

Responsible Data Science

Wrapping up

May 8, 2023

Prof. Julia Stoyanovich

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

Recall: applied ethics
(week 9)

Tuskegee Syphilis Study

In 1932, researchers from the US Public Health Service (PHS) enrolled 400 black men from Tuskegee, Alabama, infected with syphilis in a study to study the effects of the disease. The study was **non-therapeutic**: designed to document, not treat!

1932	Approximately 400 men with syphilis are enrolled in the study; they are not informed of the nature of the research
1937-38	The PHS sends mobile treatment units to the area, but treatment is withheld for the men in the study
1942-43	In order to prevent the men in the study from receiving treatment, PHS intervenes to prevent them from being drafted for WWII
1950s	Penicillin becomes a widely available and effective treatment for syphilis; the men in the study are still not treated (<u>Brandt 1978</u>)

The Belmont Report

Belmont Report: Summary

THE BELMONT REPORT

Office of the Secretary

Ethical Principles and Guidelines for the Protection of Human
Subjects of Research

The National Commission for the Protection of Human Subjects of
Biomedical and Behavioral Research

April 18, 1979

- Boundaries between research and practice
- Ethical principles
 - Respect for persons
 - Beneficence
 - Justice
- Applications

Boundaries between research & practice

- **Research** seeks generalizable knowledge, **practice** includes everyday treatment and activities

“For the most part, the term ‘practice’ refers to interventions that are designed solely to enhance the wellbeing of an individual patient or client and that have a reasonable expectation of success. The purpose of medical or behavioral practice is to provide diagnosis, preventive treatment or therapy to particular individuals ... By contrast, the **term 'research' designates an activity designed to test a hypothesis**, permit conclusions to be drawn, and thereby to develop or contribute to generalizable knowledge (expressed, for example, in theories, principles, and statements of relationships). Research is usually described in a formal protocol that sets forth an objective and a set of procedures designed to reach that objective.”

- Belmont Report sets out the ethical principles that apply **only** to research

Principles: Respect for persons (autonomy)

Individuals should be treated as autonomous agents

“To **respect autonomy** is to give weight to autonomous persons' considered opinions and choices while refraining from obstructing their actions unless they are clearly detrimental to others. To show lack of respect for an autonomous agent is to repudiate that person's considered judgments, to deny an individual the freedom to act on those considered judgments, or to **withhold information necessary to make a considered judgment**, when there are no compelling reasons to do so. “

Principles: Autonomy / Respect for Persons

Persons with diminished autonomy are entitled to protection

“In some situations, however, application of the principle is not obvious. The involvement of prisoners as subjects of research provides an instructive example. On the one hand, it would seem that the principle of respect for persons requires that prisoners not be deprived of the opportunity to volunteer for research. On the other hand, under prison conditions they may be subtly coerced or unduly influenced to engage in research activities for which they would not otherwise volunteer. Respect for persons would then dictate that prisoners be protected. Whether to allow prisoners to ‘volunteer’ or to ‘protect’ them presents a dilemma.

Respecting persons, in most hard cases, is often a matter of balancing competing claims urged by the principle of respect itself. “

Principles: Beneficence

Do not harm

Maximize possible benefits and minimize possible harm

“The Hippocratic maxim ‘do no harm’ has long been a fundamental principle of medical ethics. **Claude Bernard extended it to the realm of research, saying that one should not injure one person regardless of the benefits that might come to others.** However, even avoiding harm requires learning what is harmful; and, in the process of obtaining this information, persons may be exposed to risk of harm. Further, the Hippocratic Oath requires physicians to benefit their patients ‘according to their best judgment.’ **Learning what will in fact benefit may require exposing persons to risk.** The problem posed by these imperatives is to decide when it is **justifiable to seek certain benefits despite the risks involved,** and when the benefits should be foregone because of the risks.”

Principles: Justice

Who ought to receive the benefits of research and bear its burdens?

“Questions of justice have long been associated with social practices such as punishment, taxation and political representation. Until recently these questions have not generally been associated with scientific research. However, they are foreshadowed even in the earliest reflections on the ethics of research involving human subjects. For example, during the 19th and early 20th centuries the **burdens of serving as research subjects fell largely upon poor ward patients, while the benefits of improved medical care flowed primarily to private patients. ...**”

Principles: Justice

Who ought to receive the benefits of research and bear its burdens?

“.... Subsequently, the **exploitation of unwilling prisoners** as research subjects in Nazi concentration camps was condemned as a particularly flagrant injustice. In this country, in the 1940's, **the Tuskegee syphilis study** used disadvantaged, rural black men to study the untreated course of a disease that is by no means confined to that population. **These subjects were deprived of demonstrably effective treatment in order not to interrupt the project, long after such treatment became generally available.**”

Application: Assessment of risks and benefits

“Risks and benefits of research may affect the **individual subjects**, the **families** of the individual subjects, and **society** at large (or special groups of subjects in society).

In balancing these different elements, the **risks and benefits affecting the immediate research subject** will normally carry special weight.

Beneficence thus requires that we protect against risk of harm to subjects and also that we be concerned about the **loss of the substantial benefits that might be gained from research.** “

Application: Selection of subjects

Just as the principle of **respect for persons** finds expression in the requirements for consent, and the principle of **beneficence** in risk/benefit assessment, the principle of **justice** gives rise to moral requirements that there be fair procedures and outcomes in the selection of research subjects.

- **Individual justice** in the selection of subjects would require that researchers exhibit fairness: thus, they should not offer potentially beneficial research only to some patients who are in their favor or select only "undesirable" persons for risky research.
- **Social justice** requires that distinction be drawn between classes of subjects that ought, and ought not, to participate in any particular kind of research, based on the ability of members of that class to bear burdens and on the appropriateness of placing further burdens on already burdened persons.

The Menlo Report

The Menlo Report: Summary

The Menlo Report

Ethical Principles Guiding Information and
Communication Technology Research

August 2012

“[...] the Menlo Report calls on researchers to **move beyond the narrow definition of ‘research involving human subjects’ from the Belmont Report** to a more general notion of ‘research with human-harming potential’.

A principles-based approach means that **researchers should not hide behind a narrow, legal definition of ‘research involving human subjects’**, even if IRBs allow it. Rather, they should adopt a more general notion of ‘research with human-harming potential’ and they should subject all of their own research with human-harming potential to ethical consideration.”

The Menlo Report: Summary

Principle	Application
Respect for Persons	Participation as a research subject is voluntary, and follows from informed consent; Treat individuals as autonomous agents and respect their right to determine their own best interests; Respect individuals who are not targets of research yet are impacted; Individuals with diminished autonomy, who are incapable of deciding for themselves, are entitled to protection.
Beneficence	Do not harm; Maximize probable benefits and minimize probable harms; Systematically assess both risk of harm and benefit.
Justice	Each person deserves equal consideration in how to be treated, and the benefits of research should be fairly distributed according to individual need, effort, societal contribution, and merit; Selection of subjects should be fair, and burdens should be allocated equitably across impacted subjects.
<i>Respect for Law and Public Interest</i>	<i>Engage in legal due diligence; Be transparent in methods and results; Be accountable for actions.</i>

Respect for law and public interest

- Implicit in the Belmont Reports' application of Beneficence, but **deserves explicit consideration**
- In Information and Communication Technology Research (ICTR), included as a separate principle with two applications - *Compliance* and *Transparency and Accountability*

“The second application refers to **transparency of methodologies and results**, and accountability for actions. Transparency and accountability serve vital roles in many ICTR contexts where it is challenging or impossible to identify stakeholders (e.g., attribution of sources and intermediaries of information), to understand interactions between highly dynamic and globally distributed systems and technologies, and consequently to **balance associated harms and benefits**. A lack of transparency and accountability risks undermining the credibility of, trust and confidence in, and ultimately support for, ICT research.”

Respect for law and public interest

- Implicit in the Belmont Reports' application of Beneficence, but **deserves explicit consideration**
- In Information and Communication Technology Research (ICTR), included as a separate principle with two applications - *Compliance* and *Transparency and Accountability*

“**Accountability demands that research methodology, ethical evaluations, data collected, and results generated should be documented and made available responsibly in accordance with balancing risks and benefits.** Data should be available for legitimate research, policy-making, or public knowledge, subject to appropriate collection, use, and disclosure controls informed by the Beneficence principle. The appropriate format, scope and modality of the data exposure will vary with the circumstances, as informed by Beneficence determinations.”

informed consent

Application: Informed Consent

“**Respect for persons** requires that subjects, to the degree that they are capable, be given the opportunity to choose what shall or shall not happen to them. This opportunity is provided when adequate standards for informed consent are satisfied.

While the importance of informed consent is unquestioned, **controversy prevails over the nature and possibility** of an informed consent. Nonetheless, there is widespread agreement that **the consent process can be analyzed as containing three elements: information, comprehension and voluntariness. ...**”

Application: Informed Consent

Information, Comprehension, Voluntariness

“Most codes of research establish specific items for disclosure intended to assure that subjects are given sufficient information. These items generally include: the research procedure, their purposes, risks and anticipated benefits, alternative procedures (where therapy is involved), and a statement offering the subject the opportunity to ask questions and to withdraw at any time from the research.

... A special problem of consent arises where informing subjects of some pertinent aspect of the research is likely to impair the validity of the research. ... In all cases of research involving incomplete disclosure, such research is justified only if it is clear that (1) incomplete disclosure is truly necessary to accomplish the goals of the research, (2) there are no undisclosed risks to subjects that are more than minimal, and (3) there is an adequate plan for debriefing subjects, when appropriate, and for dissemination of research results to them. “

Recall: 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.

Recall: Racial bias in resume screening

Research question: Does an employer unlawfully discriminate against applicants based on membership in protected groups?

Employers don't provide consent, in fact, they are actively deceived!

Field experiments to study discrimination are legally permissible **if:**

1. the harm to employers is limited, **and**
2. there is great social benefit to having a reliable measure of discrimination, **and**
3. other methods of measuring discrimination are weak; **and**
4. deception does not strongly violate the norms of that setting.

Application: Informed Consent

Information, **Comprehension**, Voluntariness

“The manner and context in which information is conveyed is as important as the information itself. For example, **presenting information in a disorganized and rapid fashion, allowing too little time for consideration or curtailing opportunities for questioning, all may adversely affect a subject's ability to make an informed choice.**

Because the subject's ability to understand is a function of intelligence, rationality, maturity and language, it is **necessary to adapt the presentation of the information to the subject's capacities.** Investigators are responsible for ascertaining that the subject has comprehended the information. “

Application: Informed Consent

Information, Comprehension, **Voluntariness**

“An agreement to participate in research constitutes a **valid consent only if voluntarily given**. This element of informed consent requires conditions free of coercion and undue influence. [...]

Unjustifiable pressures usually occur when persons in positions of authority or commanding influence -- especially where possible sanctions are involved -- urge a course of action for a subject. A continuum of such influencing factors exists, however, and it is **impossible to state precisely where justifiable persuasion ends and undue influence begins**. But undue influence would include actions such as manipulating a person's choice through the controlling influence of a close relative and threatening to withdraw health services to which an individual would otherwise be entitled.”

Case study: Emotional contagion

Experimental evidence of massive-scale emotional contagion through social networks

Adam D. I. Kramer, Jamie E. Guillory, and Jeffrey T. Hancock

PNAS June 17, 2014 111 (24) 8738-8790; first published June 2, 2014 <https://doi.org/10.1073/pnas.1320040111>

Edited by Susan T. Fiske, Princeton University, Princeton, NJ, and approved March 25, 2014 (received for review October 23, 2013)



Significance

We show, via a massive ($N = 689,003$) experiment on Facebook, that emotional states can be transferred to others via emotional contagion, leading people to experience the same emotions without their awareness. We provide experimental evidence that emotional contagion occurs without direct interaction between people (exposure to a friend expressing an emotion is sufficient), and in the complete absence of nonverbal cues.

Case study: Emotional contagion

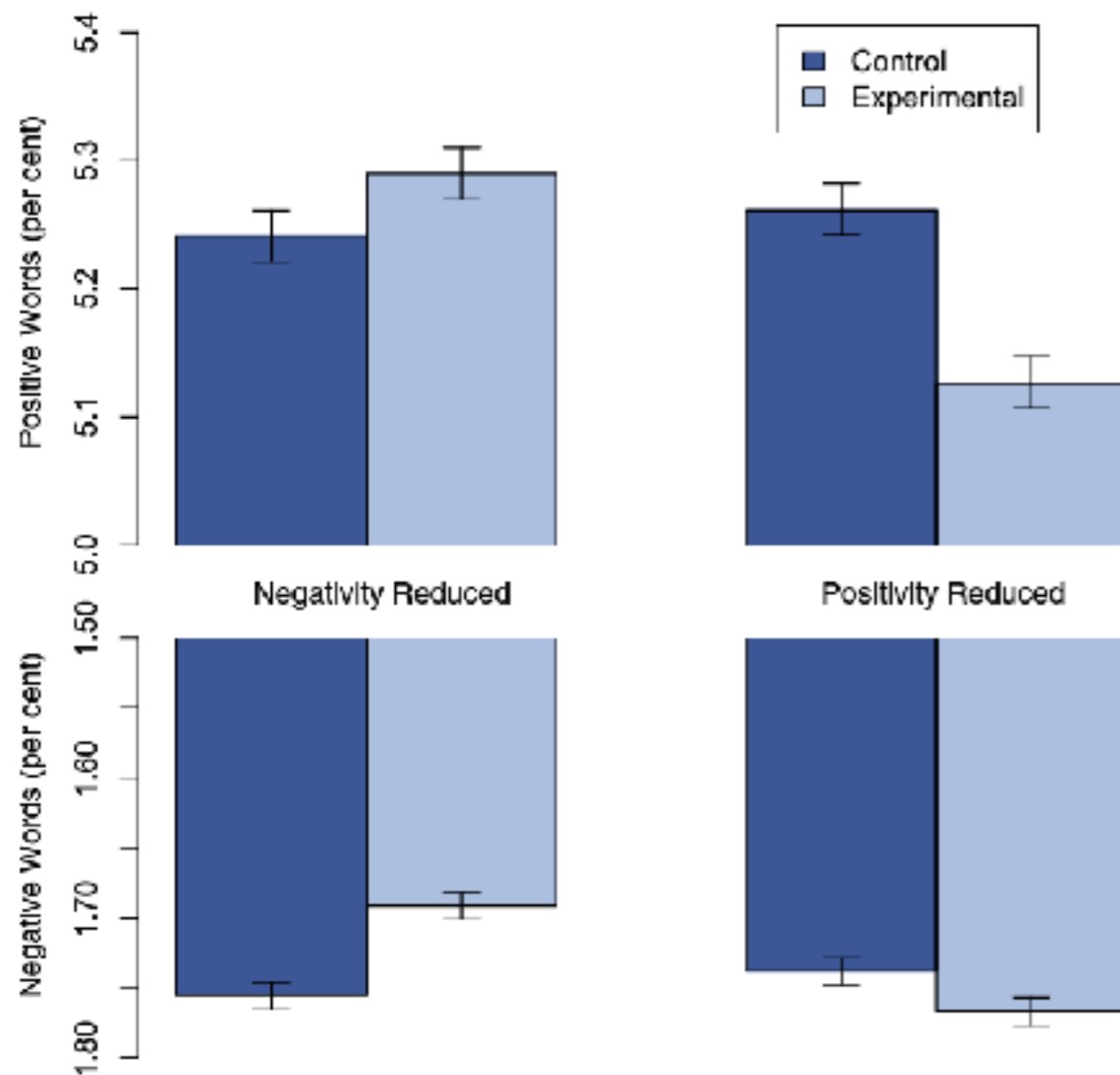


Fig. 1. Mean number of positive (*Upper*) and negative (*Lower*) emotion words (percent) generated people, by condition. Bars represent standard errors.

PNAS

Proceedings of the
National Academy of Sciences
of the United States of America

Experimental evidence of massive-scale emotional contagion through social networks

Adam D. I. Kramer, Jamie E. Guillory, and Jeffrey T. Hancock

PNAS June 17, 2014 111 (24) 8738–8790; first published June 2, 2014 <https://doi.org/10.1073/pnas.1320040111>

Edited by Susan T. Fiske, Princeton University, Princeton, NJ, and approved March 25, 2014 (received for review October 23, 2013)

Case study: Emotional contagion

- **Criticism** in the research community, in the press
 - users were not consented to participate in the study
 - there was no third-party review of study design - Facebook did not even have an IRB at the time
- **Result**
 - PNAS placed a disclaimer on the article
 - Facebook instituted an internal ethics review board

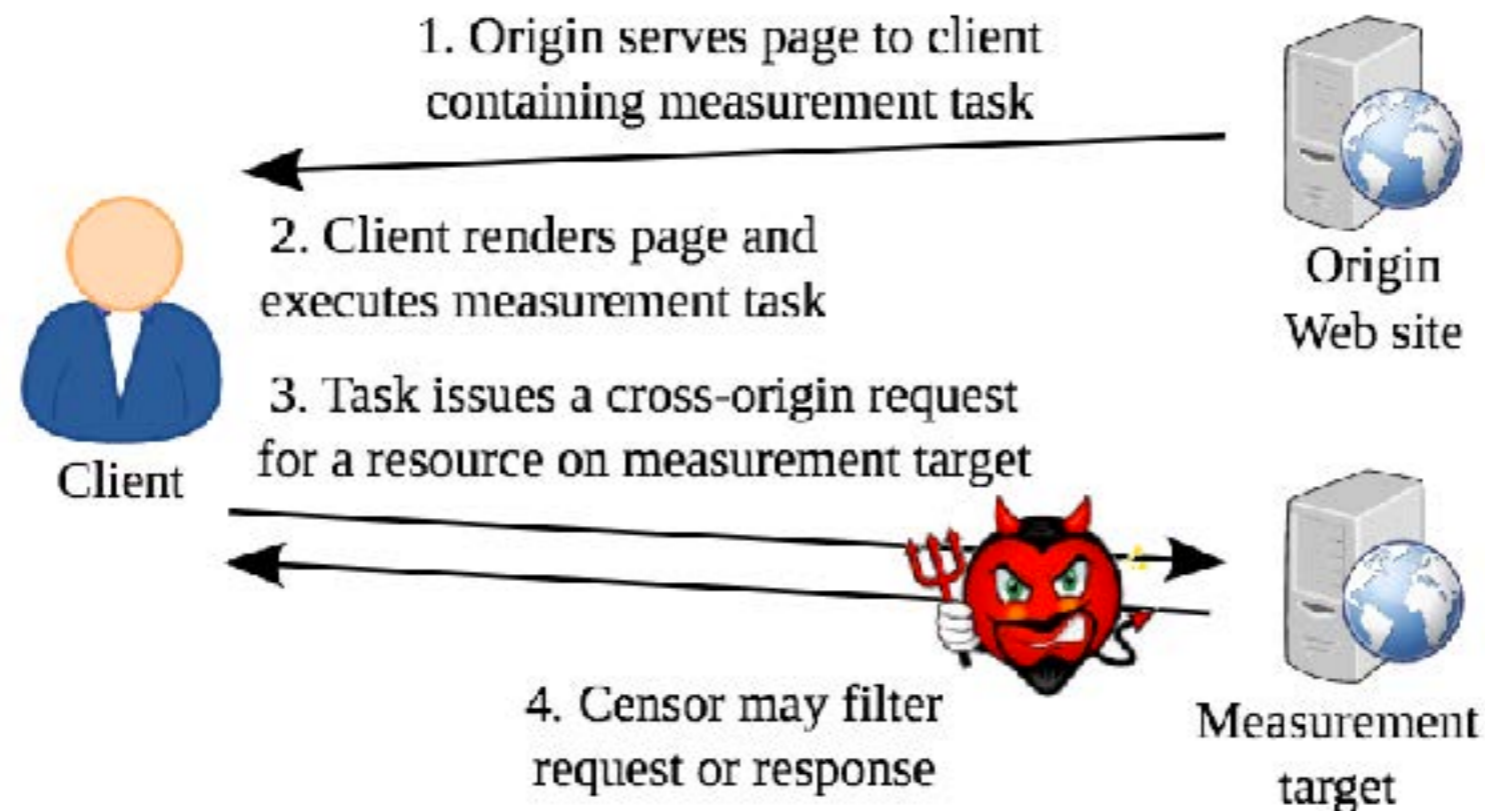
Did Facebook stop running these types of experiments?

Case study: Encore

Encore: Lightweight Measurement of Web Censorship with Cross-Origin Requests

Sam Burnett
School of Computer Science, Georgia Tech
sam.burnett@gatech.edu

Nick Feamster
Department of Computer Science, Princeton
feamster@cs.princeton.edu



Case study: Encore

Encore: Lightweight Measurement of Web Censorship with Cross-Origin Requests ACM SIGCOMM 2015

Sam Burnett

School of Computer Science, Georgia Tech
sam.burnett@gatech.edu

Nick Feamster

Department of Computer Science, Princeton
feamster@cs.princeton.edu

“...We present Encore, a system that harnesses cross-origin requests to **measure Web filtering** from a diverse set of vantage points without requiring users to install custom software, enabling longitudinal measurements from many vantage points. We explain how Encore **induces Web clients to perform cross-origin requests** that measure Web filtering, design a distributed platform for scheduling and collecting these measurements, show the feasibility of a global-scale deployment with a pilot study and an **analysis of potentially censored Web content**, identify several cases of filtering in six months of measurements, and **discuss ethical concerns** that would arise with widespread deployment.”

Case study: Encore

Encore: Lightweight Measurement of Web Censorship with Cross-Origin Requests

ACM SIGCOMM 2015

Sam Burnett

School of Computer Science, Georgia Tech
sam.burnett@gatech.edu

Nick Feamster

Department of Computer Science, Princeton
feamster@cs.princeton.edu

Statement from the SIGCOMM 2015 Program Committee: The SIGCOMM 2015 PC appreciated the technical contributions made in this paper, but found the paper controversial because some of the experiments the authors conducted raise ethical concerns. The controversy arose in large part because the networking research community does not yet have widely accepted guidelines or rules for the ethics of experiments that measure online censorship. In accordance with the published submission guidelines for SIGCOMM 2015, had the authors not engaged with their Institutional Review Boards (IRBs) or had their IRBs determined that their research was unethical, the PC would have rejected the paper without review. But the authors did engage with their IRBs, which did not flag the research as unethical. The PC hopes that discussion of the ethical concerns these experiments raise will advance the development of ethical guidelines in this area. It is the PC's view that future guidelines should include as a core principle that researchers should not engage in experiments that subject users to an appreciable risk of substantial harm absent informed consent. The PC endorses neither the use of the experimental techniques this paper describes nor the experiments the authors conducted.

research - check!
but what about everyday
decision-making?

Example: Algorithmic rankers

Input: database of items (individuals, colleges, cars, ...)

Score-based ranker: computes the score of each item using a **known formula**, often a monotone aggregation function, then sorts items on score

Output: permutation of the items, complete or top-k

\mathcal{D}			f
id	x_1	x_2	$x_1 + x_2$
t_1	0.63	0.71	1.34
t_2	0.72	0.65	1.37
t_3	0.58	0.78	1.36
t_4	0.7	0.68	1.38
t_5	0.53	0.82	1.35
t_6	0.61	0.79	1.4

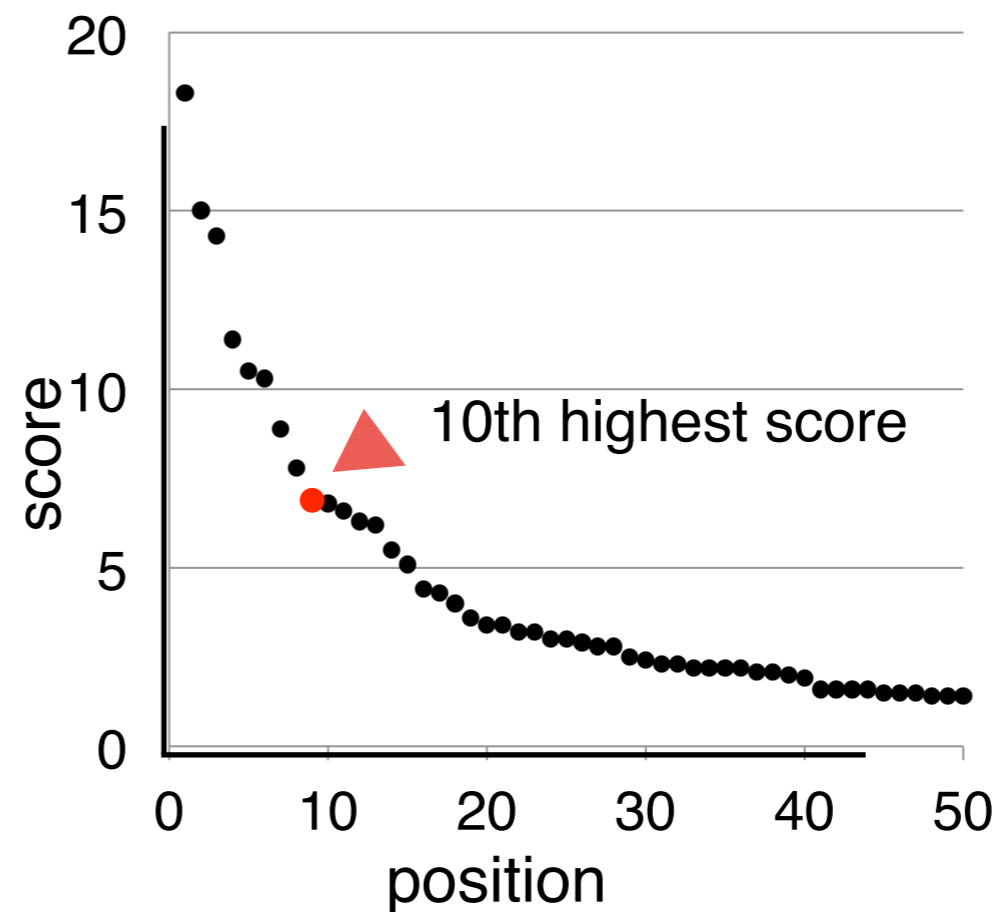
Do we have transparency?

We have syntactic transparency, but lack interpretability!

Opacity in algorithmic rankers

Reason 1: The scoring formula alone does not indicate the relative rank of an item.

Scores are absolute, rankings are relative. Is 5 a good score? What about 10? 15?



Opacity in algorithmic rankers

Reason 2: A ranking may be unstable if there are tied or nearly-tied items.

Rank	Institution	Average Count	Faculty
1	▶ Carnegie Mellon University	18.4	123
2	▶ Massachusetts Institute of Technology	15.6	64
3	▶ Stanford University	14.8	56
4	▶ University of California - Berkeley	11.5	50
5	▶ University of Illinois at Urbana-Champaign	10.6	56
6	▶ University of Washington	10.3	50
7	▶ Georgia Institute of Technology	8.9	81
8	▶ University of California - San Diego	8	51
9	▶ Cornell University	7	45
10	▶ University of Michigan	6.8	63
11	▶ University of Texas - Austin	6.6	43
12	▶ University of Massachusetts - Amherst	6.4	47

Opacity in algorithmic rankers

Reason 3: A ranking methodology may be unstable: small changes in weights can trigger significant re-shuffling.

THE NEW YORKER

DEPT. OF EDUCATION FEBRUARY 14 & 21, 2011 ISSUE

THE ORDER OF THINGS

What college rankings really tell us.



By Malcolm Gladwell

1. Porsche Cayman 193

2. Chevrolet Corvette 186

3. Lotus Evora 182

1. Chevrolet Corvette 205

2. Lotus Evora 195

3. Porsche Cayman 195

1. Lotus Evora 205

2. Porsche Cayman 198

3. Chevrolet Corvette 192

Opacity in algorithmic rankers

Reason 4: The weight of an attribute in the scoring formula does not determine its impact on the outcome.

Rank	Name	Avg Count	Faculty	Pubs	GRE
1	CMU	18.3	122	2	791
2	MIT	15	64	3	772
3	Stanford	14.3	55	5	800
4	UC Berkeley	11.4	50	3	789
5	UIUC	10.5	55	3	772
6	UW	10.3	50	2	796
		...			
39	U Chicago	2	28	2	779
40	UC Irvine	1.9	28	2	787
41	BU	1.6	15	2	783
41	U Colorado Boulder	1.6	32	1	761
41	UNC Chapel Hill	1.6	22	2	794
41	Dartmouth	1.6	18	2	794

Scoring function
 $0.2 * faculty +$
 $0.3 * avg\ cnt +$
 $0.5 * gre$

Ranking matter!

THE NEW YORKER

DEPT. OF EDUCATION FEBRUARY 14 & 21, 2011 ISSUE

THE ORDER OF THINGS

What college rankings really tell us.



By Malcolm Gladwell

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



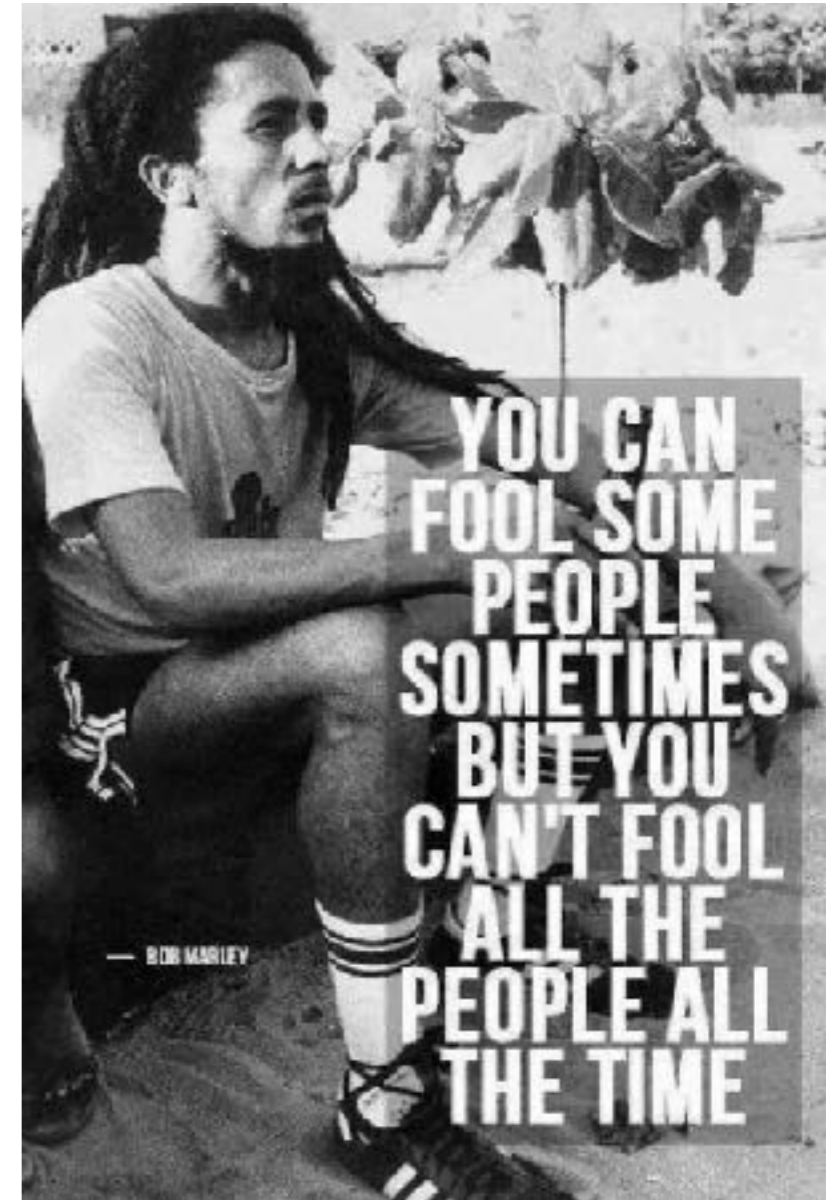
Interpretability in the service of trust!

Gladwell makes the point that rankings are claiming objectivity, yet are comparing apples and oranges.

In that sense, **a score-based ranker is a quintessential “black box” of data science**, and perhaps the simplest possible such black box.

AI is a red herring, privacy / IP / gaming arguments are overused. The truly difficult issues are that:

1. using math to pretend that we are correct when making intrinsically subjective decisions reinforcing the balance of power in society
2. math / objectivity is used as a substitute for trust, but **trust must run deeper than math!**
3. need to find the kind of an interpretability that support **informed consent, recourse, agency**, enable **trust!**



data protection:
the GDPR

GDPR

Chapter 1 (Art. 1 – 4)

General provisions

Chapter 2 (Art. 5 – 11)

Principles

Chapter 3 (Art. 12 – 23)

Rights of the data subject

Chapter 4 (Art. 24 – 43)

Controller and processor

Chapter 5 (Art. 44 – 50)

Transfers of personal data to third countries or international organisations

Chapter 6 (Art. 51 – 59)

Independent supervisory authorities

Chapter 7 (Art. 60 – 76)

Cooperation and consistency

Chapter 8 (Art. 77 – 84)

Remedies, liability and penalties

Chapter 9 (Art. 85 – 91)

Provisions relating to specific processing situations

Chapter 10 (Art. 92 – 93)

Delegated acts and implementing acts

Chapter 11 (Art. 94 – 99)

Final provisions

General Data Protection Regulation

GDPR

Welcome to gdpr-info.eu. Here you can find the official [PDF](#) of the Regulation (EU) 2016/679 (General Data Protection Regulation) in the current version of the OJ L 119, 04.05.2016; cor. OJ L 127, 23.5.2018 as a neatly arranged website. All Articles of the GDPR are linked with suitable recitals. The European Data Protection Regulation is applicable as of May 25th, 2018 in all member states to harmonize data privacy laws across Europe. If you find the page useful, feel free to support us by sharing the project.

Quick Access

[Chapter 1](#) – [1](#) [2](#) [3](#) [4](#)[Chapter 2](#) – [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [11](#)[Chapter 3](#) – [12](#) [13](#) [14](#) [15](#) [16](#) [17](#) [18](#) [19](#) [20](#) [21](#) [22](#) [23](#)[Chapter 4](#) – [24](#) [25](#) [26](#) [27](#) [28](#) [29](#) [30](#) [31](#) [32](#) [33](#) [34](#) [35](#) [36](#) [37](#) [38](#) [39](#) [40](#) [41](#) [42](#) [43](#)[Chapter 5](#) – [44](#) [45](#) [46](#) [47](#) [48](#) [49](#) [50](#)[Chapter 6](#) – [51](#) [52](#) [53](#) [54](#) [55](#) [56](#) [57](#) [58](#) [59](#)[Chapter 7](#) – [60](#) [61](#) [62](#) [63](#) [64](#) [65](#) [66](#) [67](#) [68](#) [69](#) [70](#) [71](#) [72](#) [73](#) [74](#) [75](#) [76](#)[Chapter 8](#) – [77](#) [78](#) [79](#) [80](#) [81](#) [82](#) [83](#) [84](#)[Chapter 9](#) – [85](#) [86](#) [87](#) [88](#) [89](#) [90](#) [91](#)

adopted in April 2016

enforced since May 25, 2018

GDPR: scope and definitions

Article 2: Material Scope

- This Regulation applies to the processing of personal data wholly or partly by automated means and to the processing other than by automated means of personal data which form part of a filing system or are intended to form part of a filing system.

Article 4: Definitions

- **‘personal data’** means any information relating to an identified or identifiable natural person (**‘data subject’**); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person;
- **‘processing’** means **any operation** or set of operations which is performed on personal data or on sets of personal data, **whether or not by automated means**, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction;

GDPR: scope and definitions

Article 4: Definitions

- **‘controller’** means the natural or legal person, public authority, agency or other body which, alone or jointly with others, **determines the purposes and means of the processing** of personal data; where the purposes and means of such processing are determined by Union or Member State law, the controller or the specific criteria for its nomination may be provided for by Union or Member State law;
- **‘processor’** means a natural or legal person, public authority, agency or other body which **processes personal data on behalf of the controller**;
- **‘consent’** of the data subject means any **freely given, specific, informed and unambiguous** indication of the data subject’s wishes by which he or she, by a statement or by a clear affirmative action, **signifies agreement to the processing of personal data** relating to him or her;

Art. 7 GDPR

Conditions for consent

1. Where processing is based on consent, the controller shall be able to demonstrate that the data subject has consented to processing of his or her personal data.
2. ¹ If the data subject's consent is given in the context of a written declaration which also concerns other matters, the request for consent shall be presented in a manner which is clearly distinguishable from the other matters, in an intelligible and easily accessible form, using clear and plain language. ² Any part of such a declaration which constitutes an infringement of this Regulation shall not be binding.

Art. 7 GDPR

Conditions for consent

3. ¹ The data subject shall have the right to withdraw his or her consent at any time.
² The withdrawal of consent shall not affect the lawfulness of processing based on consent before its withdrawal. ³ Prior to giving consent, the data subject shall be informed thereof. ⁴ It shall be as easy to withdraw as to give consent.
4. When assessing whether consent is freely given, utmost account shall be taken of whether, *inter alia*, the performance of a contract, including the provision of a service, is conditional on consent to the processing of personal data that is not necessary for the performance of that contract.

Chapter 3

Rights of the data subject

Section 1 – Transparency and modalities

Article 12 – Transparent information, communication and modalities for the exercise of the rights of the data subject

Section 2 – Information and access to personal data

Article 13 – Information to be provided where personal data are collected from the data subject

Article 14 – Information to be provided where personal data have not been obtained from the data subject

Article 15 – Right of access by the data subject

Chapter 3

Rights of the data subject

Section 3 – Rectification and erasure

Article 16 – Right to rectification

Article 17 – Right to erasure ('right to be forgotten')

Article 18 – Right to restriction of processing

Article 19 – Notification obligation regarding rectification or erasure of personal data or restriction of processing

Article 20 – Right to data portability

Removing personal data

The right to be forgotten (Article 17)

- Similar laws exist in other jurisdictions, e.g., Argentina (since 2006)
- Resulted in many dereferencing requests to search engines
- Often seen as controversial: **reasons?**
- May conflict with other legal requirements, or with technical requirements

Also, technically challenging:

- have to re-engineer the data management stack, **what are the issues?**
- what about models?

Chapter 3

Rights of the data subject

Section 3 – Rectification and erasure

Article 16 – Right to rectification

Article 17 – Right to erasure ('right to be forgotten')

Article 18 – Right to restriction of processing

Article 19 – Notification obligation regarding rectification or erasure of personal data or restriction of processing

Article 20 – Right to data portability

Moving personal data

The right to data portability (Article 20)

- Aims to prevent vendor lock-in
- What are some technical difficulties?
 - Suppose you want to move your photos from Service A to Service B?
 - What about moving your social interactions from Service A to Service B?
- Can we look at this from the point of view of **inter-operability** rather than moving data?

Moving personal data



[Download White Paper](#)

[About](#) [Community](#) [Documentation](#) [Updates](#) [FAQ](#)

About us

The Data Transfer Project was launched in 2018 to create an open-source, service-to-service data portability platform so that all individuals across the web could easily move their data between online service providers whenever they want.

The contributors to the Data Transfer Project believe portability and interoperability are central to innovation. Making it easier for individuals to choose among services facilitates competition, empowers individuals to try new services and enables them to choose the offering that best suits their needs.

Current contributors include:



What is the Data Transfer Project

Data Transfer Project (DTP) is a collaboration of organizations committed to building a common framework with open-source code that can connect any two online service providers, enabling a seamless, direct, user-initiated portability of data between the two platforms.

[Learn More](#)



Chapter 3

Rights of the data subject

Section 4 – **Right to object and automated individual decision-making**

Article 21 – **Right to object**

Article 22 – **Automated individual decision-making, including profiling**

Recital 58

The principle of transparency*

¹ The principle of transparency requires that any information addressed to the public or to the data subject be concise, easily accessible and easy to understand, and that clear and plain language and, additionally, where appropriate, visualisation be used. ² Such information could be provided in electronic form, for example, when addressed to the public, through a website. ³ This is of particular relevance in situations where the proliferation of actors and the technological complexity of practice make it difficult for the data subject to know and understand whether, by whom and for what purpose personal data relating to him or her are being collected, such as in the case of online advertising.

⁴ Given that children merit specific protection, any information and communication, where processing is addressed to a child, should be in such a clear and plain language that the child can easily understand.

from data to impacts:
algorithmic impact
statements

Regulating ADS?

Precautionary



@FalaahArifKhan

Nah! I'm fine!



@FalaahArifKhan

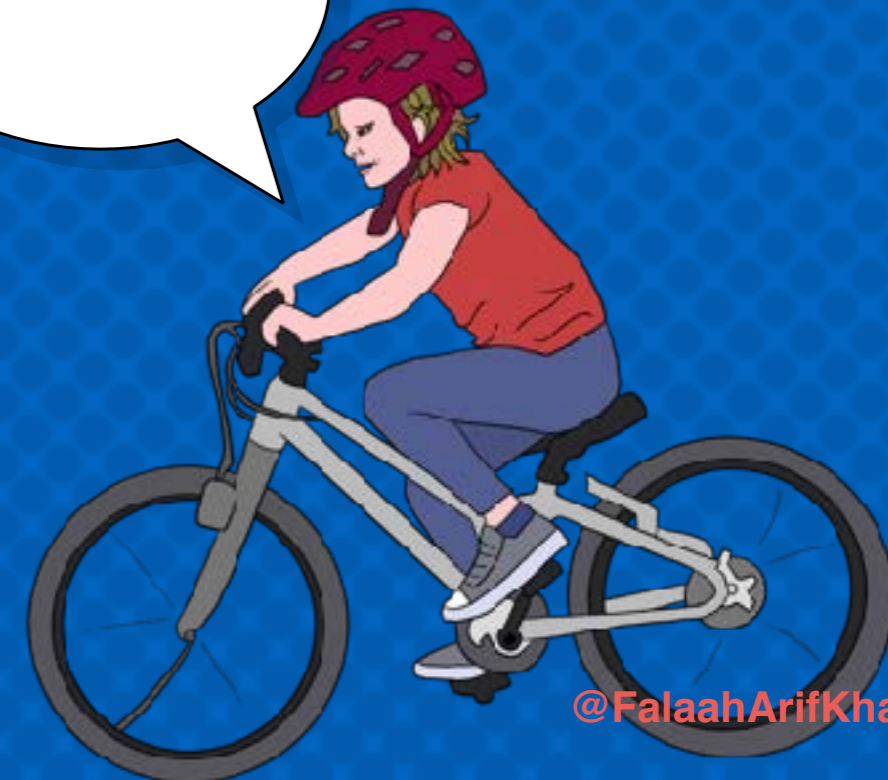


The Anti-Elon 
@antiElon

Regulation rocks!

 2.3K  9.2K  126K

Risk-based



@FalaahArifKhan

Setting the stage: “Big Data Policing”

“Despite its growing popularity, predictive policing is in its relative infancy and is still mostly hype. Current prediction is akin to early weather forecasting, and, like Big Data approaches in other sectors, mixed evidence exists about its effectiveness.

Cities such as Los Angeles, Atlanta, Santa Cruz, and Seattle have enlisted the predictive policing software company PredPol to predict where property crimes will occur. Santa Cruz reportedly “saw burglaries drop by 11% and robberies by 27% in the first year of using [PredPol’s] software.” Similarly, Chicago’s Strategic Subject List—or “heat list”—of people most likely to be involved in a shooting had, as of mid-2016, predicted more than 70% of the people shot in the city, according to the police.

But two rigorous academic evaluations of predictive policing experiments, one in Chicago and another in Shreveport, have shown no benefit over traditional policing. **A great deal more study is required to measure both predictive policing’s benefits and its downsides.** “

what are the potential benefits?

what are the potential downsides?

How to regulate “Big Data Policing”

“While policing is just one of many aspects of society being upended by machine learning, and potentially exacerbating disparate impact in a hidden way as a result, it is a particularly useful case study because of how little our legal system is set up to regulate it.”

The Fourth Amendment: The right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated, and no Warrants shall issue, but upon probable cause, supported by Oath or affirmation, and particularly describing the place to be searched, and the persons or things to be seized.

“[...] the Fourth Amendment’s reasonable suspicion requirement is inherently a “small data doctrine,” rendering it impotent in even its primary uses when it comes to data mining.”

new legal strategies are needed

How to regulate “Big Data Policing”

“ Regarding predictive policing specifically, **society lacks basic knowledge and transparency about both the technology’s efficacy and its effects on vulnerable populations**. Thus, this Article proposes a regulatory solution designed to fill this knowledge gap—to **make the police do their homework** and show it to the public before buying or building these technologies.”

Main contribution: Algorithmic Impact Statements (AISs)

“Impact statements are designed to **force consideration of the problem at an early stage**, and to document the process so that the public can learn what is at stake, **perhaps as a precursor to further regulation**. The primary problem is that no one, including the police using the technology, yet knows what the results of its use actually are.”

Algorithmic Impact Statements (AISs)

- Modeled on the Environmental Impact Statements (EISs) of the 1969 National Environmental Policy Act (NEPA)
- GDPR requires “data protection impact assessments (DPIAs) whenever data processing “is likely to result in a high risk to the rights and freedoms of natural persons”
- Privacy impact statements (PIAs) are used to assess the risks of using personally identifiable information by IT systems

The gist:

- Explore and evaluate all reasonable alternatives
- Include the alternative of “No Action”
- Include appropriate mitigation measures
- Provide opportunities for public comment

Canadian directive on
automated decision-making



Government
of Canada

Gouvernement
du Canada



[Home](#) → [How government works](#) → [Policies, directives, standards and guidelines](#)

Directive on Automated Decision-Making

The Government of Canada is increasingly looking to utilize artificial intelligence to make, or assist in making, administrative decisions to improve service delivery. The Government is committed to doing so in a manner that is compatible with core administrative law principles such as transparency, accountability, legality, and procedural fairness. Understanding that this technology is changing rapidly, this Directive will continue to evolve to ensure that it remains relevant.

Date modified: 2019-02-05

- Took effect on **April 1, 2019**, compliance by **April 1, 2020**
- Applies to any ADS developed or procured after April 1, 2020
- Reviewed automatically every 6 months

Definitions

Appendix A: Definitions

- **Administrative Decision** Any decision that is made by an authorized official of an institution as identified in section 9 of this Directive pursuant to powers conferred by an Act of Parliament or an order made pursuant to a prerogative of the Crown that affects legal rights, privileges or interests.
- **Algorithmic Impact Assessment** A framework to help institutions better understand and reduce the risks associated with Automated Decision Systems and to provide the appropriate governance, oversight and reporting/audit requirements that best match the type of application being designed.
- **Automated Decision System** Includes any technology that either assists or replaces the judgement of human decision-makers. These systems draw from fields like statistics, linguistics, and computer science, and use techniques such as rules-based systems, regression, predictive analytics, machine learning, deep learning, and neural nets.

Objectives

Section 4: Objectives and Expected Results

- **4.1** The objective of this Directive is to ensure that Automated Decision Systems are deployed in a manner that **reduces risks** to Canadians and federal institutions, and **leads to more efficient, accurate, consistent, and interpretable decisions** made pursuant to Canadian law.
- **4.2** The expected results of this Directive are as follows:
 - Decisions made by federal government departments are data-driven, responsible, and complies with procedural fairness and due process requirements.
 - Impacts of algorithms on administrative decisions are assessed and negative outcomes are reduced, when encountered.
 - Data and information on the use of Automated Decision Systems in federal institutions are made available to the public, when appropriate.

Requirements

Section 6.1: Algorithmic Impact Assessment (excerpt)

- **6.1.1 Completing** an Algorithmic Impact Assessment **prior to the production** of any Automated Decision System.
- **6.1.2 ...**
- **6.1.3 Updating** the Algorithmic Impact Assessment when system functionality or the scope of the Automated Decision System changes.
- **6.1.4 Releasing the final results of Algorithmic Impact Assessments** in an accessible format via Government of Canada websites and any other services designated by the Treasury Board of Canada Secretariat pursuant to the Directive on Open Government.

Requirements

Section 6.2: Transparency

- providing notice **before** decisions
- providing explanations **after** decisions
- access to components
- release of source code, unless it's classified Secret, Top Secret or Protected C

Impact Assessment Levels

Decisions classified w.r.t. impact on:

- the rights of individuals or communities,
- the health or well-being of individuals or communities,
- the economic interests of individuals, entities, or communities,
- the ongoing sustainability of an ecosystem.

Level I: no impact: impacts are reversible and brief

Level II: moderate: impacts are likely reversible and short-term

Level III: high: impacts are difficult to reversible and ongoing

Level IV: very high: impacts are irreversible and perpetual

higher impact levels lead to more stringent requirements

regulating ADS in
New York City

How it started: The Vacca bill

Int. No. 1696

August 16, 2017

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:

1 Section 1. Section 23-502 of the administrative code of the city of New York is amended

2 to add a new subdivision g to read as follows:

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

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

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

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

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

How it started: The Vacca bill

THE
NEW YORKER

By Julia Powles December 20, 2017

ELEMENTS

NEW YORK CITY'S BOLD, FLAWED ATTEMPT TO MAKE ALGORITHMS ACCOUNTABLE



Automated systems guide the allocation of everything from firehouses to food stamps. So why don't we know more about them?

Photograph by Mario Tama / Getty

October 16, 2017



https://dataresponsibly.github.io/documents/Stoyanovich_VaccaBill.pdf

How it's going: NYC Local Law 49

January 11, 2018



The screenshot shows the official website of the New York City Council. The header includes the Council's name, the Speaker's name (Corey Johnson), and the Legislative Research Center. Navigation tabs for Council Home, Legislation, Calendar, City Council, and Committees are visible. The main content area displays details for a specific bill, including its file number, type, name, status, committee, and enactment date. A list of sponsors and a summary of the bill are also provided.

THE NEW YORK CITY COUNCIL
Corey Johnson, Speaker

LEGISLATIVE RESEARCH CENTER

Council Home Legislation Calendar City Council Committees

RSS Alerts

Details Reports

File #: Int 1696-2017 Version: A
Name: Automated decision systems used by agencies.
Type: Introduction Status: Enacted
Committee: [Committee on Technology](#)
On agenda: 8/24/2017
Enactment date: 1/11/2018 Law number: 2018/049
Title: A Local Law in relation to automated decision systems used by agencies
Sponsors: [James Vacca](#), [Helen K. Rosenthal](#), [Corey D. Johnson](#), [Rafael Salamanca, Jr.](#), [Vincent J. Gentile](#), [Robert E. Cornegy, Jr.](#), [Jumaane D. Williams](#), [Ben Kallos](#), [Carlos Menchaca](#)
Council Member Sponsors: 9
Summary: This bill would require the creation of a task force that provides recommendations on how information on agency automated decision systems may be shared with the public and how agencies may address instances where people are harmed by agency automated decision systems.
Indexes: Oversight
Attachments: 1. [Summary of Int. No. 1696-A](#), 2. [Summary of Int. No. 1696](#), 3. [Int. No. 1696](#), 4. [August 24, 2017 - Stated Meeting Agenda with Links to Files](#), 5. [Committee Report 10/16/17](#), 6. [Hearing Testimony 10/16/17](#), 7. [Hearing Transcript 10/16/17](#), 8. [Proposed Int. No. 1696-A - 12/12/17](#), 9. [Committee Report 12/7/17](#), 10. [Hearing Transcript 12/7/17](#), 11. [December 11, 2017 - Stated Meeting Agenda with Links to Files](#), 12. [Hearing Transcript - Stated Meeting 12-11-17](#), 13. [Int. No. 1696-A \(FINA\)](#), 14. [Fiscal Impact Statement](#), 15. [Legislative Documents - Letter to the Mayor](#), 16. [Local Law 49](#), 17. [Minutes of the Stated Meeting - December 11, 2017](#)

How it's going: NYC Local Law 49

January 11, 2018

An **Automated Decision System (ADS)** is a “computerized implementation of algorithms, including those derived from machine learning or other data processing or artificial intelligence techniques, which are used to make or assist in making decisions.”

Form task force that surveys the current use of ADS in City agencies and develops procedures for:

- requesting and receiving an **explanation** of an algorithmic decision affecting an individual (3(b))
- interrogating ADS for **bias and discrimination** against members of legally-protected groups (3(c) and 3(d))
- allowing the **public** to **assess** how ADS function and are used (3(e)), and archiving ADS together with the data they use (3(f))

The ADS task force

May 16, 2018

Visit alpha.nyc.gov to help us test out new ideas for NYC's website.

The Official Website of the City of New York **NYC** 简体中文 Translate Text Size

Home NYC Resources NYC311 **Office of the Mayor** Events Connect Jobs Search

Mayor First Lady News Officials

SHARE

Mayor de Blasio Announces First-In-Nation Task Force To Examine Automated Decision Systems Used By The City

May 16, 2018

NEW YORK— Today, Mayor de Blasio announced the creation of the Automated Decision Systems Task Force which will explore how New York City uses algorithms. The task force, the first of its kind in the U.S., will work to develop a process for reviewing "automated decision systems," commonly known as algorithms, through the lens of equity, fairness and accountability.

"As data and technology become more central to the work of city government, the algorithms we use to aid decision making must be aligned with our goals and values," said **Mayor de Blasio**. "The establishment of the Automated Decision Systems Task Force is an important first step towards greater transparency and equity in our use of technology."

Email Print

The ADS task force

April 15, 2019

POLICY \ REPORT \ IS & WORLD

New York City's algorithm task force is fracturing

Some members say the city isn't being transparent

By Colin Lecher | @colinlecher | Apr 15, 2019, 8:43am EDT



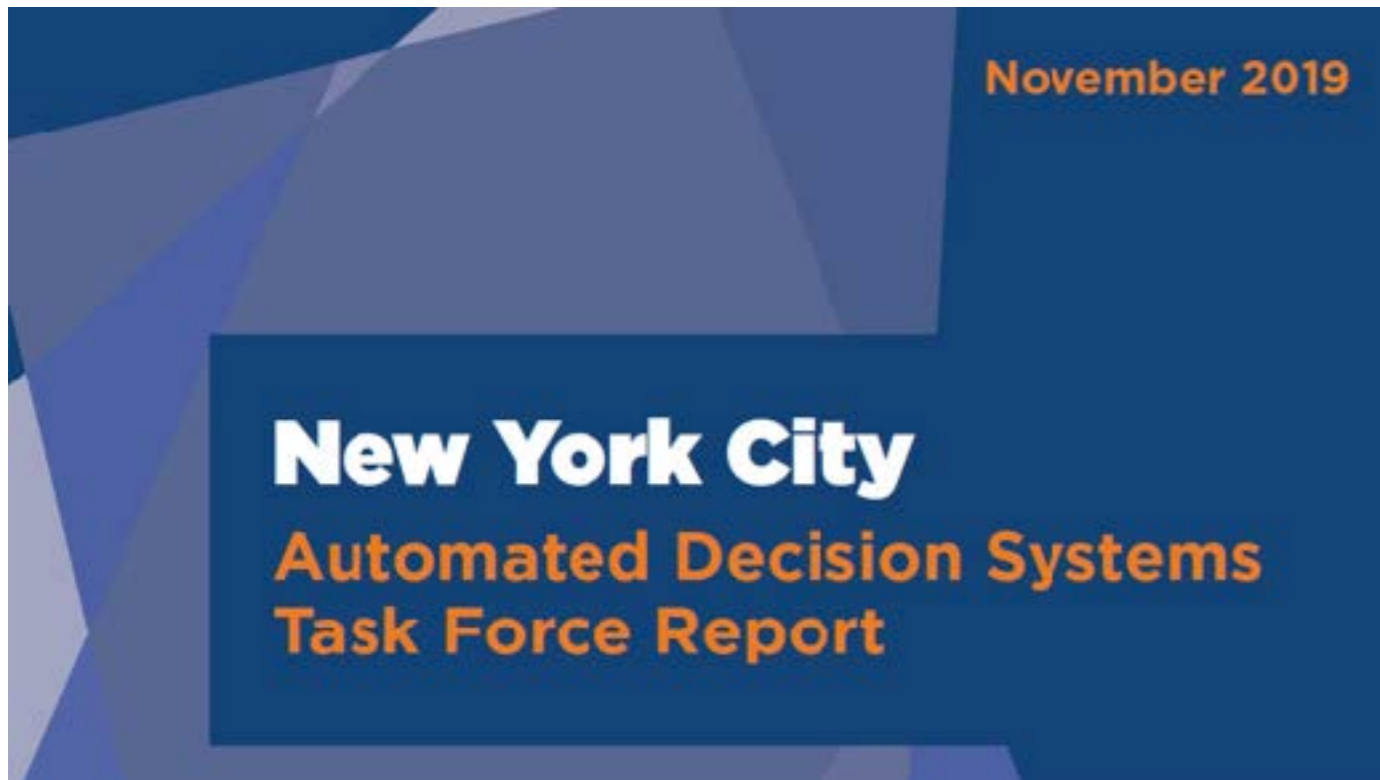
With nothing to study, critics say, the task force is toothless and able to provide only broad policy recommendations ...

New York University assistant professor and task force member Julia Stoyanovich told *The Verge* that **if no examples are forthcoming, “then there was really no point in forming the task force at all.”**

https://dataresponsibly.github.io/documents/StoyanovichBarocas_April4,2019testimony.pdf

The ADS task force

November 19, 2019



THE CITY OF NEW YORK
OFFICE OF THE MAYOR
NEW YORK, N. Y. 10007

EXECUTIVE ORDER No. 50

November 19, 2019

ESTABLISHING AN

ALGORITHMS MANAGEMENT AND POLICY OFFICER

<https://www1.nyc.gov/assets/adstaskforce/downloads/pdf/ADS-Report-11192019.pdf>

<https://www1.nyc.gov/assets/home/downloads/pdf/executive-orders/2019/eo-50.pdf>

ADS task force report



@FalaahArifKhan

Principles

- using ADS **where** they promote innovation and efficiency in service delivery
- promoting **fairness, equity, accountability, and transparency** in the use of ADS
- reducing potential harm **across the lifespan** of ADS

Recommendations

- formalize ADS management functions
- build the City's ADS management capacity
- broaden public conversation on ADS

so what's algorithmic
transparency?

Point 1

algorithmic transparency is not
synonymous with releasing the source
code

publishing source code helps, but it is sometimes
unnecessary and often insufficient

Point 2

**algorithmic transparency requires data
transparency**

data is used in training, validation, deployment

validity, accuracy, applicability can only be
understood in the data context

data transparency is necessary for all ADS, not
only for ML-based systems

Point 3

**data transparency is not synonymous
with making all data public**

release data whenever possible;

also release:

data selection, collection and pre-processing methodologies; data provenance and quality information; known sources of bias; privacy-preserving statistical summaries of the data

Data Synthesizer



input

uid	sex	race	MarriageSta	DateOfBirth	age	lev	fel	conv	decile	score
1	1	0	1	4/18/47	69	0	0	0	1	
2	2	0	2	1/22/82	34	0	0	0	8	
3	3	0	2	5/14/91	24	0	0	0	4	
4	3	0	2	1/21/99	23	0	0	0	8	
5	4	0	1	2/22/73	43	0	0	0	1	
6	5	0	1	3/22/71	44	0	0	0	1	
7	6	0	1	7/23/74	41	0	0	0	4	
8	7	0	3	2/25/73	43	0	0	0	4	
9	8	0	1	8/10/94	21	0	0	0	8	
10	9	0	3	6/1/88	27	0	0	0	4	
11	10	0	3	8/22/78	37	0	0	0	1	
12	11	1	3	12/27/74	41	0	0	0	4	
13	12	0	2	6/14/68	47	0	0	0	1	
14	13	1	3	3/25/85	31	0	0	0	8	
15	14	0	2	1/25/79	37	0	0	0	1	
16	15	0	4	4/22/90	25	0	0	0	10	
17	16	0	2	12/24/84	31	0	0	0	5	
18	17	0	3	1/8/85	31	0	0	0	3	
19	18	0	2	6/28/51	64	0	0	0	6	
20	19	0	2	11/29/94	21	0	0	0	9	
21	20	0	2	8/6/88	27	0	0	0	2	
22	21	0	3	3/22/95	21	0	0	0	4	
23	22	1	3	1/23/92	24	0	0	0	4	
24	23	0	4	1/10/75	43	0	0	0	1	
25	24	0	1	8/24/83	32	0	0	0	3	
26	25	0	1	2/8/89	27	0	0	0	3	
27	26	0	1	5/3/79	36	0	0	0	3	
28	27	1	3	2/28/82	32	0	0	0	3	

Data
Describer



summary

age	int	min=23 max=60	32% mis	
name	str	length 10 to 98	no mis	
sex	str	cat	10% mis	

Data
Generator



output

uid	sex	race	MarriageSta	DateOfBirth	age	lev	fel	conv	decile	score
1	1	0	1	4/18/47	69	0	0	0	1	
2	2	0	2	1/22/82	34	0	0	0	8	
3	3	0	2	5/14/91	24	0	0	0	4	
4	3	0	2	1/21/99	23	0	0	0	8	
5	4	0	1	2/22/73	43	0	0	0	1	
6	5	0	1	3/22/71	44	0	0	0	1	
7	6	0	1	7/23/74	41	0	0	0	4	
8	7	0	3	2/25/73	43	0	0	0	4	
9	8	0	1	8/10/94	21	0	0	0	8	
10	9	0	3	6/1/88	27	0	0	0	4	
11	10	0	3	8/22/78	37	0	0	0	1	
12	11	1	3	12/27/74	41	0	0	0	4	
13	12	0	2	6/14/68	47	0	0	0	1	
14	13	1	3	3/25/85	31	0	0	0	8	
15	14	0	2	1/25/79	37	0	0	0	1	
16	15	0	4	4/22/90	25	0	0	0	10	
17	16	0	2	12/24/84	31	0	0	0	5	
18	17	0	3	1/8/85	31	0	0	0	3	
19	18	0	2	6/28/51	64	0	0	0	6	
20	19	0	2	11/29/94	21	0	0	0	9	
21	20	0	2	8/6/88	27	0	0	0	2	
22	21	0	3	3/22/95	21	0	0	0	4	
23	22	1	3	1/23/92	24	0	0	0	4	
24	23	0	4	1/10/75	43	0	0	0	1	
25	24	0	1	8/24/83	32	0	0	0	3	
26	25	0	1	2/8/89	27	0	0	0	3	
27	26	0	1	5/3/79	36	0	0	0	3	
28	27	1	3	2/28/82	32	0	0	0	3	

Model
Inspector



comparison

	before	after
age	int min=23 max=60 32% mis	int min=23 max=60 32% mis
name	str length 10 to 98 no mis	str length 10 to 98 no mis
sex	str cat 10% mis	str cat 10% mis

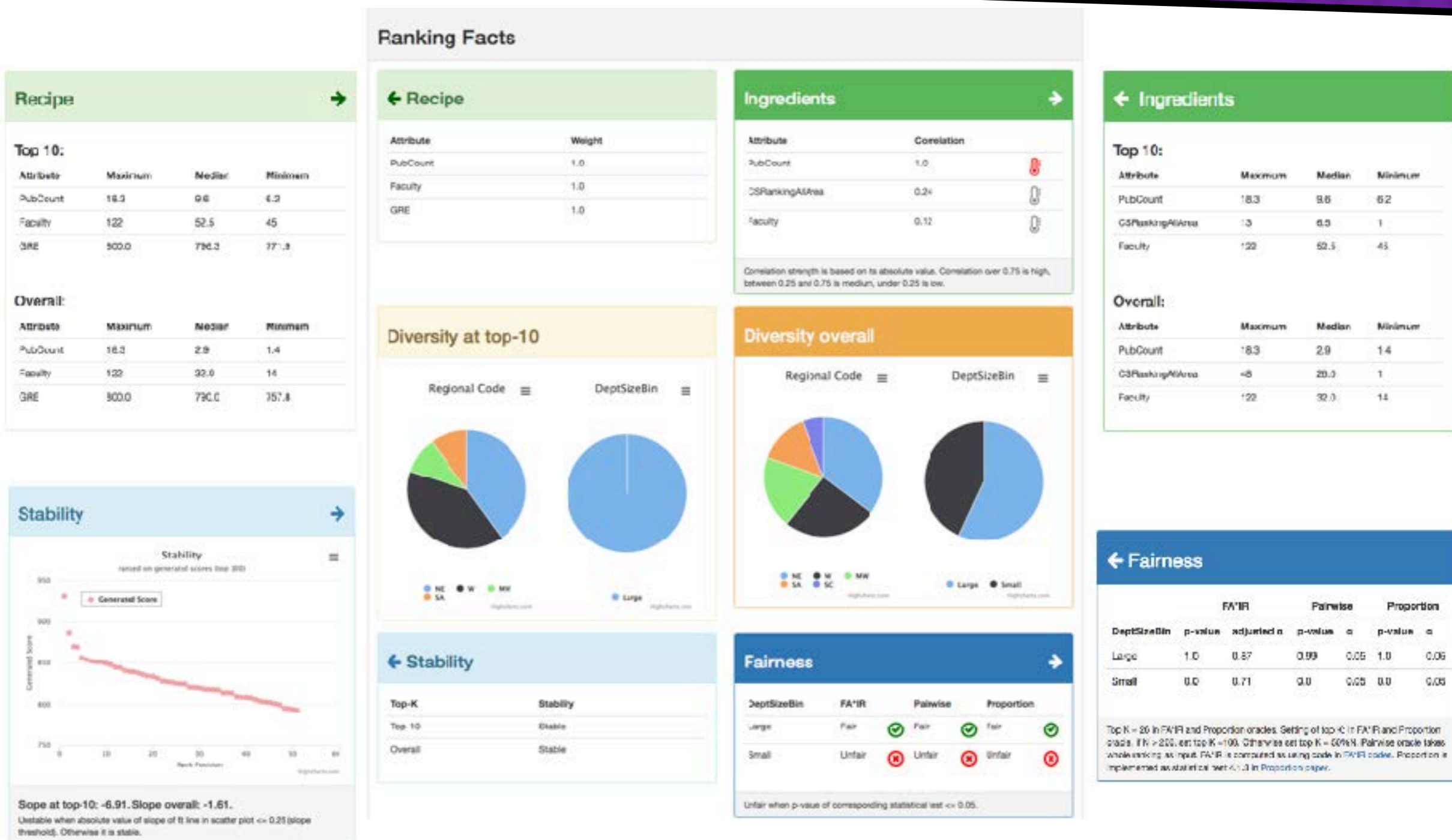
Point 4

actionable transparency requires
interpretability

explain assumptions and effects, not details of
operation

engage the public - technical and non-technical

“Nutritional labels” for data and models



http://demo.dataresponsibly.com/rankingfacts/nutrition_facts/

[K. Yang, J. Stoyanovich, A. Asudeh, B. Howe, HV Jagadish, G. Miklau; 2018]



Properties of a nutritional label

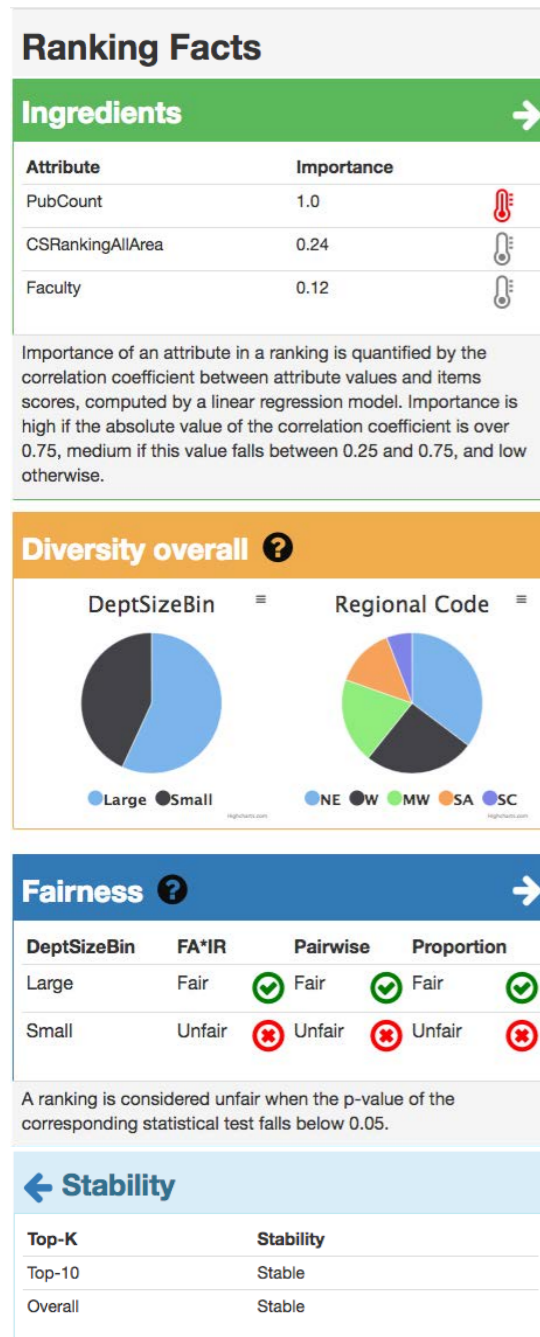
comprehensible: short, simple, clear

consultative: provide actionable info

comparable: implying a standard

concrete: helps determine a dataset's fitness for use for a given task

computable: produced as a “by-product” of computation - interpretability-by-design



Point 5

**transparency / interpretability by design,
not as an afterthought**

provision for transparency and interpretability at
every stage of the data lifecycle

useful internally during development, for
communication and coordination between
agencies, and for accountability to the public

*interpretability in the
eye of the
stakeholder*

Interpretability for different stakeholders



What are we explaining?

To **Whom** are we explaining?

Why are we explaining?

What are we explaining?

process (same for everyone? **why** is this the process?) vs. outcome

procedural justice aims to ensure that algorithms are perceived as fair and legitimate

data transparency is unique to algorithm-assisted decision-making, relates to the justification dimension of interpretability

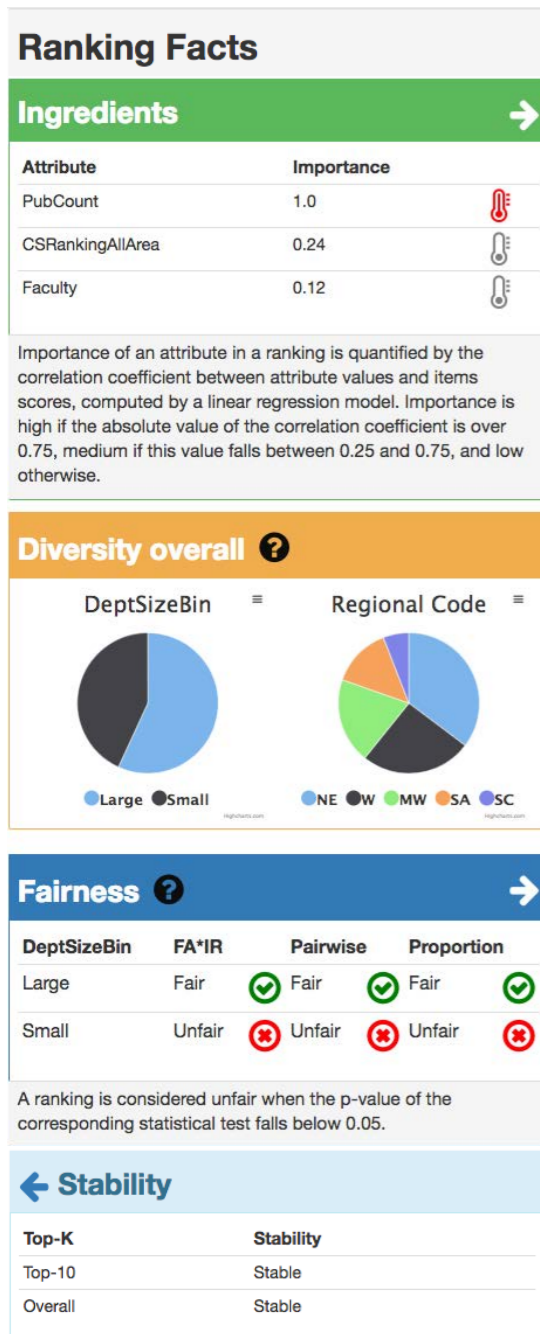
To whom are we explaining and why?

accounting for the needs of different stakeholders

social identity - people trust their in-group members more

moral cognition - is a decision or outcome morally right or wrong?

How do we know that we explained well?



nutritional labels! :)

... but do they work?

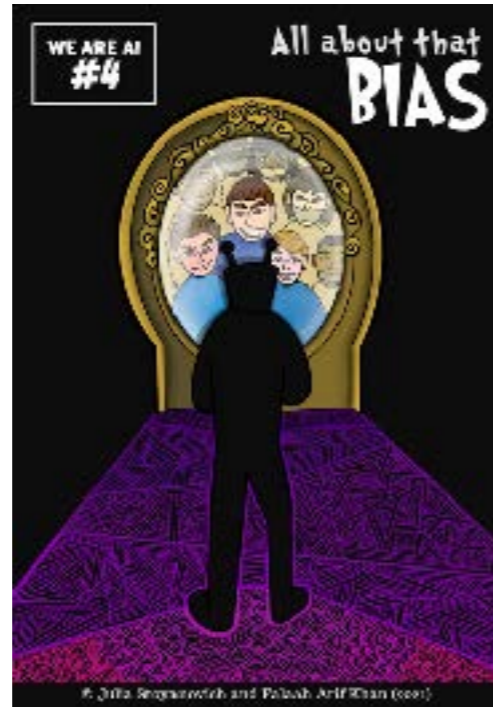
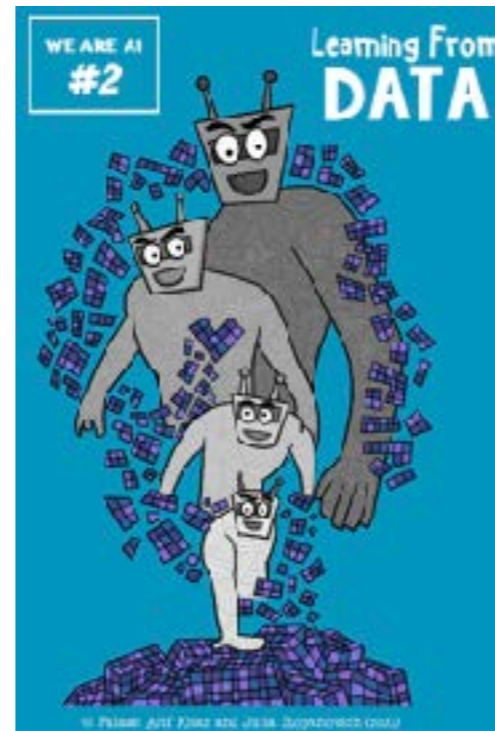
We are AI

taking control of technology
powered by NYU Center for Responsible AI

r/ai center
for
responsible
ai



AI comics for the general public



regulating automated
hiring systems

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.

Hiring ADS regulation

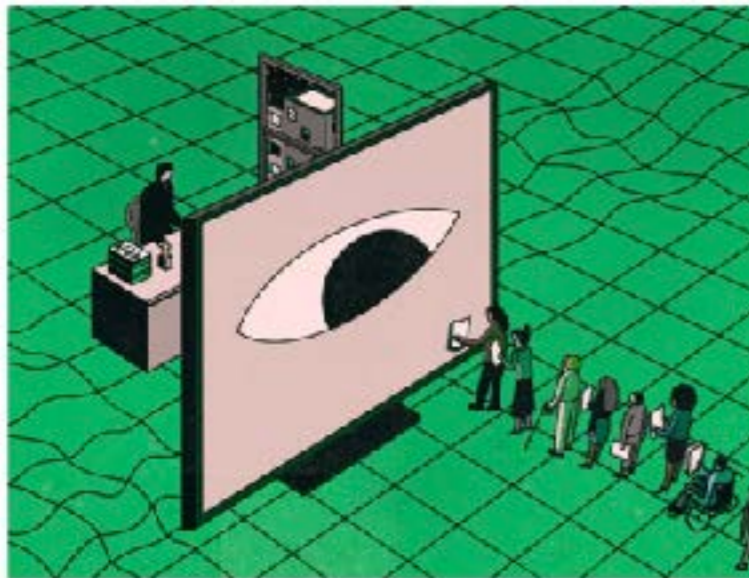
The New York Times

March 17, 2021

March 17, 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.



The measure must require companies to **publicly disclose what they find when they audit their tech for bias**. Despite pressure to limit its scope, the City Council must ensure that the bill would address discrimination in all forms — on the basis of not only race or gender but also disability, sexual orientation and other protected characteristics.

These audits should consider the circumstances of **people who are multiply marginalized** — for example, Black women, who may be discriminated against because they are both Black and women. Bias audits conducted by companies typically don't do this.

By Alexandra Reeve Givens, Hilke Schellmann and Julia Stoyanovich

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

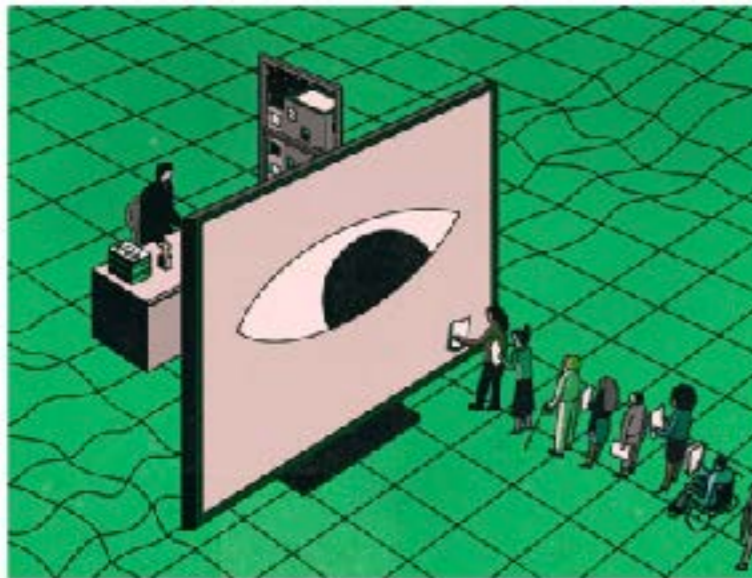
Hiring ADS regulation

The New York Times

March 17, 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.



March 17, 2021

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

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

By Alexandra Reeve Givens, Hilke Schellmann and Julia Stoyanovich

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

*but do the tools
work?*



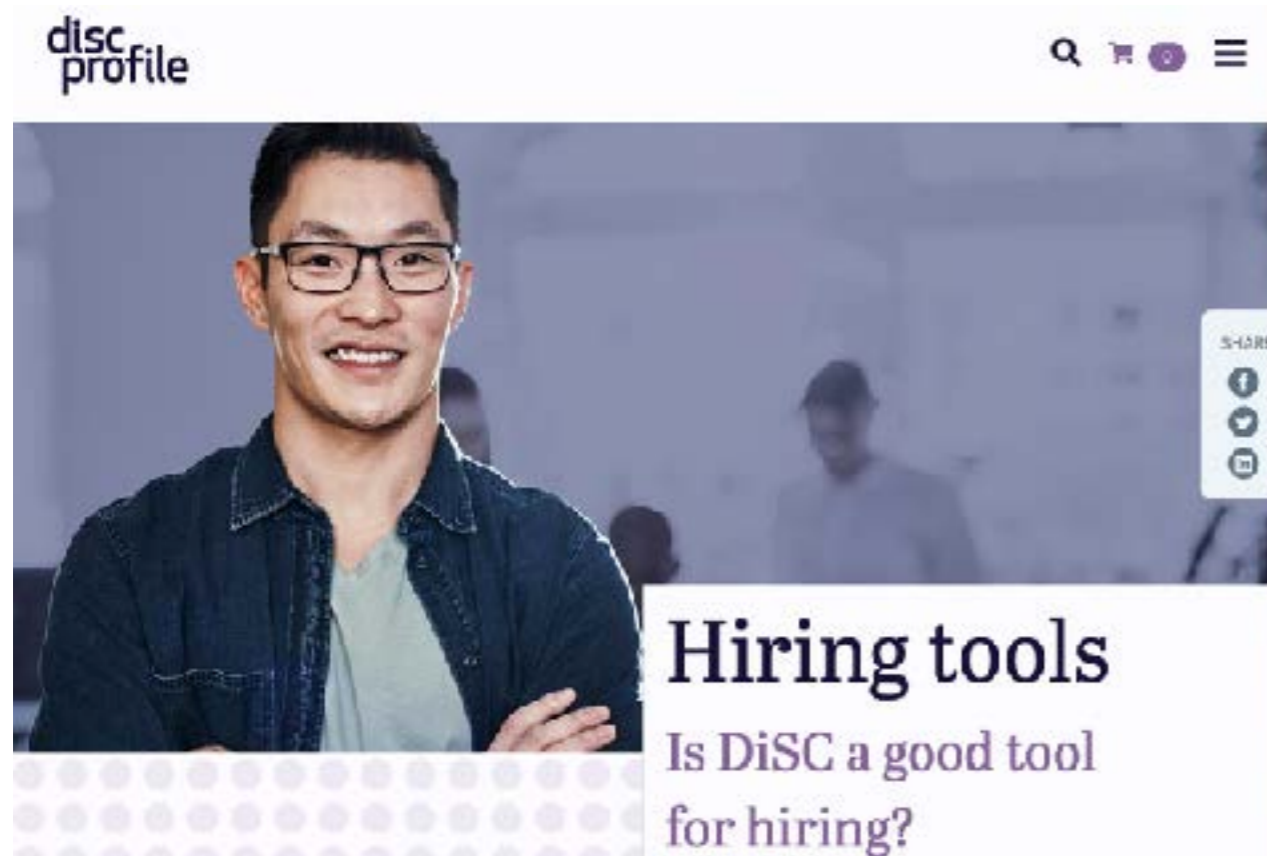
Personality prediction in hiring

DISC: Dominance (D), Influence (I), Steadiness (S), and Conscientiousness (C)

The Big Five: Openness (O), Conscientiousness (C), Extraversion (E), Agreeableness (A), and Neuroticism (N)



Personality prediction in hiring



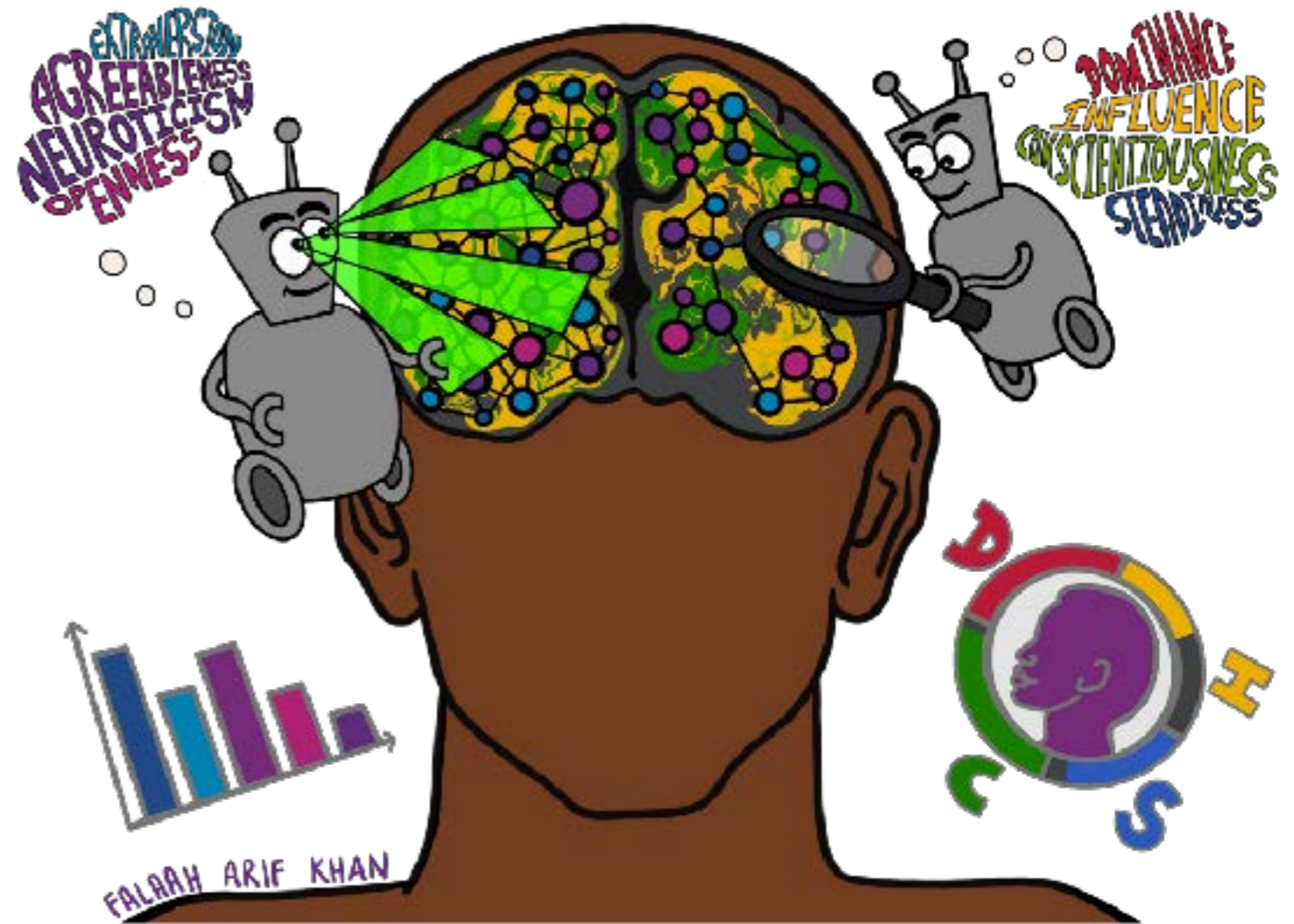
Although DiSC® profiles are often used as part of the hiring and onboarding process, they're **not recommended for pre-employment screening.**

DiSC does not measure specific skills, aptitudes, or other factors critical for a position; it describes one's natural work behavior patterns or styles to help improve productivity, teamwork, and communication.

Algorithmic personality tests

Input: resume or LinkedIn handle (both systems) or Twitter (Humantic AI)

Output: a personality profile + a job fit score (Crystal) or match score (Humantic AI)









Stability audit framework

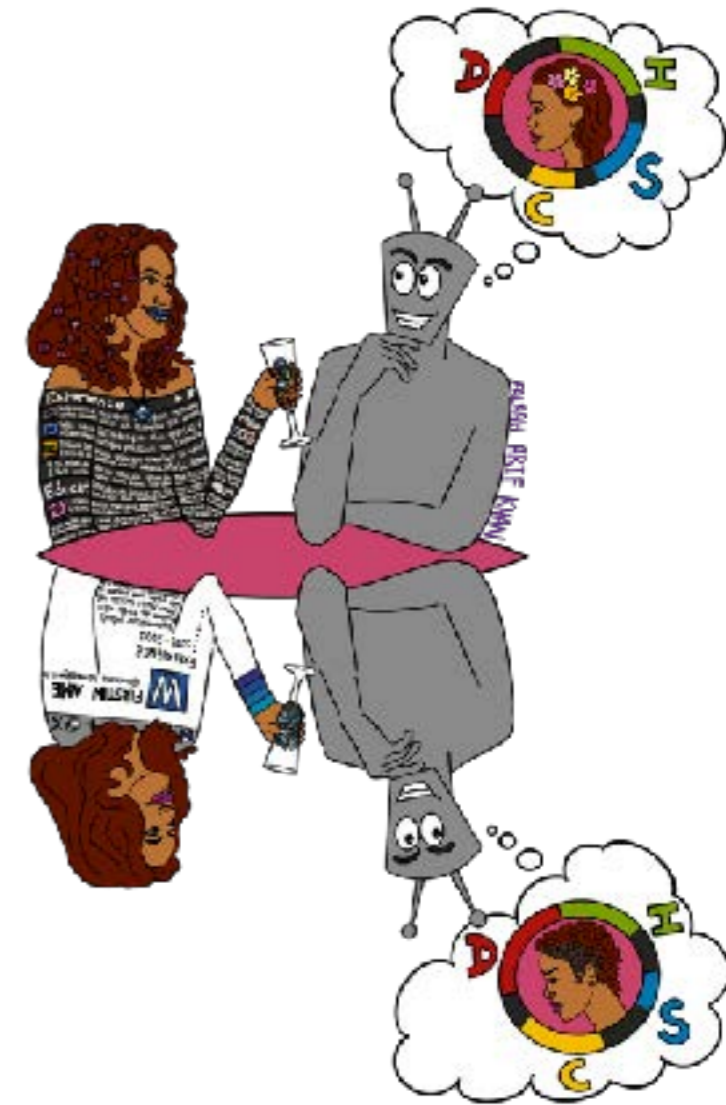
Data Mining and Knowledge Discovery (2022) 36:2153–2193
<https://doi.org/10.1007/s10618-022-00861-0>



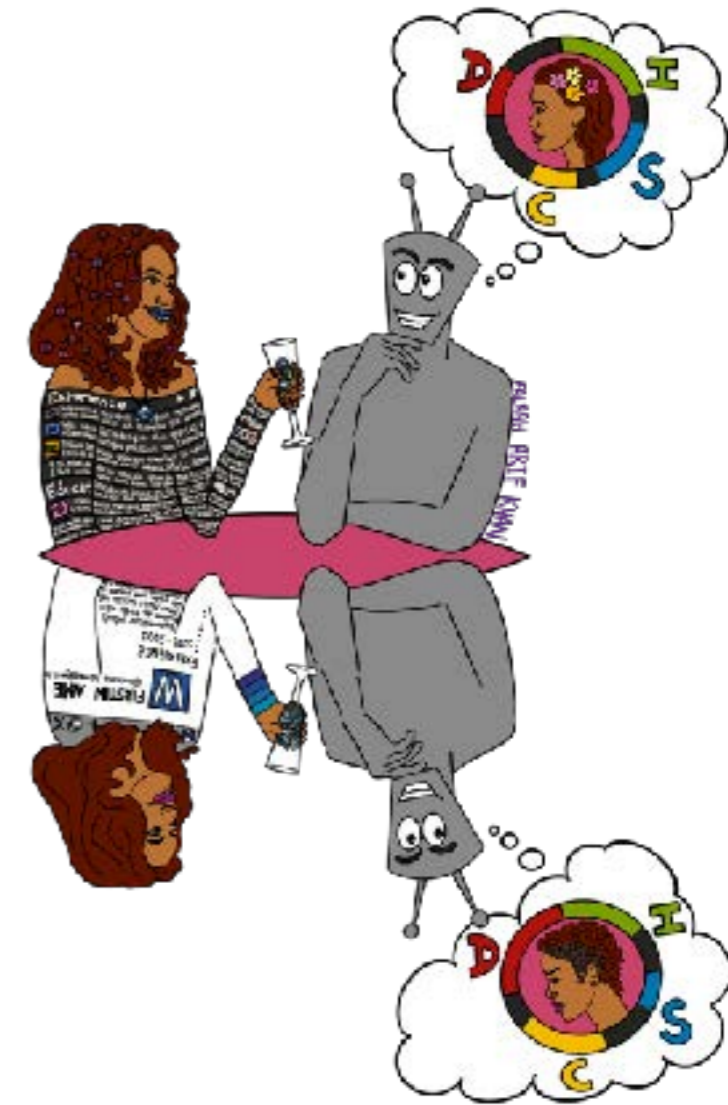
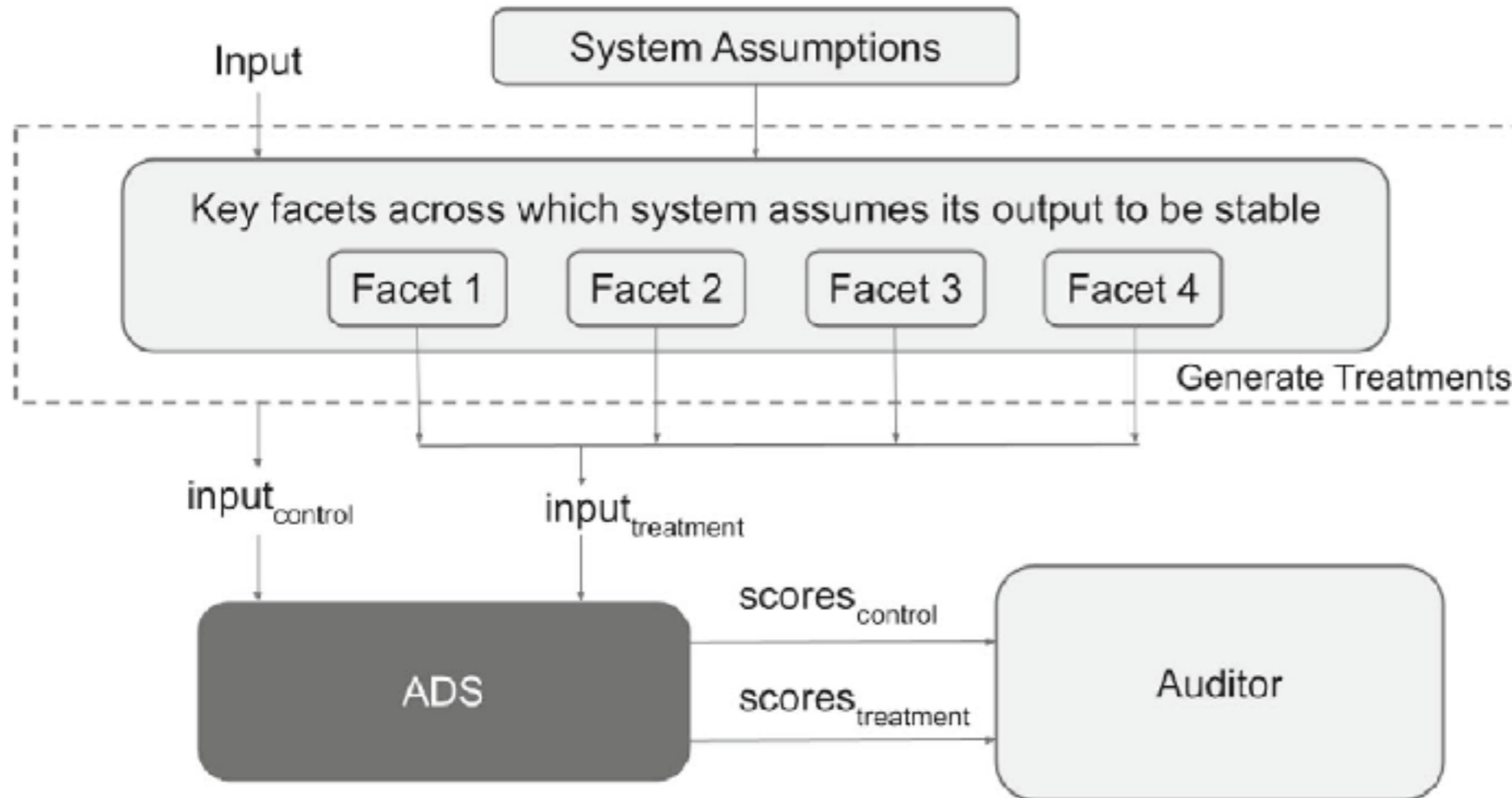
An external stability audit framework to test the validity of personality prediction in AI hiring

Alene K. Rhea^{1,2}  · Kelsey Markey^{1,2}  · Lauren D'Arinzo^{1,2,3}  ·
Hilke Schellmann⁴ · Mona Sloane²  · Paul Squires⁵ · Falaah Arif Khan^{1,2}  ·
Julia Stoyanovich^{1,2,6} 

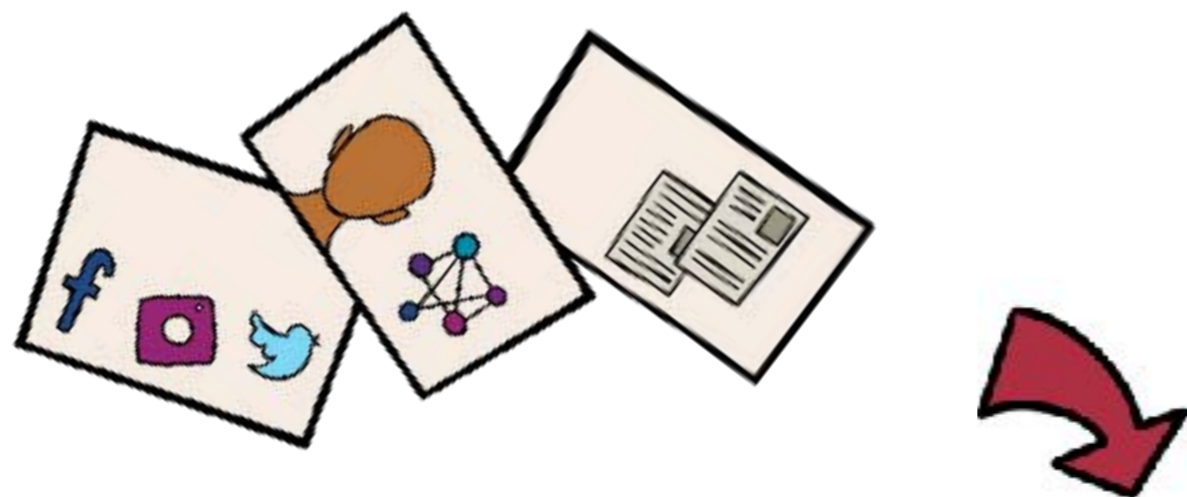
Received: 6 October 2021 / Accepted: 5 August 2022 / Published online: 17 September 2022
© The Author(s) 2022



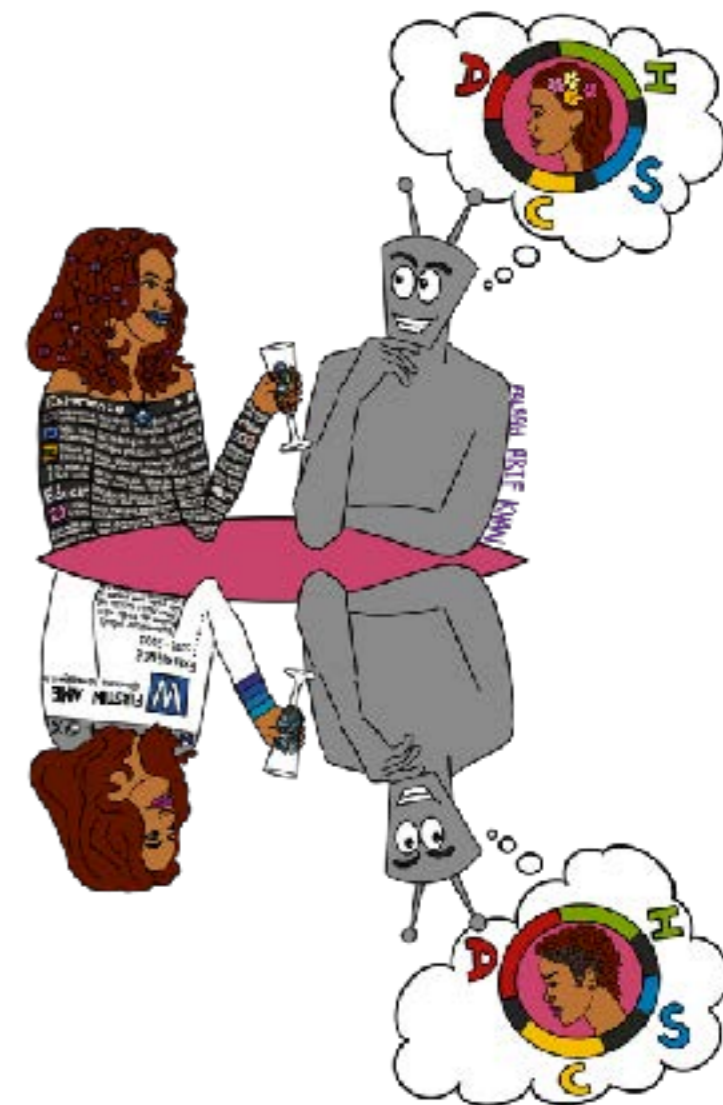
Stability audit framework



Stability audit framework



Facet	Crystal	Humantic
Resume file format	X	✓
LinkedIn URL in resume	?	X
Source context	X	X
Algorithm-time / immediate	✓	✓
Algorithm-time / 31 days	✓	X
Participant-time / LinkedIn	X	X
Participant-time / Twitter	N/A	✓



we need interpretability!

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.

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

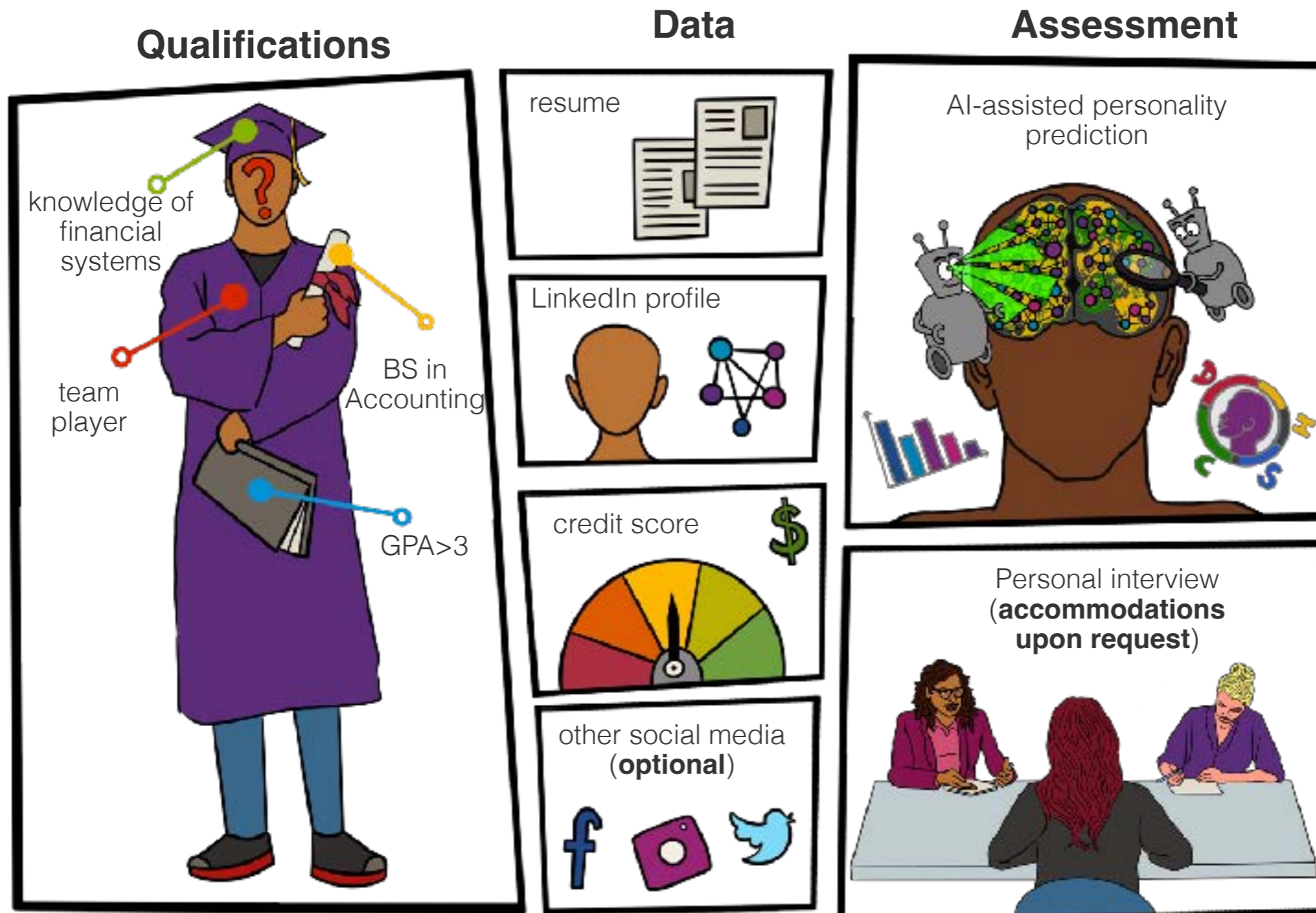
Qualifications: BS in accounting, GPA >3.0, Knowledge of financial and accounting systems and applications

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

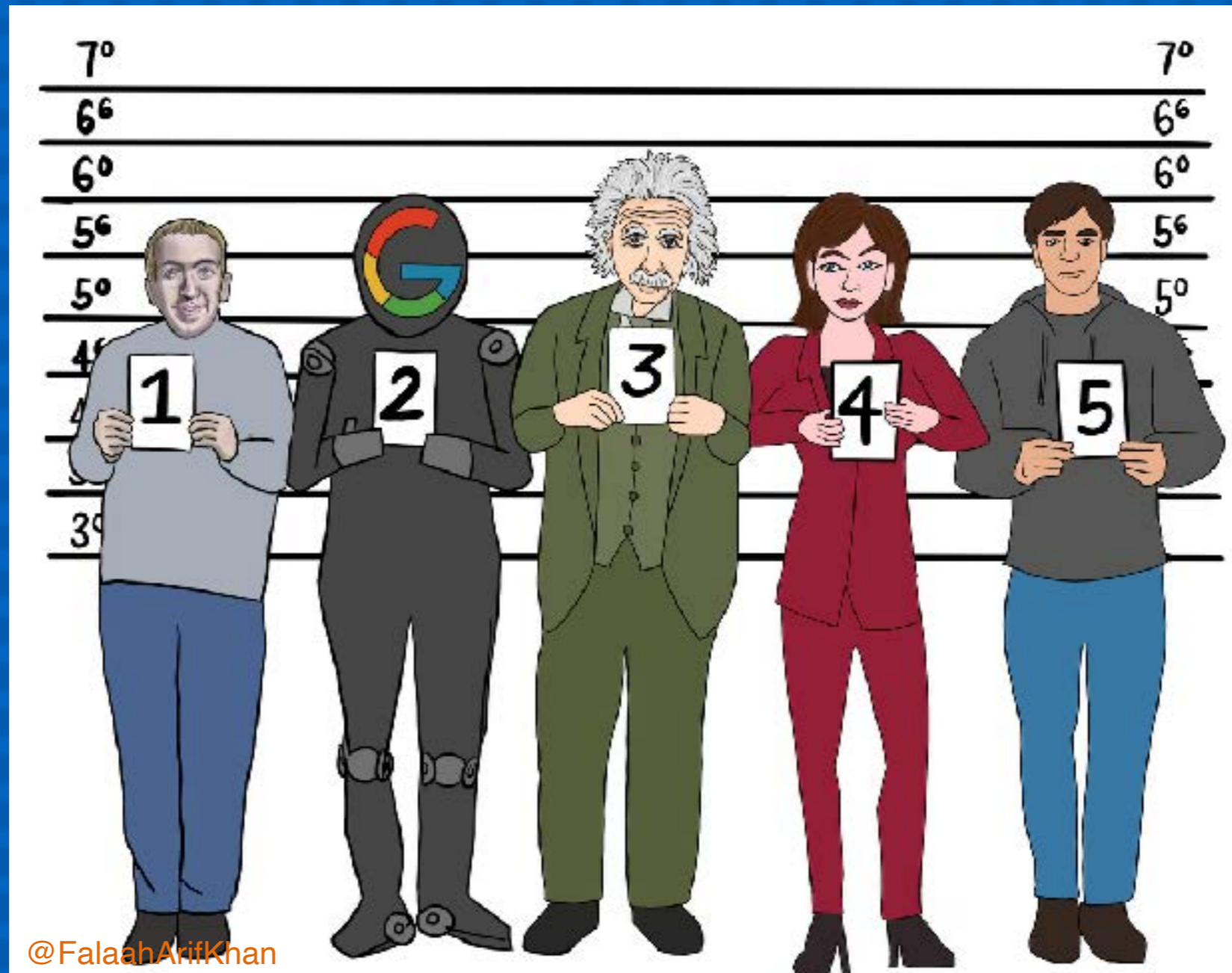
ALERT: Applicants for this position DO NOT have the option to selectively decline use of AI analysis for any of their personal data or to review and challenge the results of such analysis.

Anatomy of a job posting label



take-aways

We all are responsible

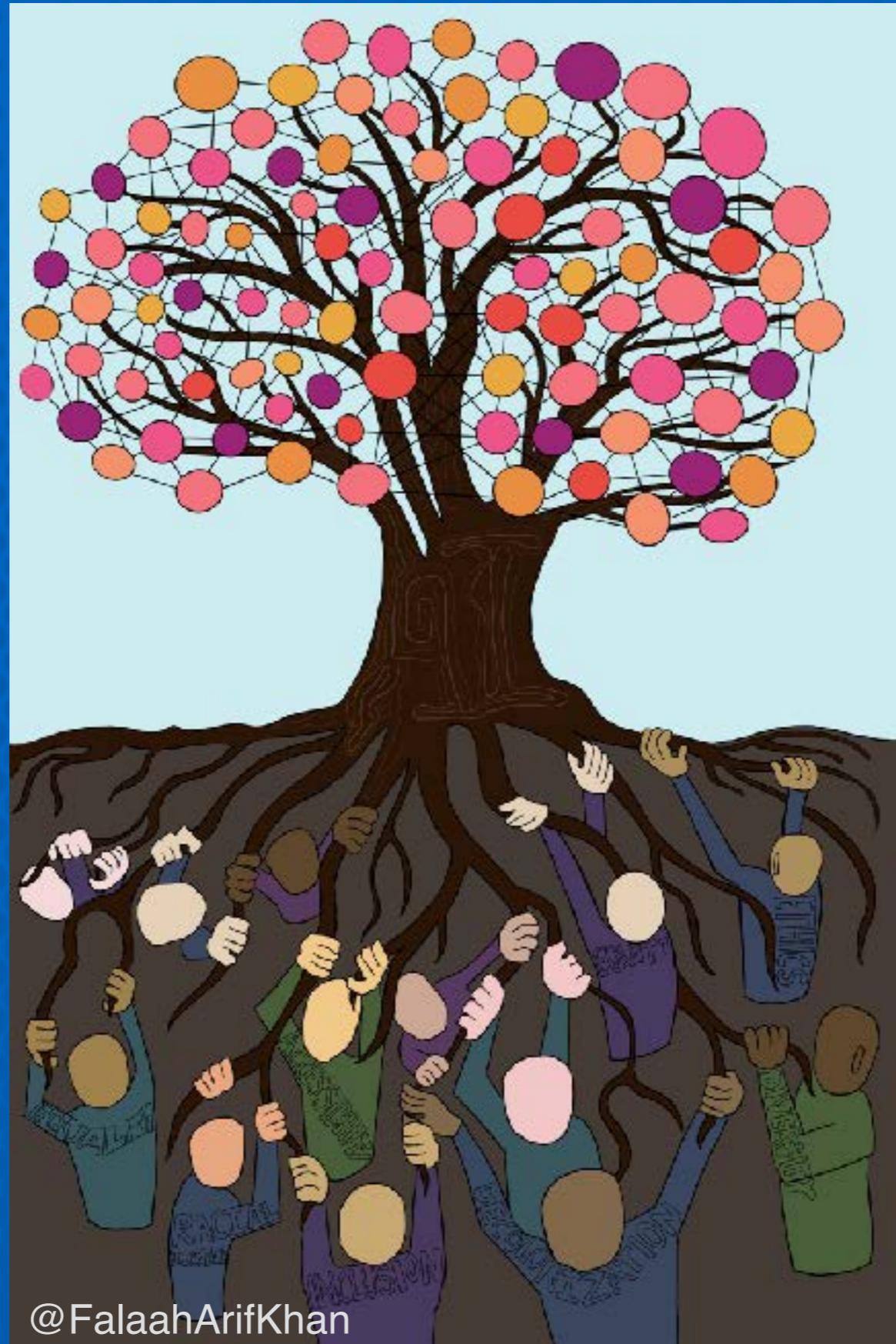


Searching for balance



@FalaahArifKhan

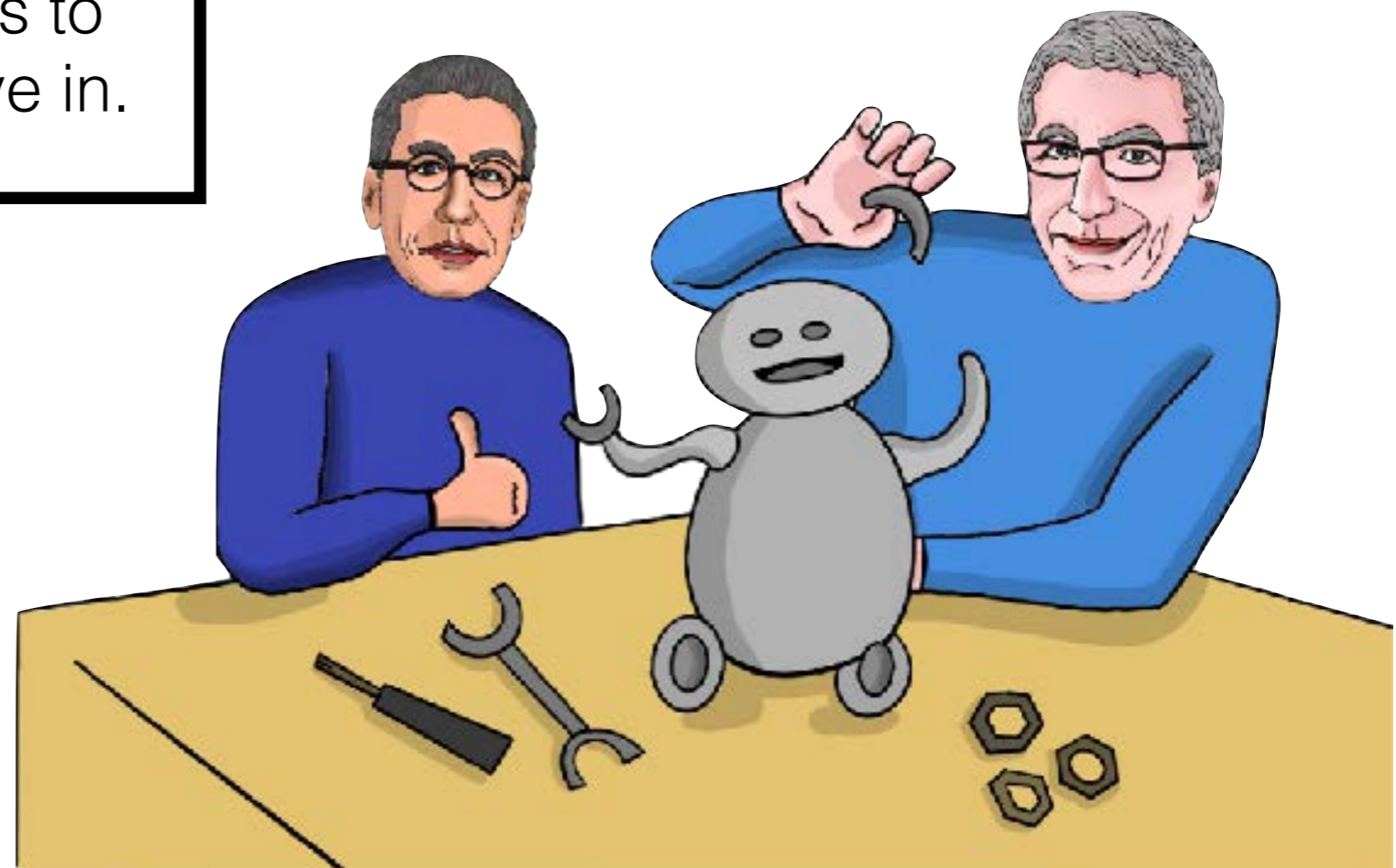
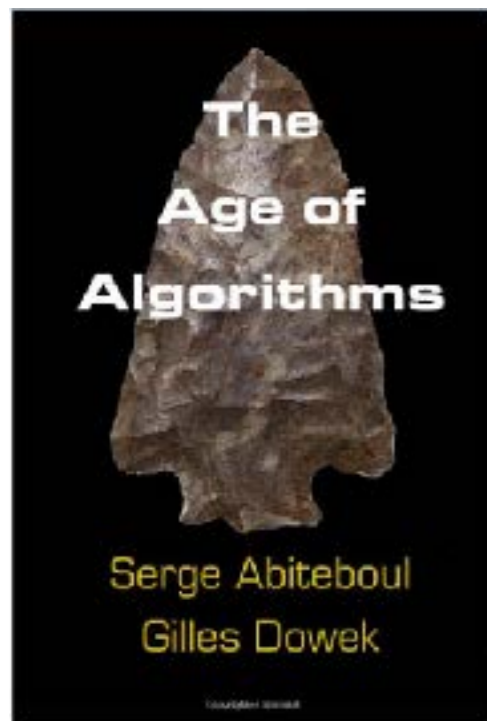
Tech rooted in people



@FalaahArifKhan

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.



Responsible Data Science

Thank you!



NYU

TANDON SCHOOL
OF ENGINEERING



NYU

Center for
Data Science

r/ai