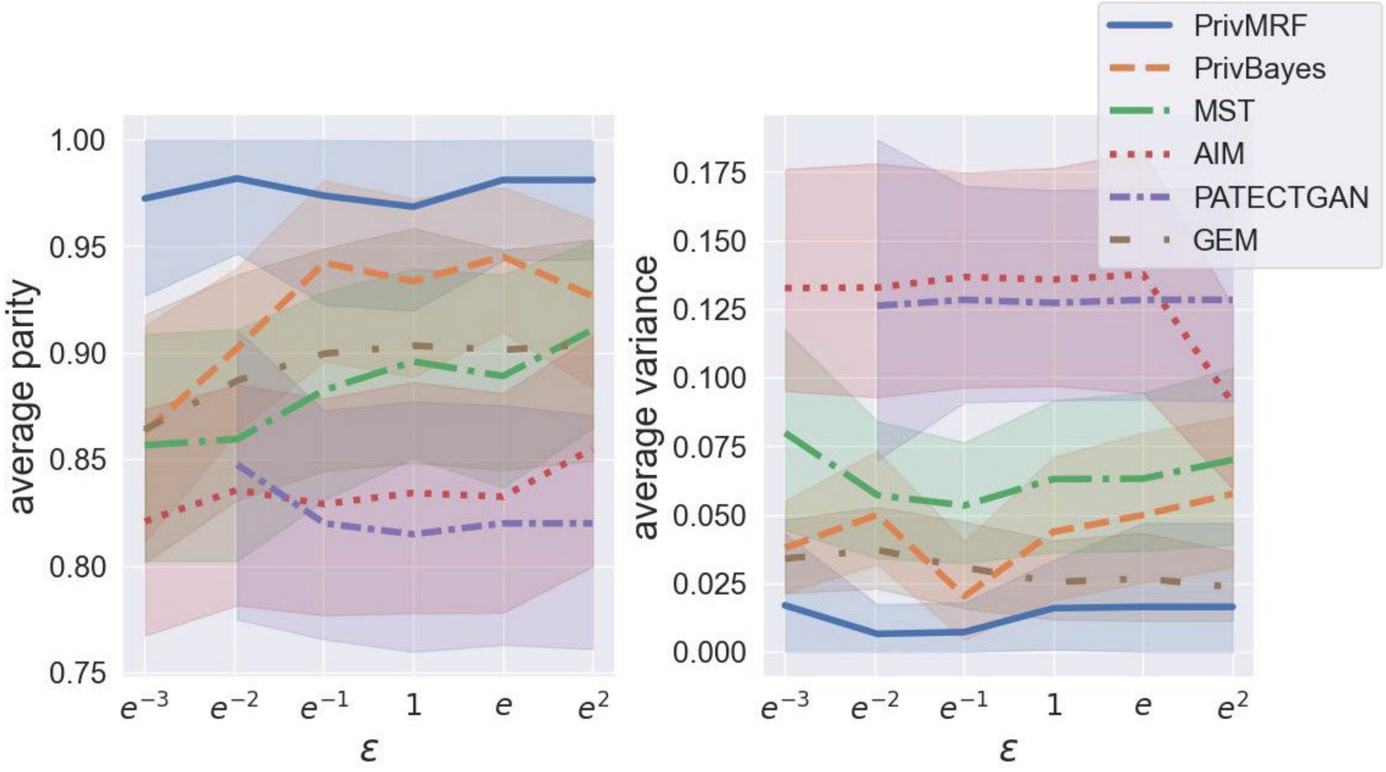
r/ai center for responsible ai

Epistemic parity: The workflow



[Rosenblatt et al., *VLDB 2023*]

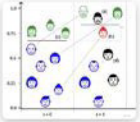
Epistemic parity: Results



[Rosenblatt et al., *VLDB 2023*]

center for responsible ai

Algorithmic fairness



Setting the Right Expectations: Algorithmic Recourse Over Time
João Fonseca, Andrew Bell, Carlo Abrate, Francesco Bonchi, and Julia Stoyanovich
In Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization, EAAMO 2023, Boston, MA, USA, 30 October 2023 - 1 November 2023 2023
CITE PDF

Fairness in Ranking: From Values to Technical Choices and Back
Julia Stoyanovich, Meike Zehlike, and Ke Yang
In Companion of the 2023 International Conference on Management of Data, SIGMOD/PODS 2023, Seattle, WA, USA, June 18-23, 2023 2023
CITE PDF

Counterfactuals for the Future
Lucius E. J. Bynum, Joshua R. Loftus, and Julia Stoyanovich
In Proceedings of the AAAI Conference on Artificial Intelligence 2023
CITE PDF

Towards Substantive Conceptions of Algorithmic Fairness: Normative Guidance from Equal Opportunity Doctrines
Falaah Arif Khan, Eleni Manis, and Julia Stoyanovich
In Equity and Access in Algorithms, Mechanisms, and Optimization, EAAMO 2022, Arlington, VA, USA, October 6-9, 2022 2022
CITE PDF

Query Refinement for Diversity Constraint Satisfaction
Jinyang Li, Yuval Moskovitch, Julia Stoyanovich, and H. V. Jagadish
Proc. VLDB Endow. 2023
CITE PDF

r/ai center for responsible ai

Fairness as equality of opportunity (EO): The principles

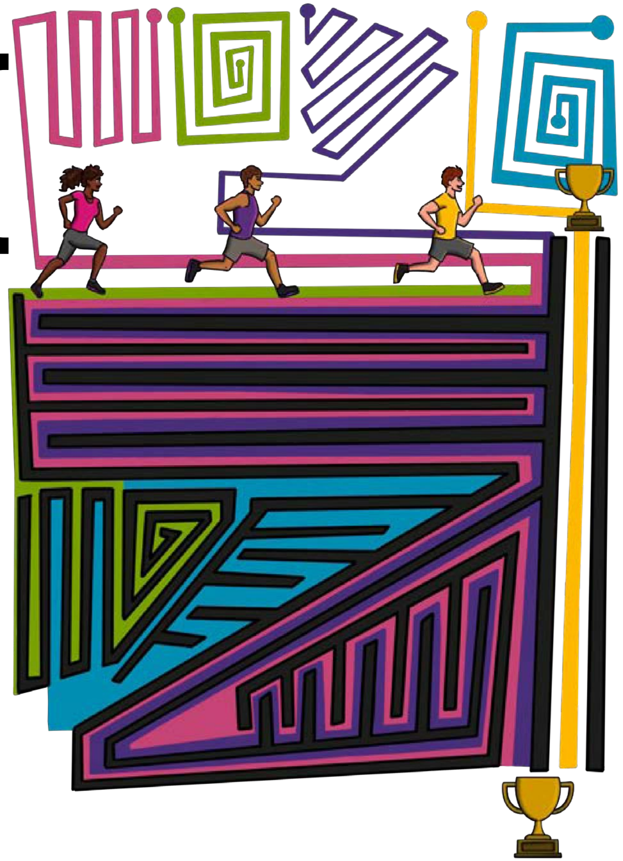
fair contests (non-discrimination)

fair life chances (leveling the playing field)

Fairness as equality of opportunity (EO): The domains

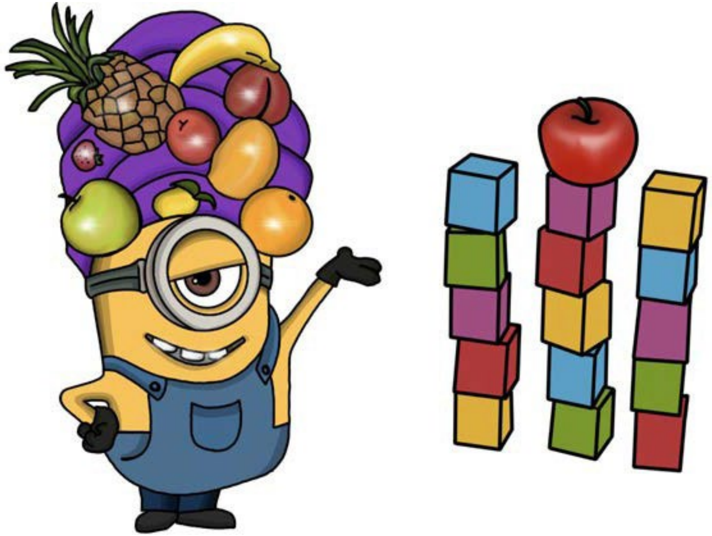
fairness in developmental opportunities

comparable opportunities over a lifetime



fairness at a decision point

Fairness in ranking



Fairness in Ranking: From Values to Technical Choices and Back

Julia Stoyanovich, Meike Zehlike, and Ke Yang

In Companion of the 2023 International Conference on Management of Data, SIGMOD/PODS 2023, Seattle, WA, USA, June 18-23, 2023 2023

CITE PDF

Fairness in Ranking, Part I: Score-Based Ranking

Meike Zehlike, Ke Yang, and Julia Stoyanovich

ACM Computing Surveys 2023

CITE PDF

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

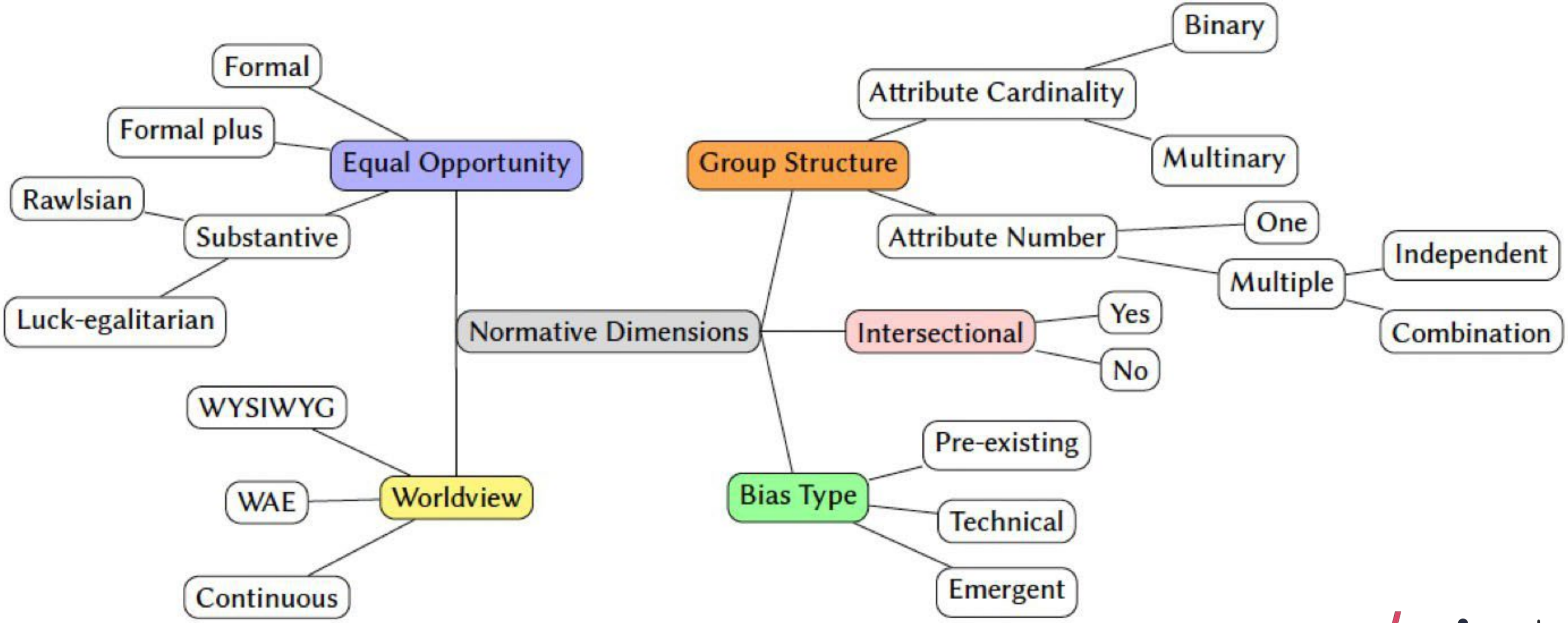
Meike Zehlike, Ke Yang, and Julia Stoyanovich

ACM Computing Surveys 2023

CITE PDF

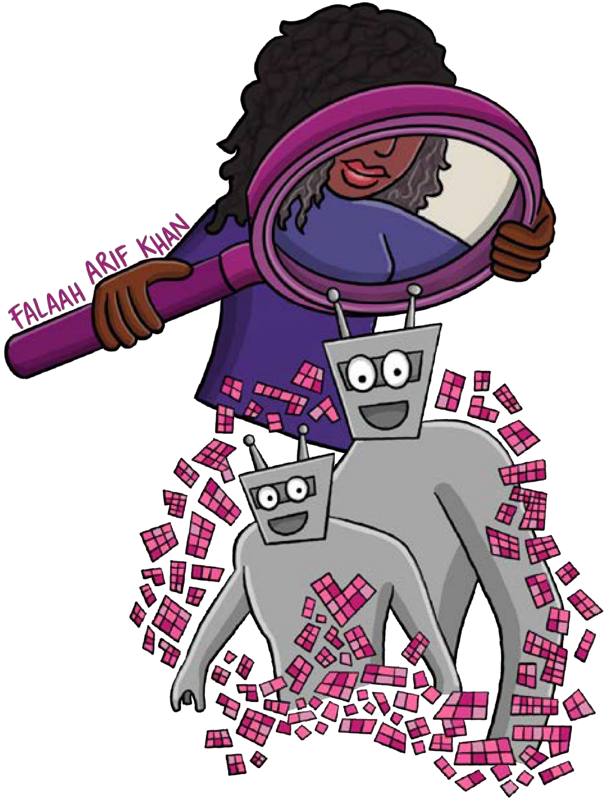
r/ai center for responsible ai

Classification of fair ranking methods



r/ai center for responsible ai

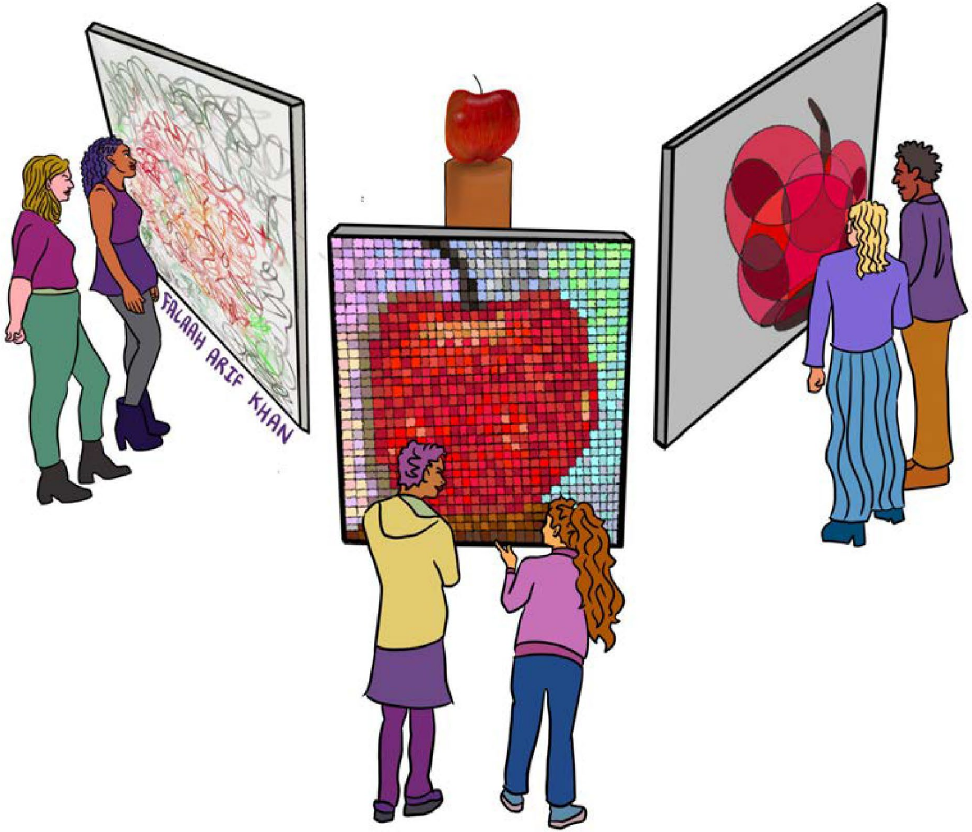
Responsible AI is about …



… exposing the knobs of responsibility to people

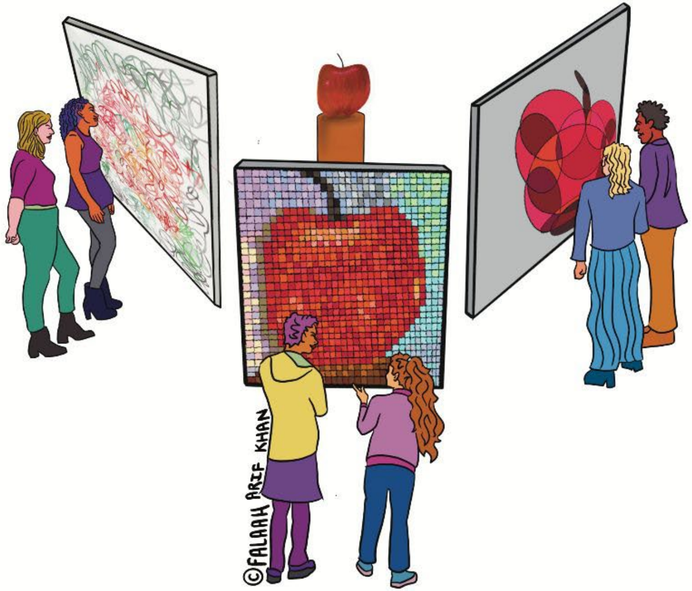
r/ai center for responsible ai

We need (responsible) AI education & training for everyone!



center
for
responsible
ai

Education & Training



r/ai center for responsible ai

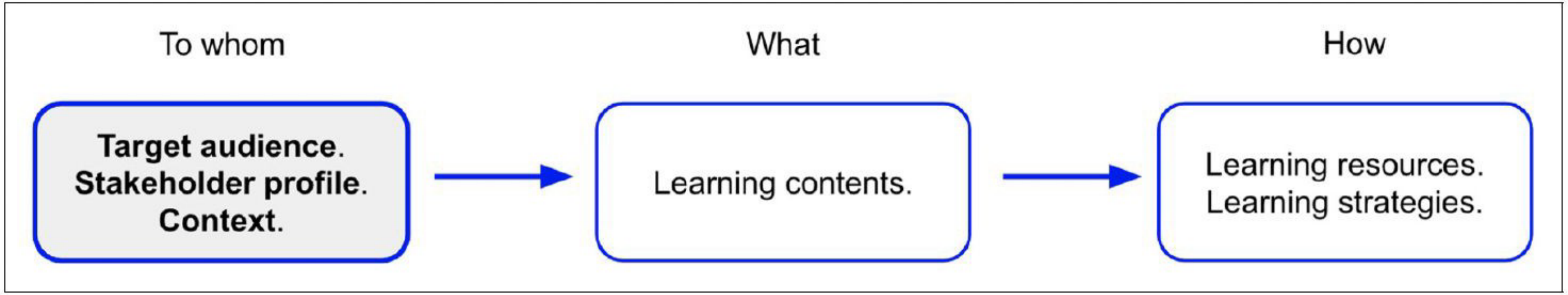
Original Research Article

Big Data & Society
July–December: 1–15

DOI: 10.1177/20539517231219958
journals.sagepub.com/home/bds
Sage

Responsible AI literacy: A stakeholder-first approach

Daniel Domínguez Figaredo[1] and Julia Stoyanovich[2]



r/ai center for responsible ai

Education & training: students (since 2019)

Audience: undergraduate and graduate students in data science, computer science

Prerequisites: introduction to data science or computer science (not machine learning!)

Challenge: reconcile technical training with interdisciplinary ethical concerns

Bonus: uses many of the technical tools I'll discuss today, all course materials are open

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.

http://r-ai.co/education

[Lewis and Stoyanovich; *IJAIED 2022*]

Practitioner training

NYU CENTER FOR RESPONSIBLE AI PRESENTS

WHAT IS RESPONSIBLE AI & HOW DOES IT APPLY TO YOUR WORK AT META?

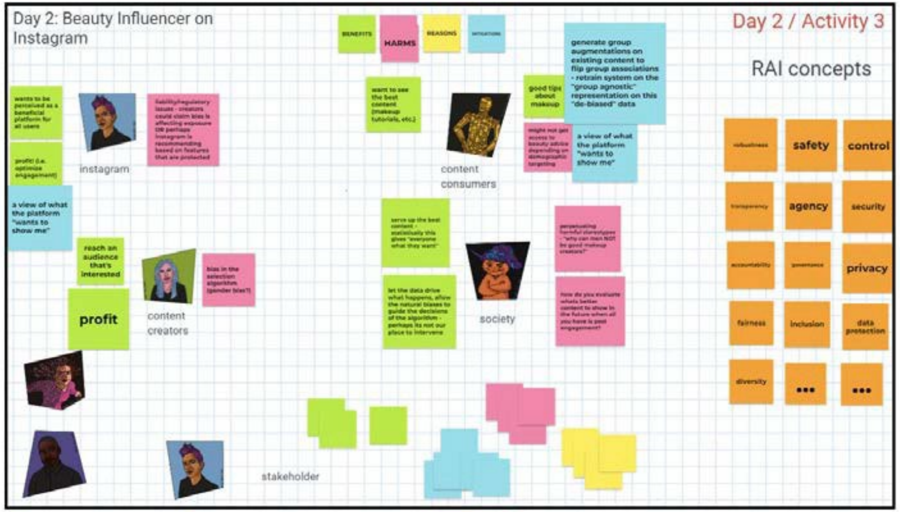
Responsible AI is the science and the practice of making the design, development, and use of AI socially sustainable. In this workshop, the **NYU Center for Responsible AI** will give an overview of the discipline, and will present and extensively discuss concrete Meta-relevant case studies on ad targeting, content ranking, machine translation, algorithmic hiring, and more. We will take deep dives into algorithmic fairness, transparency and interpretability, and privacy and data protection, while keeping the conversation relevant to your work at Meta. No prior knowledge of responsible AI concepts or techniques is required, or even expected, for participation. Whether you are a responsible AI aficionado or a skeptic, and whether or not your role at Meta is technical — this workshop is for you!

This workshop will consist of two 120-minute sessions. When signing up for the workshop, you commit to attending both sessions and participating in the discussion. Note that the NYU team is conducting an educational research study in conjunction with the workshop, to learn more about effective ways to teach responsible AI concepts and techniques to industry practitioners. They welcome your participation in the workshop even if you do not wish to participate in the research study. You will have an opportunity to find out more about the research and opt into the study at the start of the workshop.

The workshop will be presented by **Dr. Julia Stoyanovich** Institute Associate Professor of Computer Science and Engineering, Associate Professor of Data Science, and Director of the Center for Responsible AI at NYU. Please reach out to her at **stoyanovich@nyu.edu** if you have any questions about the workshop.

NYU Center for Responsible AI aims to make responsible AI synonymous with AI. We engage in basic and applied research, participate in technology policy & regulation efforts, and teach responsible AI to data science students, practitioners, and members of the public. For more information, visit **https://airesponsibly.com**.

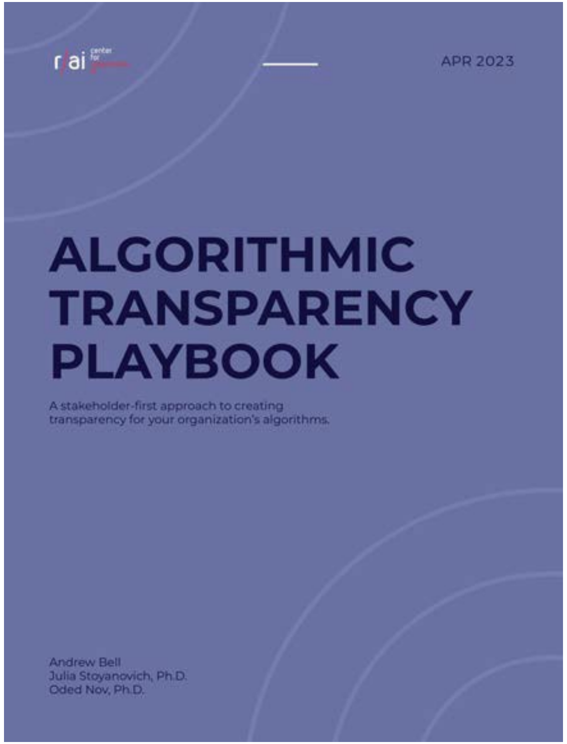
r/ai

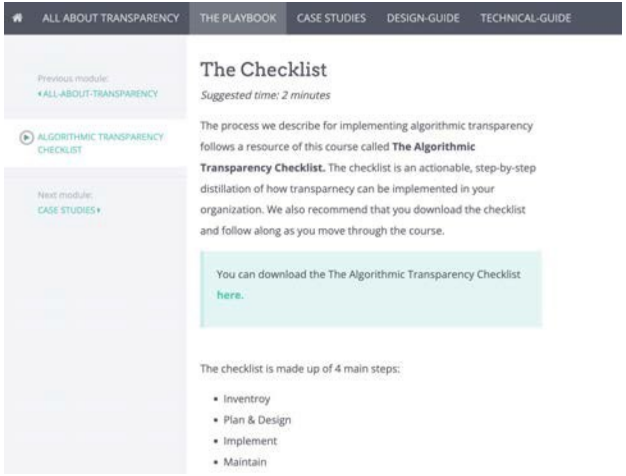


[Stoyanovich, de Paula, Lewis, Zheng; *EAAI 2025*]

r/ai center for responsible ai

Practitioner training

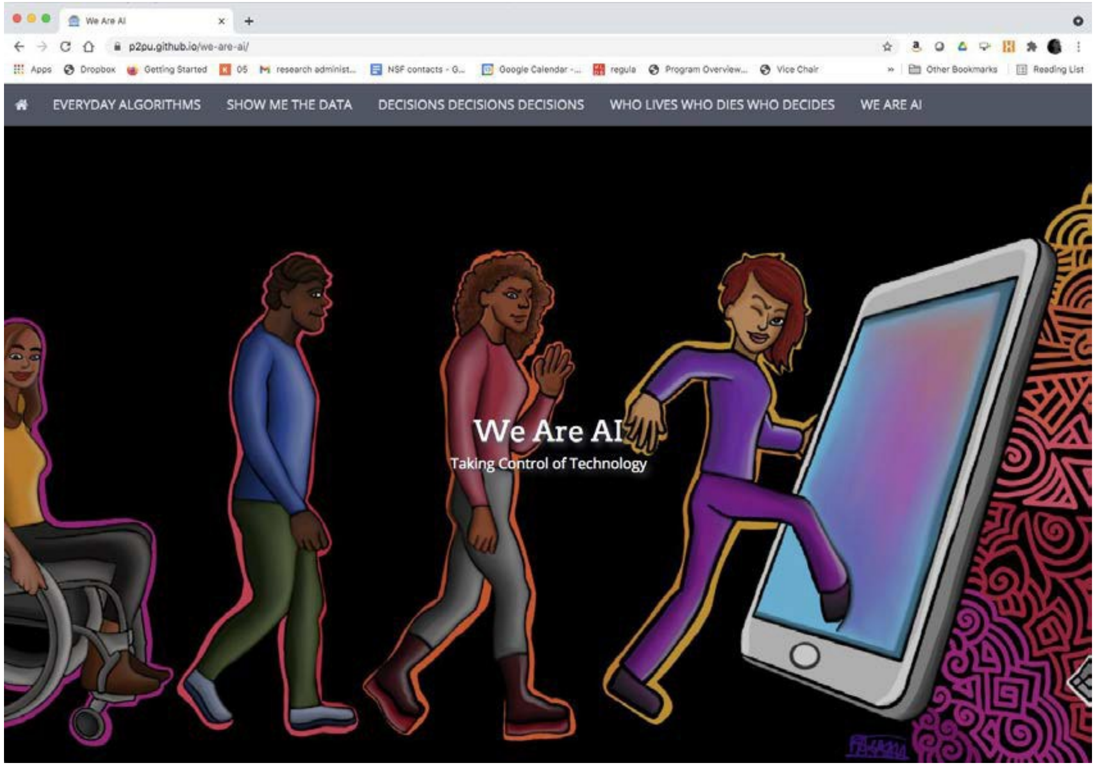




[Bell and Stoyanovich, *EAAI 2025*]

center for responsible ai

We are AI: Taking control of technology

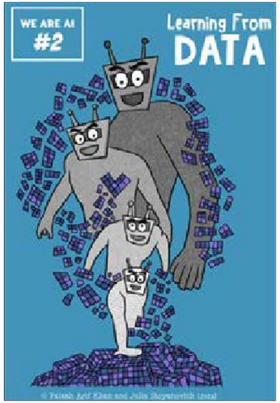


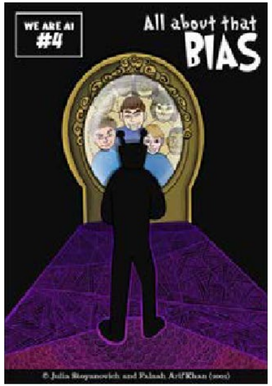
[Stoyanovich, Lewis, Corbett, et al., *EAAI 2025*]

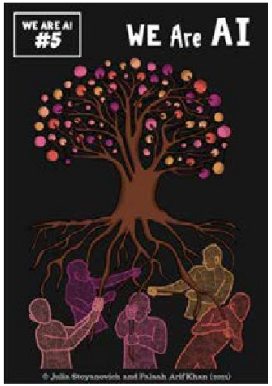
http://r-ai.co/We-are-AI

r/ai center for responsible ai

We are AI comics



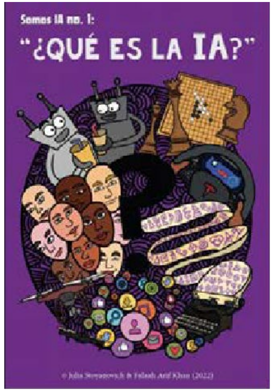


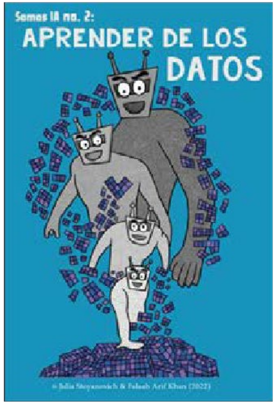


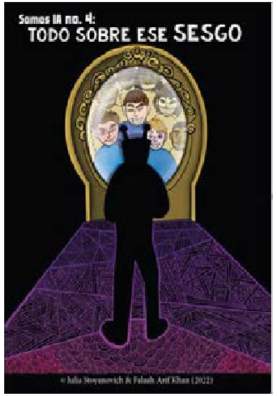
http://r-ai.co/comics

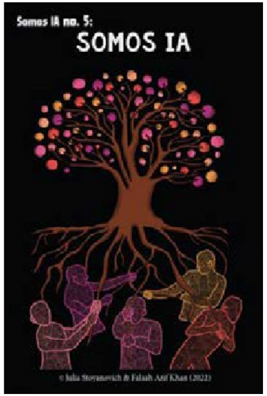
r/ai center for responsible ai

We are AI comics / Spanish





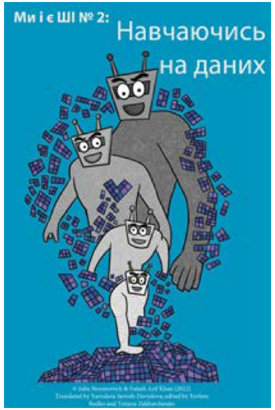


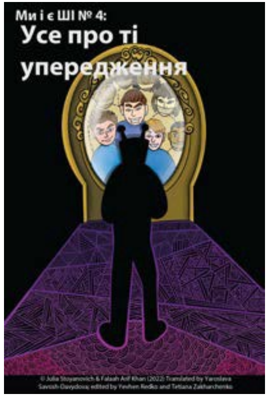


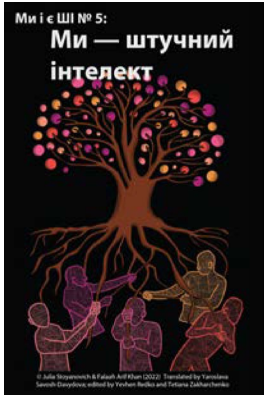
http://r-ai.co/comics

r/ai center for responsible ai

We are AI comics / Ukrainian







http://r-ai.co/comics

r/ai center for responsible ai

2023

ALL ABOARD! MAKING AI EDUCATION ACCESSIBLE

Authors:

Falaah Arif Khan, Lucius Bynum, Amy Hurst, Lucas Rosenblatt, Meghana Shanbhogue, Mona Sloane, Julia Stoyanovich

http://r-ai.co/AllAboard

center for responsible ai

http://r-ai.co/alice

r/ai center for responsible ai

Looking ahead: A responsible AI “Sputnik” moment

https://science.house.gov/2025/4/deepseek-a-deep-dive

- Robust trustworthy AI cyberinfrastructure - **the NAIRR**
- Responsible data engineering tools and primitives across the lifecycle
- Data protection tools, primitives, guardrails

center for responsible ai

Looking ahead: A responsible AI “Sputnik” moment

Correspondence

https://doi.org/10.1038/s41591-025-03680-0

Promoting transparency in AI for biomedical and behavioral research

Check for updates

Recent advancements in artificial intelligence (AI) in healthcare have highlighted the need for transparency, including explainability, interpretability, and accountability across the AI lifecycle[1,2]. Transparency ensures stakeholders can make informed decisions about data and model reuse, fostering trust and fairness while aligning with regulatory frameworks. However, the concept of transparency lacks a clear definition for both biomedical research and clinical care, resulting in inconsistent practices.

This Correspondence focuses on transparency within the realm of AI-driven biomedical and behavioral research. Although the effect of AI on clinical care is crucial, this discussion centers on its implications for research, addressing gaps in data reuse, model generalization and fairness. The National Institutes of Health (NIH) Office of Data Science Strategy (ODSS) convened a workshop that brought together leading experts in AI, healthcare and ethics to examine transparency in this context[3]. The workshop findings highlight practical solutions tailored to research contexts, addressing documentation standards, patient and community co-design, and oversight mechanisms to achieve equitable outcomes.

Transparency in AI extends beyond technical documentation to include ethical and societal dimensions. We define transparency as a multidimensional construct comprising three pillars: clarity of processes, stakeholder engagement, and accountability mechanisms. Clarity of processes includes documenting model development, training data sources, and decision-making pathways. Stakeholder engagement, such as actively involving patients and communities in the co-design of models and tools, ensures that AI is aligned with societal values. Finally, researchers need to establish oversight frameworks for ethical AI deployment. This approach enables transparency to address specific challenges in biomedical research ensuring ethical use and reuse practices.

Transparency empowers patient and community engagement by enabling active involvement in AI decision-making processes across the AI lifecycle. Including these perspectives in AI governance ensures that research outputs align with ethical standards, fostering trust and accountability between developers, researchers and the communities that they serve. Seeking input on data sovereignty and benefit-sharing further strengthens transparency and ethical decision-making across the healthcare ecosystem. Transparency should extend to educating stakeholders about the impacts of AI on healthcare, incorporating principles such as data sovereignty and benefit-sharing. These strategies support the development of patient-centered AI by ensuring alignment with societal values, fostering trust, and promoting active participation in AI decision-making processes.

Although co-design is not a new concept, the integration of patients and communities in AI development introduces unique challenges. For example, AI systems often involve complex datasets and technical methodologies that may alienate non-technical stakeholders. To address this, it is important to simplify technical concepts through educational resources tailored for diverse audiences. Governance structures are needed to ensure meaningful input from underrepresented communities, and it is essential to establish feedback loops to evaluate the long-term impacts of community involvement. These strategies enable patient-centered AI development, ensuring ethical alignment and fostering trust.

A comprehensive training framework for researchers and clinicians is essential to understand AI's capabilities, limitations and societal impacts[4]. With AI becoming more prevalent and sophisticated, there is an increasing need to educate researchers and practitioners on both the capabilities and limitations of AI technologies. Furthermore, it is important to understand how AI interacts with human decision-making in research and clinical practice. Developing a living compendium of ethical challenges and best practices can guide future work in responsible AI[5–9].

Comprehensive documentation standards, such as model cards and datasheets, are crucial for ethical AI practices[10]. Workshop findings propose augmenting existing tools to further transparency. Establishing metadata registries for datasets and models can track provenance, consent and lifecycle information, ensuring that AI tools are used ethically and effectively. This registry would serve as a centralized record of datasets and models, including their provenance, tracking their origin, data ownership, consent and any modifications. This requires refining existing metadata standards, creating lifecycle-centric methodologies, and integrating semi-automated documentation tools. A registry of proxy variables would also be integral components of the AI lifecycle metadata. Proxy variables are stand-in measures used to represent underlying factors or characteristics that are difficult to observe directly, such as using zip code as a proxy for socioeconomic status or access to healthcare. Equally important are harm-incident databases that track adverse outcomes linked to AI tools to guide future improvements. These databases can identify known biases and potential harms from AI data use. These resources will provide a foundation for ethical AI development and help to mitigate unintended consequences in healthcare AI applications.

Improving the effectiveness, safety and inclusivity of AI in research and practice requires a distributed accountability framework in which all stakeholders contribute to the responsible design, development and oversight of AI technologies. Building institutional capacities in education, training, public engagement and standardization is essential to support these efforts. Institutions should prioritize the development of specialized curricula and certification programs tailored to AI ethics, transparency and bias mitigation. These programs should target diverse stakeholders, including researchers, clinicians and administrators, to ensure a shared understanding of responsible AI practices. Public engagement initiatives must foster open dialogue between the scientific community and society, addressing concerns about AI's impact and building trust.

Expanding Institutional Review Boards (IRBs) to include AI expertise is crucial to address transparency, bias and ethical

U.S. Department of Health and Human Services (HHS)
National Institutes of Health (NIH)
Office of the Director (OD)
Division of Program Coordination, Planning, and Strategic Initiatives (DPCPSI)
Office of Data Science Strategy (ODSS)

Report from the Community Workshop *Toward an Ethical Framework for Artificial Intelligence in Biomedical and Behavioral Research: Transparency for Data and Model Reuse*

Authors

Workshop Co-Chairs *Dr. Tina Hernandez-Boussard, Dr. Aaron Lee, Prof. Julia Stoyanovich;*
Workshop Breakout Leads *Drs. Ansu Chatterjee, Caroline Chung, Maia Hightower, Sajid Hussain, H.V. Jagadish, Jayashree Kalpathy-Cramer, Vincent Liu, Courtney Lyles, Shazia Siddique, Eric Stahlberg, and Colin Walsh*

January 31 - February 2, 2024
Rockledge II, 6701 Rockledge Drive
Bethesda, MD
& Virtual

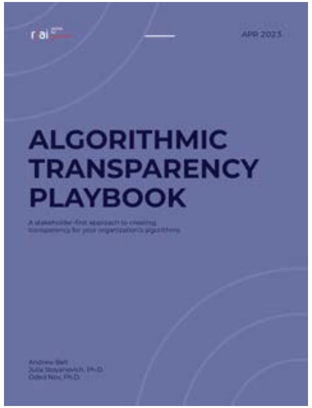
Community-Informed Policies and Best-Practices for the National Artificial Intelligence Research Resource (NAIRR) Workshop Report

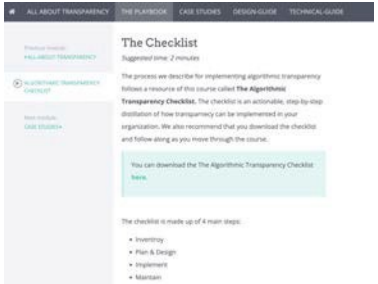
July 29–31, 2024
New York University
New York, NY

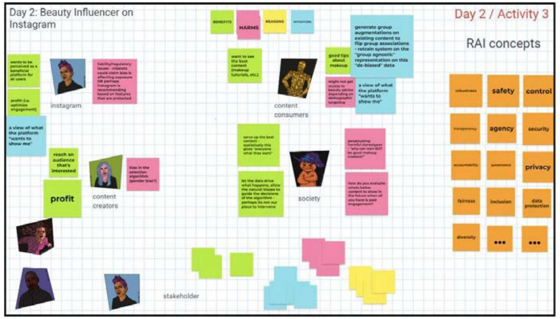
center for responsible ai

What need education and practitioner training on RAI

It's more expensive to do things right (i.e., responsibly) than to do them somehow. **Small and medium-sized businesses that develop and/or use AI need support, training, tools, data, compute, infrastructure.**



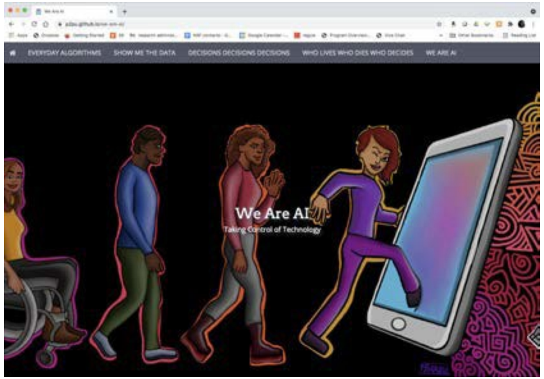




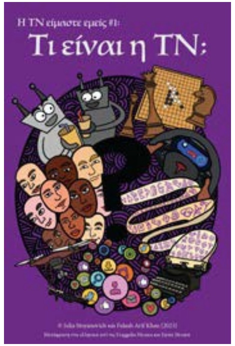
r/ai center for responsible ai

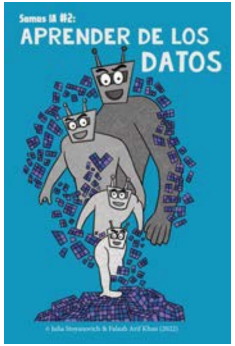
What need AI literacy for everyone!

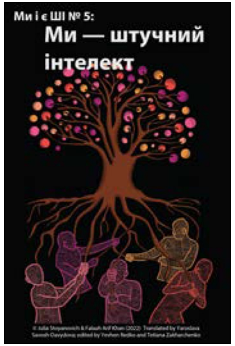
Many of us are passionate about AI literacy, and have invested time and effort into pilot projects.
And now, we need to scale up and scale out.



r/ai center for responsible ai | NYU Ability Project | NORTHEAST BIG DATA INNOVATION HUB

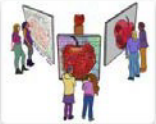






2023
ALL ABOARD!
MAKING AI EDUCATION ACCESSIBLE

Authors:
Falaah Arif Khan, Lucius Bynum, Amy Hurst, Lucas Rosenblatt, Meghana Shanbhogue, Mona Sloane, Julia Stoyanovich



Responsible AI literacy: A stakeholder-first approach
Daniel Dominguez, and Julia Stoyanovich
Big Data and Society 2023
CITE PDF

r/ai center for responsible ai

Thank you to my collaborators!

- Serge Abiteboul
- HV Jagadish
- Bill Howe
- Sebastian Schelter
- Steven Whang
- Joshua Loftus
- Daniel Dominguez
- Armanda Lewis
- George Fletcher
- Andrew Bell
- Lucas Rosenblatt
- Lucius Bynum
- Falaah Arif Khan
- Venetia Pliatsika
- Joao Fonseca
- Stefan Grafberger
- Ke Yang

- Rodrigo Kreis de Paula
- Chloe Zheng
- Yuval Moskovitch
- Meike Zehlike
- Tetiana Zakharchenko
- Denys Herasymuk
- Nazar Protsiv
- Nazar Drushchak
- Katia Akhynko
- Ivan Shevchenko
- Matvii Prytula
- Olha Nahurna
- Anastasia (Stacy) Holovenko
- Andrii Stadnik
- Taras Rumezhak
- Oleksandra Konopatska
- Daniel Hwang

r/ai center for responsible ai

U.S. National Science Foundation

Thank you!

NYU

r/ai center for responsible ai