Testimony of Julia Stoyanovich before the New York City Department of Consumer and Worker Protection regarding Local Law 144 of 2021 in Relation to Automated Employment Decision Tools

June 6, 2022

Dear Chair and members of the Department:

My name is Julia Stoyanovich. I hold a Ph.D. in Computer Science from Columbia University. I am an Associate Professor of Computer Science and Engineering at the Tandon School of Engineering, and an Associate Professor of Data Science at the Center for Data Science, and the founding Director of the Center for Responsible AI at New York University. In my research and public engagement activities, I focus on incorporating legal requirements and ethical norms, including fairness, accountability, transparency, and data protection, into data-driven algorithmic decision making.[1] I teach responsible data science courses to graduate and undergraduate students at NYU.[2] Most importantly, I am a devoted and proud New Yorker.

I actively participated in the deliberations leading up to the adoption of Local Law 144 of 2021[3] and have carried out several public engagement activities around this law when it was proposed [4]. Informed by my research and by opinions of members of the public, I have written extensively on the auditing and disclosure requirements of this Law, including an opinion article in the New York Times[5] and an article in the Wall Street Journal[6]. I have also been teaching members of the public about the impacts of AI and about its use in hiring, most recently by

[1] See https://dataresponsibly.github.io/ for information about this work, funded by the National Science Foundation through NSF Awards #1926250, 1934464, and 1922658.
[2] All course materials are publicly available at https://dataresponsibly.github.io/courses/
[3] Testimony of Julia Stoyanovich before New York City Council Committee on Technology regarding Int 1894-2020, November 12, 2020, available at https://dataresponsibly.github.io/documents/Stoyanovich_Int1894Testimony.pdf
[4] Public engagement showreel, Int 1894, NYU Center for Responsible AI, December 15, 2022 available at https://dataresponsibly.github.io/documents/Bill1894Showreel.pdf
[5] We need laws to take on racism and sexism in hiring technology, Alexandra Reeve Givens, Hilke Schellmann and Julia Stoyanovich, The New York Times, March 17, 2021, available at https://www.nytimes.com/2021/03/17/opinion/ai-employment-bias-nyc.html
[6] Hiring and AI: Let job candidates know why they were rejected, Julia Stoyanovich, The Wall Street Journal Reports: Leadership, September 22, 2021, available at https://www.wsj.com/articles/hiring-job-candidates-ai-11632244313

offering a free in-person course at the Queens Public Library called “We are AI”[7]. Course materials are available online[8].

In my statement today I would like to make three recommendations regarding the enforcement of Local Law 144 of 2021:

1. **Auditing**: The scope of auditing for bias should be expanded beyond disparate impact to include other dimensions of discrimination, and also contain information about a tool’s effectiveness - about whether a tool works. Audits should be based on a set of uniform publicly available criteria.
2. **Disclosure**: Information about job qualifications or characteristics for which the tool screens the job seeker should be disclosed to them in a manner that is comprehensible and actionable. Specifically, job seekers should see simple, standardized labels that show the factors that go into the AI’s decision both before they apply and after a decision on their application is made.
3. **An informed public**: To be truly effective, this law requires an informed public. I recommend that New York City invests resources into informing members of the public about data, algorithms, and automated decision making, using hiring ADS as a concrete and important example.

In what follows, I will give some background on automated hiring systems, and will then expand on each of my recommendations.

Automated hiring systems

Since the 1990s, and increasingly so in the last decade, commercial tools are being used by companies large and small to hire more efficiently: source and screen candidates faster and with less paperwork, and successfully select candidates who will perform well on the job. These tools are also meant to improve efficiency for the job applicants, matching them with relevant positions, allowing them to apply with a click of a button, and facilitating the interview process.

In their 2018 report, Bogen and Rieke[9] describe the hiring process from the point of view of an employer as a series of decisions that form a funnel: “Employers start by *sourcing* candidates,

---

[7] “We are AI” series by NYU Tandon Center for Responsible AI and Queens Public Library helps citizens take control of tech, March 14 2022, available at https://engineering.nyu.edu/news/we-are-ai-series-nyu-tandon-center-responsible-ai-queens-public-library

[8] “We are AI: Taking control of technology”, NYU Center for Responsible AI, available https://dataresponsibly.github.io/we-are-ai/

[9] Bogen and Rieke, “*Help Wanted: An Examination of Hiring Algorithms, Equity, and Bias”*, Upturn, (2018) https://www.upturn.org/static/reports/2018/hiring-algorithms/files/Upturn%20--%20Help%20Wanted%20-%20An%20Exploration%20of%20Hiring%20Algorithms,%20Equity%20and%20Bias.pdf

attracting potential candidates to apply for open positions through advertisements, job postings, and individual outreach. Next, during the *screening* stage, employers assess candidates—both before and after those candidates apply—by analyzing their experience, skills, and characteristics. Through *interviewing* applicants, employers continue their assessment in a more direct, individualized fashion. During the *selection* step, employers make final hiring and compensation determinations." Importantly, while a comprehensive survey of the space lacks, we have reason to believe that automated hiring tools are in broad use in all stages of the hiring process.

Despite their potential to improve efficiency for both employers and job applicants, hiring ADS are also raising concerns. I will recount two well-known examples here.

Sourcing: One of the earliest indications that there is cause for concern came in 2015, with the results of the AdFisher study out of Carnegie Mellon University[10] that was broadly circulated by the press[11]. Researchers ran an experiment, in which they created two sets of synthetic profiles of Web users who were the same in every respect — in terms of their demographics, stated interests, and browsing patterns — with a single exception: their stated gender, male or female. In one experiment, the AdFisher tool stimulated an interest in jobs in both groups, and showed that Google displays ads for a career coaching service for high-paying executive jobs far more frequently to the male group (1,852 times) than to the female group (318 times). This brings back memories of the time when it was legal to advertise jobs by gender in newspapers. This practice was outlawed in the US 1964, but it persists in the online ad environment.

Screening: In late 2018 it was reported that Amazon's AI resume screening tool, developed with the stated goal of increasing workforce diversity, in fact did the opposite thing: the system taught itself that male candidates were preferable to female candidates.[12] It penalized resumes that included the word "women's," as in "women's chess club captain," and downgraded graduates of two all-women's colleges. These results aligned with, and reinforced, a stark gender imbalance in the workforce at Amazon and other platforms, particularly when it comes to technical roles.

---

[10] Datta, Tschantz, Datta, "*Automated experiments on ad privacy settings",* Proceedings of Privacy Enhancing Technology (2015) https://content.sciendo.com/view/journals/popets/2015/1/article-p92.xml

[11] Gibbs, "*Women less likely to be shown ads for high-paid jobs on Google, study shows",* The Guardian (2015) https://www.theguardian.com/technology/2015/jul/08/women-less-likely-ads-high-paid-jobs-google-study

[12] Dastin, "*Amazon scraps secret AI recruiting tool that showed bias against women*", Reuters (2018) https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

Numerous other cases of discrimination based on gender, race, and disability status during screening, interviewing, and selection stages have been documented in recent reports[13][14]. These and other examples show that, if left unchecked, automated hiring tools will replicate, amplify, and normalize results of historical discrimination.

Recommendation 1: Expanding the scope of auditing

Bias audits should take a broader view, going beyond disparate impact when considering fairness of outcomes. Others surely spoke to this point, and I will not dwell on it here. Instead, I will focus on another important dimension of due process that is closely linked to discrimination — substantiating the use of particular features in decision-making.

Regarding the use of predictive analytics to screen candidates, Jenny Yang states: "Algorithmic screens do not fit neatly within our existing laws because algorithmic models aim to identify statistical relationships among variables in the data whether or not they are understood or job related.[...] Although algorithms can uncover job-related characteristics with strong predictive power, they can also identify correlations arising from statistical noise or undetected bias in the training data. Many of these models do not attempt to establish cause-and-effect relationships, creating a risk that employers may hire based on arbitrary and potentially biased correlations."[15]

In other words, identifying what features are impacting a decision is important, but it is insufficient to alleviate due process and discrimination concerns. I recommend that an audit of an automated hiring tool should also include information about the job relevance of these features.

A subtle but important point is that even features that can legitimately be used for hiring may capture information differently for different population groups. For example, it has been documented that the mean score of the math section of the SAT (Scholastic Assessment Test) differs across racial groups, as does the shape of the score distribution.[16] These disparities are

---

[13] Emerging Technology from the arXiv, "*Racism is Poisoning Online Ad Delivery, Says Harvard Professor*", MIT Technology Review (2013)
https://www.technologyreview.com/s/510646/racism-is-poisoning-online-ad-delivery-says-harvard-professor/

[14] Stains, *"Are Workplace Personality Tests Fair?*", Wall Street Journal (2014)
http://www.wsj.com/articles/are-workplace-personality-tests-fair-1412044257

[15] Yang, "*Ensuring a Future that Advances Equity in Algorithmic Employment Decisions*", Urban Institute (2020)
https://www.urban.org/research/publication/ensuring-future-advances-equity-algorithmic-employment-decisions

[16] Reeves and Halikias "*Race gaps in SAT scores highlight inequality and hinder upward mobility*", Brookings (2017)
https://www.brookings.edu/research/race-gaps-in-sat-scores-highlight-inequality-and-hinder-upward-mobility

often attributed to racial and class inequalities encountered early in life, and are thought to present persistent obstacles to upward mobility and opportunity.

Some automated hiring tools used today claim to predict job performance by analyzing an interview video for body language and speech patterns. Arvind Narayanan refers to tools of this kind as "fundamentally dubious" and places them in the category of AI snake oil.[17] The premise of such tools, that (a) it is possible to predict social outcomes based on a person's appearance or demeanor and (b) it is ethically defensible to try, reeks of scientific racism and is at best an elaborate random number generator.

The AI snake oil example brings up a related point: that an audit should also evaluate the effectiveness of the tool. Does the tool work? Is it able to identify promising job candidates better than a random coin flip? What were the specific criteria for the evaluation, and what evaluation methodology was used? Was the tool's performance evaluated on a population with demographic and other characteristics that are similar to the New York City population on which it will be used? Without information about the statistical properties of the population on which the tool was trained (in the case of machine learning) and validated, we cannot know whether the tool will have similar performance when deployed.[18]

In my own work, I recently evaluated the validity of two algorithmic personality tests that are used by employers for pre-employment assessment[19]. This work was done by a large interdisciplinary team that included several data scientists, a sociologist, an industrial-organizational (I-O) psychologist, and an investigative journalist. My colleagues and I developed a methodology for an external audit of stability of algorithmic personality tests, and used it to audit two systems, Humantic AI and Crystal. Importantly, rather than challenging or affirming the assumptions made in psychometric testing — that personality traits are meaningful and measurable constructs, and that they are indicative of future success on the job— we framed our methodology around testing the underlying assumptions made by the vendors of the algorithmic personality tests themselves.

In our audits of Humantic AI and Crystal, we found that both systems show substantial instability on key facets of measurement, and so cannot be considered valid testing instruments. For example, Crystal frequently computes different personality scores if the same resume is given in PDF vs. in raw text, while Humantic AI gives different personality scores on a LinkedIn profile vs. a resume of the same job seeker. This violated the assumption that the output of a personality test is stable across job-irrelevant input variations. Among other notable findings is

---

[17]Narayanan, "*How to recognize AI snakeoil*" (2019) https://www.cs.princeton.edu/~arvindn/talks/MIT-STS-AI-snakeoil.pdf

[18] Stoyanovich and Howe, "*Follow the data: Algorithmic transparency starts with data transparency*" (2019) https://ai.shorensteincenter.org/ideas/2018/11/26/follow-the-data-algorithmic-transparency-starts-with-data-transparency

[19] An external stability audit of framework to test the validity of personality prediction in AI hiring, Rhea et al., 2022, available at https://arxiv.org/abs/2201.09151

evidence of persistent — and often incorrect —data linkage by Humantic AI. A summary of our results are presented in **Table 1**.

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

Table 1: *Summary of stability results for Crystal and Humantic AI, with respect to facets of measurement: ✔ indicates sufficient rank-order stability in all traits, while ✗ indicates insufficient rank-order stability or significant locational instability in at least one trait, and **N/A** indicates the facet was not tested in our audit. Results are detailed in https://arxiv.org/abs/2201.09151.*

In summary, I recommend that the scope of auditing for bias should be expanded beyond disparate impact to include other dimensions of discrimination, and also contain information about a tool's effectiveness. To support compliance and enable a comparison between tools during procurement, these audits should be based on a set of uniform criteria. To enable public input and deliberation, these criteria should be made publicly available.

Recommendation 2: Explaining decisions to the job applicant

Information about job qualifications or characteristics that the tool uses for screening should be provided in a manner that allows the job applicant to understand, and, if necessary, correct and contest the information. As I argued in Recommendation 1, it is also important to disclose why these specific qualifications and characteristics are considered job relevant.

I recommend that explanations for job seekers are built around the popular nutritional label metaphor, drawing an analogy to the food industry, where simple, standardized labels convey information about the ingredients and production processes.[20]

[20] Stoyanovich and Howe, "*Nutritional labels for data and models*", IEEE Data Engineering BUlletin 42(3): 13-23 (2019) http://sites.computer.org/debull/A19sept/p13.pdf

An applicant-facing nutritional label for an automated hiring system should be comprehensible: short, simple, and clear. It should be consultative, providing actionable information. Based on such information, a job applicant may, for example, take a certification exam to improve their chances of being hired for this or similar position in the future. Labels should also be comparable: allowing a job applicant to easily compare their standing across vendors and positions, and thus implying a standard.

Nutritional labels are a promising metaphor for other types of disclosure, and can be used to represent the process or the result of an automated hiring system for auditors, technologists, or employers.[21]

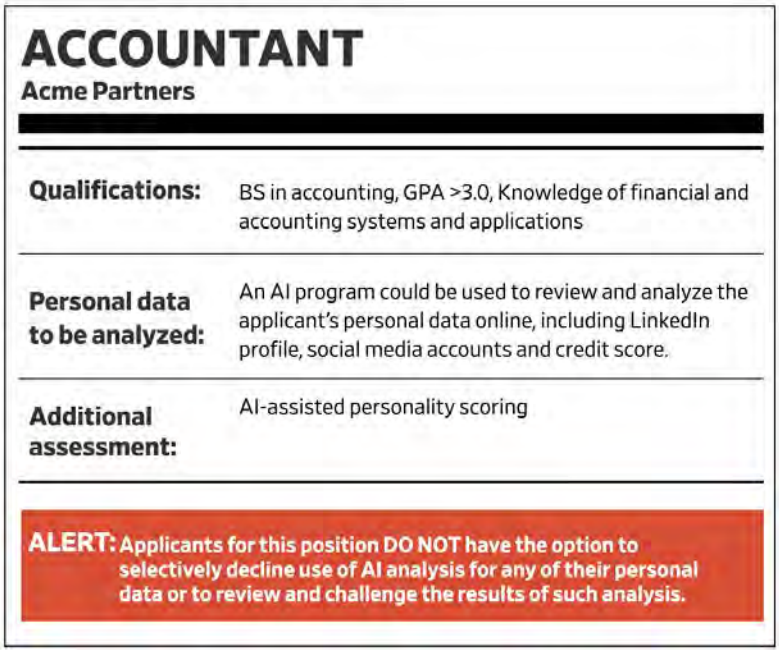


Figure 1: *A posting label is a short, simple, and clear summary of the screening process. This label is presented to a job seeker before they apply, supporting informed consent, allowing them to opt out of components of the process or to request accommodations.*

Figure 1 shows a posting label, a short and clear summary of the screening process. This label is presented to a job seeker before they apply, supporting informed consent, allowing them to opt out of components of the process or to request accommodations. Giving job seekers an opportunity to request accommodations is particularly important in light of the recent guidance

[21] Stoyanovich, Howe, Jagadish, *"Responsible Data Management"*, PVLDB 13(12): 3474-3489 (2020) https://dataresponsibly.github.io/documents/mirror.pdf

by the Equal Employment Opportunity Commission (EEOC) on the Americans with Disabilities Act and the use of AI to assess job applicants and employees [22].

If a job seeker applies for the job but isn't selected, then he or she would receive a"decision label" along with the decision. This label would show how the applicant's qualifications measured up to the job requirements; how the applicant compared with other job seekers; and how information about these qualifications was extracted.

Recommendation 3: Creating an informed public

My final recommendation will be brief. To be truly effective, this law requires an informed public. Individual job applicants should be able to understand and act on the information disclosed to them. In Recommendation 1, I spoke about the need to make auditing criteria for fairness and effectiveness publicly available. Empowering members of the public to weigh in on these standards will strengthen the accountability structures and help build public trust in the use of ADS in hiring and beyond. In Recommendation 2, I spoke about nutritional labels as a disclosure method. We should help job seekers, and the public at large, to understand and act upon information about data and ADS.

I recommend that New York City invests resources into informing members of the public about data, algorithms, and automated decision making, using hiring ADS as a concrete and important example. I already started this work, having developed "We are AI", a free public education course on AI and its impacts in society. This course is accompanied by a comic book series, available in English and Spanish.

Conclusion

In conclusion, I would like to quote from the recently released position statement by IEEE-USA, titled "Artificial Intelligence: Accelerating Inclusive Innovation by Building Trust".[23] IEEE is the largest professional organization of engineers in the world; I have the pleasure of serving on their AI/AS (Artificial Intelligence and Autonomous Systems) Policy Committee.

"We now stand at an important juncture that pertains less to what new levels of efficiency AI/AS can enable, and more to whether these technologies can become a force for good in ways that go beyond efficiency. We have a critical opportunity to use AI/AS to help make society more equitable, inclusive, and just; make government operations more transparent and

---

[22] The Americans with Disabilities Act and the use of software, algorithms, and AI to assess job applicants and employees, US Equal Employment Opportunity Commission, 2022, https://www.eeoc.gov/laws/guidance/americans-disabilities-act-and-use-software-algorithms-and-artificial-intelligence

[23] IEEE-USA, "*Artificial Intelligence: Accelerating Inclusive Innovation by Building Trust*" (2020) https://ieeeusa.org/wp-content/uploads/2020/10/AITrust0720.pdf

accountable; and encourage public participation and increase the public's trust in government. When used according to these objectives, AI/AS can help reaffirm our democratic values.

If, instead, we miss the opportunity to use these technologies to further human values and ensure trustworthiness, and uphold the status quo, we risk reinforcing disparities in access to goods and services, discouraging public participation in civic life, and eroding the public's trust in government. Put another way: Responsible development and use of AI/AS to further human values and ensure trustworthiness is the only kind that can lead to a sustainable ecosystem of innovation. It is the only kind that our society will tolerate."

We Need Laws to Take On Racism and Sexism in

Artificial intelligence used to evaluate job candidates must not become a tool

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 and Dr. Stoyanovich are professors at New York University focusing on artificial intelligence.

American democracy depends on everyone having equal access to work. But in reality, people of color, women, those with disabilities and other marginalized groups experience unemployment or underemployment at disproportionately high rates, especially amid the economic fallout of the Covid-19 pandemic. Now the use of artificial intelligence technology for hiring may exacerbate those problems and further bake bias into the hiring process.

At the moment, the New York City Council is debating a proposed new law that would regulate automated tools used to evaluate job candidates and employees. If done right, the law could make a real difference in the city and have wide influence nationally: In the absence of federal regulation, states and cities have used models from other localities to regulate emerging technologies.

Over the past few years, an increasing number of employers have started using artificial intelligence and other automated tools to speed up hiring, save money and screen job applicants without in-person interaction. These are all features that are increasingly attractive during the pandemic. These technologies include screeners that scan résumés for key words, games that claim to assess attributes such as generosity and appetite for risk, and even emotion analyzers that claim to read facial and vocal cues to predict if candidates will be engaged and team players.

In most cases, vendors train these tools to analyze workers who are deemed successful by their employer and to measure whether job applicants have similar traits. This approach can worsen underrepresentation and social divides if, for example, Latino men or Black women are inadequately represented in the pool of employees. In another case, a résumé-screening tool could identify Ivy League schools on successful employees' résumés and then downgrade résumés from historically Black or women's colleges.

In its current form, the council's bill would require vendors that sell automated assessment tools to audit them for bias and discrimination, checking whether, for example, a tool selects male candidates at a higher rate than female candidates. It would also require vendors to tell job applicants the characteristics the test claims to measure. This approach could be helpful: It would shed light on how job applicants are screened and force vendors to think critically about potential discriminatory effects. But for the law to have teeth, we recommend several important additional protections.

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.

The bill should also 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. Mandatory validity testing would also eliminate bad actors whose hiring tools do arbitrary things like assess job applicants' personalities differently based on subtle changes in the background of their video interviews.

In addition, the 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. The proposed legislation currently before the City Council would require companies to alert candidates within 30 days if they have been evaluated using A.I., but only after they have taken the test.

Finally, the bill must cover not only the sale of automated hiring tools in New York City but also their use. Without that stipulation, hiring-tool vendors could escape the obligations of this bill by simply locating sales outside the city. The council should close this loophole.

With this bill, the city has the chance to combat new forms of employment discrimination and get closer to the ideal of what America stands for: making access to opportunity more equitable for all. Unemployed New Yorkers are watching.

Alexandra Reeve Givens is the chief executive of the Center for Democracy & Technology. Hilke Schellmann is a repo
professor of journalism at New York University. Julia Stoyanovich is an assistant professor of computer science and e
Center for Responsible AI at New York University.

*The Times is committed to publishing a diversity of letters to the editor. We'd like to hear what you think about this o
email: letters@nytimes.com.*

Follow The New York Times Opinion section on Facebook, Twitter (@NYTopinion) and Instagram.

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

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

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.

How would this work? When people apply for a job, they will see a list of the hiring criteria, such as degree requirements, specific skills and the number of years of experience, so that they know precisely what a company is looking for. Then, if the applicant is rejected, the AI will present them with another list, showing where they didn't meet the criteria or compared unfavorably to other applicants—the reasoning behind the decision.

In other words, we should show people very clearly what factors are used to judge them, just as we show people the ingredients that go into their food.

We desperately need such a system. AI's widespread use in hiring far outpaces our collective ability to keep it in check—to understand, verify and oversee it. Is a résumé screener identifying promising candidates, or is it picking up irrelevant, or even discriminatory, patterns from historical data? Is a job seeker participating in a fair competition if he or she is unable to pass an online personality test, despite having other qualifications needed for the job?

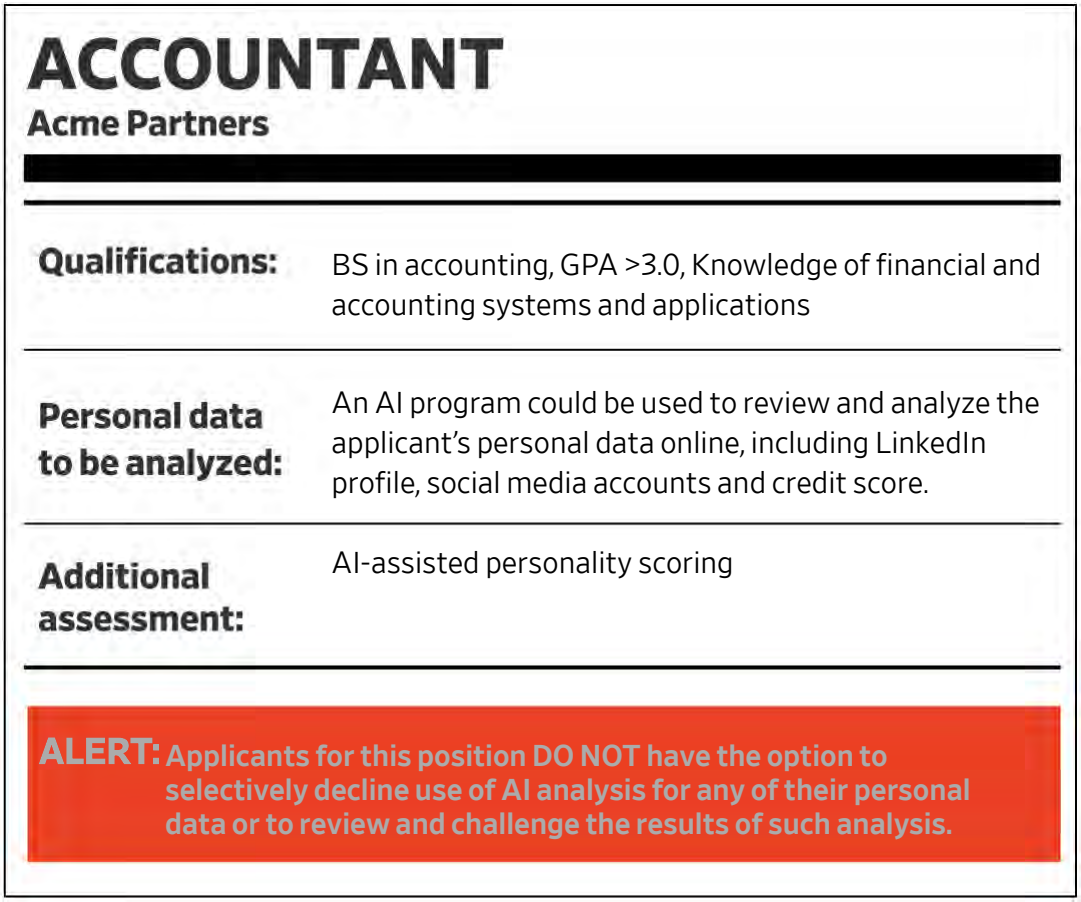
A two-tier system

The labels I propose would come in two parts. First, the "posting label," a short, simple and clear set of requirements that an AI screener will be looking for in applicants. For example, the posting label for an art-director position might list "B.S. in communications or similar," "two years of full-time experience" and "expert knowledge of Adobe Design Suite."

The posting label also would explain the assessment process. Will the AI consider only submitted résumés, or also use applicants' public LinkedIn profiles and Twitter feeds? What about their credit histories? Will a video interview be required? Which parts of the application are processed by a machine and which by a human?

Disclosing AI in Hiring

The author proposes a standardized method akin to nutritional labels that employers could use to inform job candidates when artificial intelligence programs will play a role in their evaluations. One example of what such a form might look like:



Source: Julia Stoyanovich

A posting label for an accountant position, for instance, may list résumé, LinkedIn, Twitter and credit scores as the sources of information, and it may state that personality scores will be used to assess the candidate, with preference given to candidates with higher S ("steady") and C ("conscientious") scores.

The label would also provide actionable information. It would state, for instance, that a job applicant is allowed to correct some data that the company uses to make decisions or contest the company's use of their personal information, such as their social-media feed.

In addition, applicants should be informed that they can request accommodations if they have reason to believe that a certain kind of assessment would discriminate against them. For example, scoring a video interview based in part on making eye contact with the camera would disadvantage people with limited vision or autism.

Critically, the posting label enables informed consent: Job seekers agree to the assessment procedure by submitting their applications, and they opt out by deciding not to apply if they object to the employer's data practices (e.g., using an applicant's credit score) or assessment methodology (e.g., constructing an estimation of an applicant's personality based on a résumé or performance in an online game).

If a job seeker applies for the job but isn't selected, then he or she would receive a "decision label" along with the rejection. This label would show how the applicant's qualifications measured up to the job requirements; how the applicant compared with other job seekers; and how information about these qualifications was extracted.

For example, a portion of the applicant's résumé may be highlighted to explain that he or she lacked sufficient experience for the position. Or a tweet may be highlighted to substantiate a low "conscientious" score on a personality test. This information would allow the applicant to accept or contest the hiring decision.

Explaining the choice

Having clear criteria for decisions not only helps applicants—it also gives employers vital information.

Many times, AI makes judgment calls that are opaque. Employers often don't know what data AI screeners are using, or how they analyze that data to make a final decision. The labels can show managers the factors that the AI is using to screen applicants—and let those managers decide if those factors need to be changed.

SHARE YOUR THOUGHTS

What do you think are the advantages and drawbacks of using AI in hiring? Join the conversation below.

For instance, does the AI need to be given more—or different—training data, covering different job roles and demographic groups, to avoid making biased and arbitrary decisions? Likewise, what is the predictive accuracy of the tool for different demographic groups? What features of past applicants' profiles led to a positive or a negative decision by the tool, and can job relevance of these features be substantiated?

One concern may be that these labels will motivate strategic manipulation or "gaming." However, there is already strategic manipulation happening: Career services routinely offer training and advice on how to make a résumé attractive to algorithmic tools. Greater transparency will help alleviate unproductive gaming and tilt the balance in favor of positive change, motivating individuals to actually improve their qualifications, rather than to make it seem like they are qualified.

Humans—and not AI—should ultimately make the final call on whom to hire. But, like it or not, many managers use AI systems at different stages of the hiring process—and that practice is only going to become more common. If managers are relying on AI, those tools should be as transparent as possible, and job seekers should have a say in how their data is used.

Ultimately, hiring is complex. It is a multistep process in which we trade off objective criteria, such as an applicant's degree requirements and measurable skills, against subjective factors such as how well they will fit into the team and pick up new skills.

We bring in AI to help alleviate some of this complexity. But we cannot forget that AI tools work to specification, and they do best when those specifications are clear. We can use AI effectively for parts of the hiring process—to identify clear requirements-based matches. But AI tools cannot exercise discretion or apply subjective judgment. My hope is that nutritional labels will help us come to a consensus on which decisions we should leave to an AI, and which we should make ourselves.

Dr. Stoyanovich is institute associate professor of computer science and engineering at the Tandon School of Engineering, associate professor of data science at the Center for Data Science and director of New York University's Center for Responsible AI. She can be reached at reports@wsj.com.

Appeared in the September 23, 2021, print edition as 'What's In AI's Hiring Black Box.'

WE ARE AI
#4

All about that BIAS

TERMS OF USE

Cite as:

Julia Stoyanovich and Falaah Arif Khan. "All about that Bias".
We are AI Comics, Vol 4 (2021)
https://dataresponsibly.github.io/we-are-ai/comics/vol4_en.pdf

Contact:

Please direct any queries about using elements from this comic to themachinelearnist@gmail.com and cc stoyanovich@nyu.edu

LET'S TALK ABOUT WHAT WE MEAN BY 'BIAS' IN AI, AND HOW IT ARISES.

WE SAY THAT AN AI IS BIASED IF ITS USE CAN LEAD TO SYSTEMATIC AND UNFAIR DISCRIMINATION AGAINST SOME INDIVIDUALS OR GROUPS IN FAVOR OF OTHERS.

BIAS CAN STEM FROM HARMFUL PATTERNS PICKED UP FROM THE DATA ITSELF,

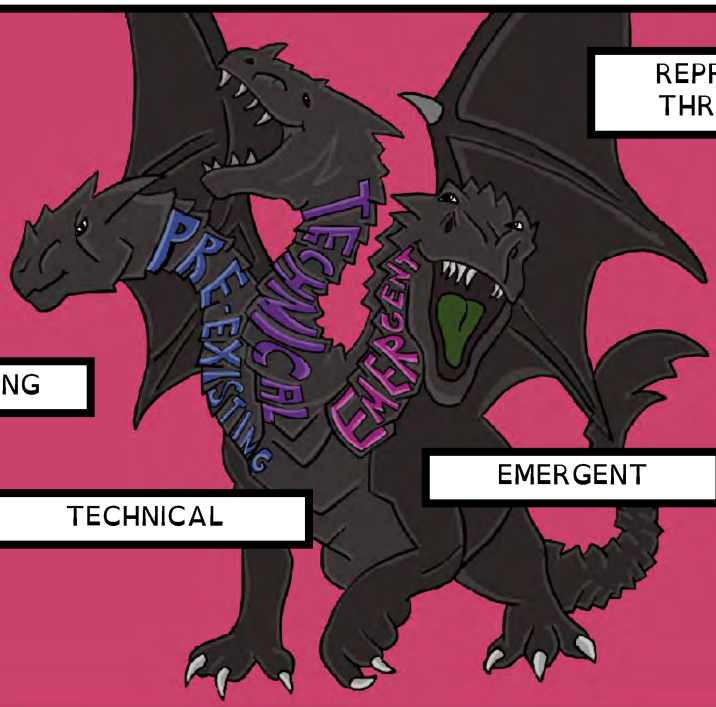
OR FROM HOW THE ALGORITHM IS DESIGNED,

OR FROM THE OBJECTIVES THAT WE SPECIFIED FOR IT,

OR FROM HOW WE USE IT.

IN THEIR SEMINAL 1996 PAPER [1], BATYA FRIEDMAN AND HELEN NISSENBAUM IDENTIFIED THREE TYPES OF BIAS THAT CAN ARISE IN COMPUTER SYSTEMS,

REPRESENTED HERE AS A THREE-HEADED DRAGON:



PRE-EXISTING

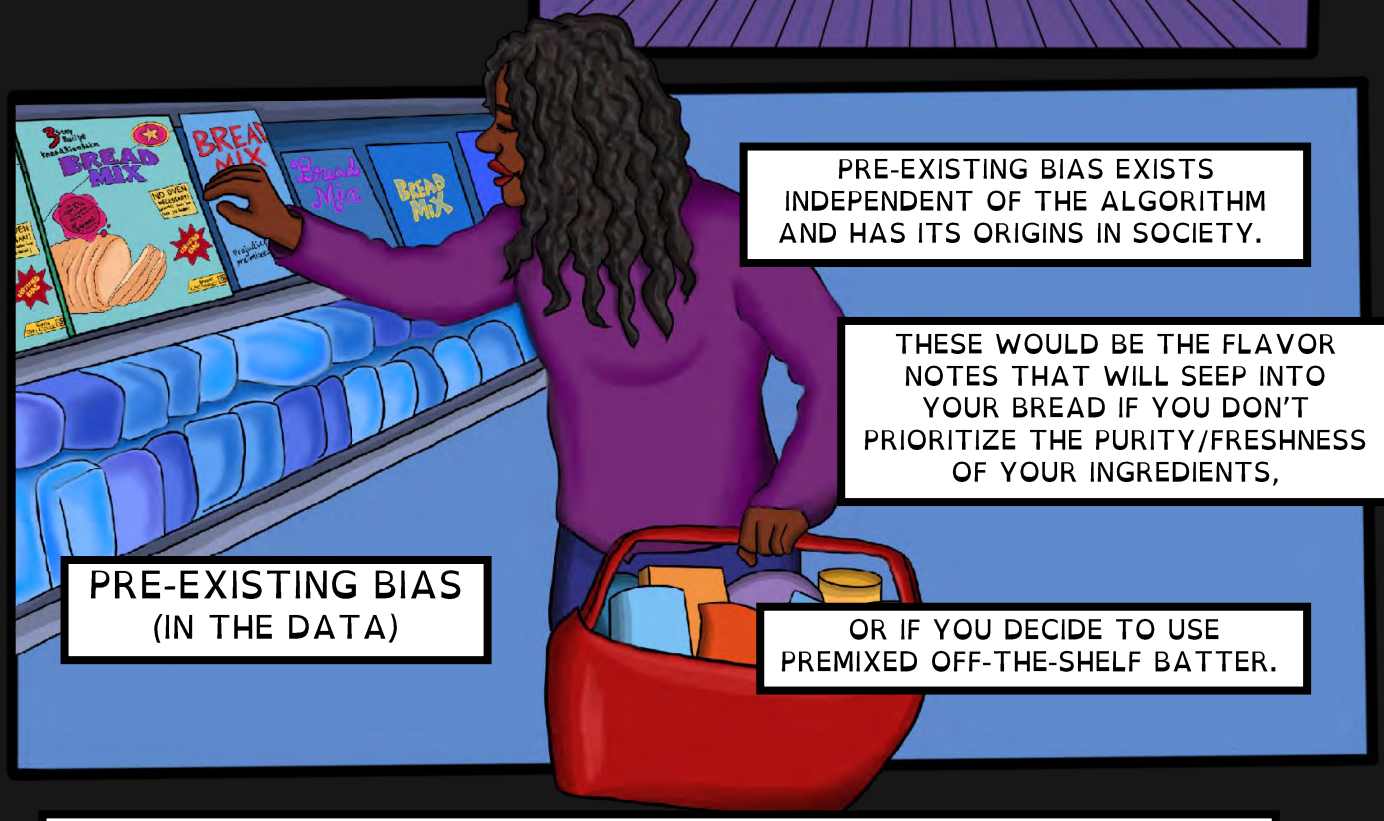
TECHNICAL

EMERGENT

[1] Batya Friedman and Helen Nissenbaum. (1996). Bias in computer systems.

RECALL THE BAKING METAPHOR WE USED TO UNDERSTAND DATA-DRIVEN ALGORITHMS IN VOLUME 1.

LET'S NOW USE THE SAME METAPHOR TO UNDERSTAND BIAS!



PRE-EXISTING BIAS EXISTS INDEPENDENT OF THE ALGORITHM AND HAS ITS ORIGINS IN SOCIETY.

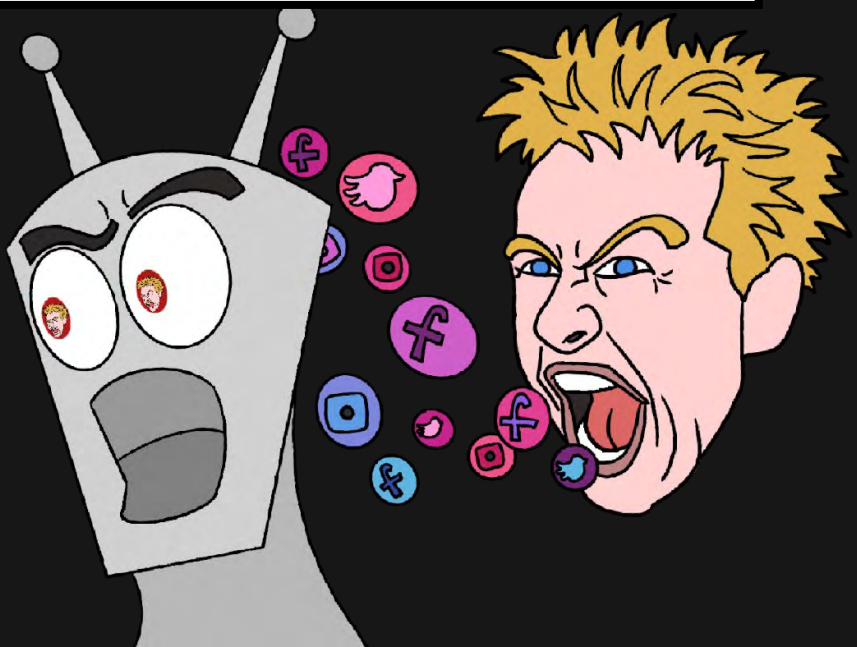
THESE WOULD BE THE FLAVOR NOTES THAT WILL SEEP INTO YOUR BREAD IF YOU DON'T PRIORITIZE THE PURITY/FRESHNESS OF YOUR INGREDIENTS,

PRE-EXISTING BIAS
(IN THE DATA)

OR IF YOU DECIDE TO USE PREMIXED OFF-THE-SHELF BATTER.

THESE BIASES EXIST IN SOCIETY AND COME 'PRE-BAKED' INTO THE ALGORITHM,

FROM THE UNDERLYING DISCRIMINATORY SYSTEM THAT THE DATA WAS COLLECTED FROM -

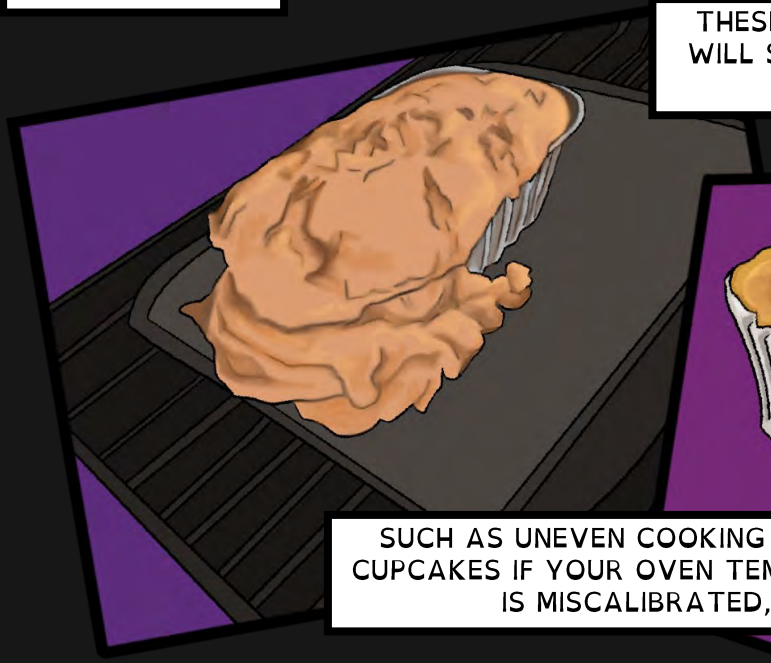


SUCH AS THE GENDER AND RACIAL STEREOTYPES THAT LANGUAGE MODELS PICK UP WHEN TRAINED ON DATA FROM SOCIAL MEDIA.

TECHNICAL BIAS

TECHNICAL BIAS IS INTRODUCED BY THE SYSTEM ITSELF - BECAUSE OF THE WAY IT IS DESIGNED OR OPERATES.

THESE WOULD BE THE IMPERFECTIONS THAT WILL SEEP INTO YOUR BREAD IF YOU USE THE WRONG EQUIPMENT -

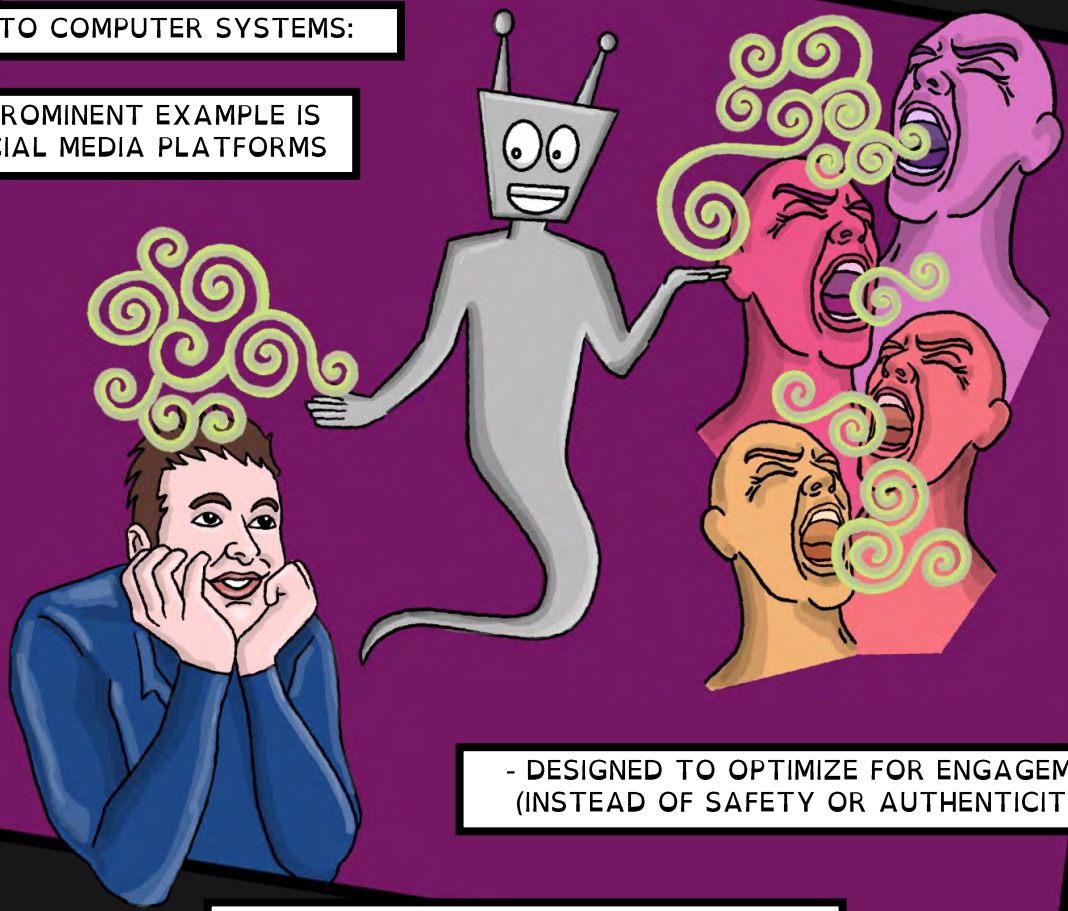


SUCH AS UNEVEN COOKING OF YOUR CUPCAKES IF YOUR OVEN TEMPERATURE IS MISCALIBRATED,

OR SPILLAGE OF BATTER IF YOUR BAKING EQUIPMENT IS OF THE WRONG SIZE.

BACK TO COMPUTER SYSTEMS:

A PROMINENT EXAMPLE IS SOCIAL MEDIA PLATFORMS



- DESIGNED TO OPTIMIZE FOR ENGAGEMENT (INSTEAD OF SAFETY OR AUTHENTICITY) -

THAT END UP PROMOTING POLARIZING ARTICLES AND FAKE NEWS.

EMERGENT BIAS
(DUE TO DECISIONS)

EMERGENT BIAS ARISES OVER TIME, BECAUSE THE DECISIONS MADE WITH THE HELP OF THE SYSTEM CHANGE THE WORLD,

WHICH IN TURN IMPACTS THE OPERATION OF THE SYSTEM GOING FORWARD.

THINK ABOUT BEHAVIORAL CHANGES THAT WILL EMERGE AS A RESULT OF YOUR BAKING -

WHAT IF YOU BECOME SUCH A MAESTRO AT BAKING THAT YOU INADVERTENTLY MAKE BREAD A STEADY PART OF YOUR DIET!

OR MAKE IT SO OFTEN, THAT YOU TURN EVERYONE AROUND YOU OFF THE THOUGHT OF EVER EATING ANOTHER SLICE!

OR THINK ABOUT HOW YOUR IDEA OF 'WHAT BREAD SHOULD TASTE LIKE' IS SHAPED BY THE POPULARITY OF PRODUCTS LIKE 'WONDER BREAD'.

IN THE SAME VEIN, THINK ABOUT HOW YOUR EXPOSURE TO NEWS - AND INFORMATION MORE BROADLY -

IS SHAPED BY ALGORITHMS THAT CURATE SOCIAL FEEDS WITH POPULAR AND 'TRENDING' POSTS.

TO MAKE OUR DISCUSSION CONCRETE, LET'S LOOK AT REAL-WORLD EXAMPLES OF ALGORITHMIC BIAS.

LET'S TAKE 'HIRING' AS A REPRESENTATIVE DOMAIN IN WHICH ALGORITHMS ARE INCREASINGLY BEING USED TO MAKE CRITICAL DECISIONS MORE 'EFFICIENTLY'.

ONE OF THE EARLIEST INDICATIONS THAT THERE IS CAUSE FOR CONCERN CAME IN 2015, WITH THE RESULTS OF THE ADFISHER STUDY OUT OF CARNEGIE MELLON UNIVERSITY. [2]

RESEARCHERS RAN AN EXPERIMENT, IN WHICH THEY CREATED TWO SETS OF SYNTHETIC PROFILES OF WEB USERS WHO WERE THE SAME IN EVERY RESPECT

— IN TERMS OF THEIR DEMOGRAPHICS, STATED INTERESTS, AND BROWSING PATTERNS —

WITH A SINGLE EXCEPTION: THEIR STATED GENDER, MALE OR FEMALE.

RESEARCHERS SHOWED THAT GOOGLE DISPLAYED ADS FOR A CAREER COACHING SERVICE FOR HIGH-PAYING EXECUTIVE JOBS FAR MORE FREQUENTLY TO THE MALE GROUP THAN TO THE FEMALE GROUP.

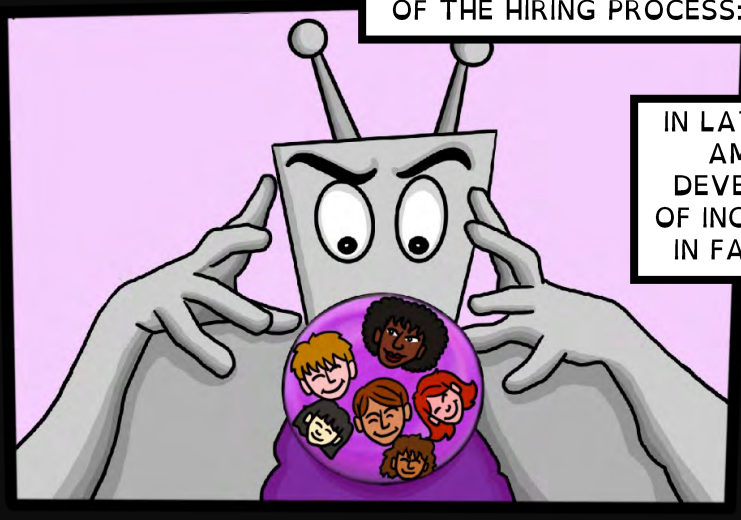
THIS BRINGS BACK MEMORIES OF THE TIME WHEN IT WAS LEGAL TO ADVERTISE JOBS BY GENDER IN NEWSPAPERS. THIS PRACTICE WAS OUTLAWED IN THE US IN 1964, BUT IT PERSISTS IN THE ONLINE AD ENVIRONMENT.

IT WAS LATER SHOWN THAT PART OF THE REASON THIS WAS HAPPENING IS THE MECHANICS OF THE ADVERTISEMENT TARGETING SYSTEM ITSELF, AS AN ARTIFACT OF THE BIDDING PROCESS.

THIS IS TECHNICAL BIAS IN ACTION!

[2] Women less likely to be shown ads for high-paid jobs on Google, study shows. Guardian (2015)

LET US MOVE FORWARD TO THE NEXT STAGE OF THE HIRING PROCESS: **RESUME SCREENING.**



IN LATE 2018 IT WAS REPORTED THAT AMAZON'S AI RECRUITING TOOL, DEVELOPED WITH THE STATED GOAL OF INCREASING WORKFORCE DIVERSITY, IN FACT DID THE OPPOSITE THING: [3]

THE SYSTEM TAUGHT ITSELF THAT MALE CANDIDATES WERE PREFERABLE TO FEMALE CANDIDATES.

IT PENALIZED RESUMES THAT INCLUDED THE WORD "WOMEN'S," AS IN "WOMEN'S CHESS CLUB CAPTAIN."

AND IT DOWNGRADED GRADUATES OF TWO ALL-WOMEN'S COLLEGES.

THE RESULTS ALIGNED WITH, AND REINFORCED, A STARK GENDER IMBALANCE IN THE WORKFORCE.

THIS IS EMERGENT BIAS IN ACTION -

A HIRING MANAGER TO WHOM AN AI TOOL REPEATEDLY SUGGEST THE SAME KIND OF JOB APPLICANT AS A GOOD FIT,

WILL OVERTIME COME TO BELIEVE THAT THIS IS WHAT A PROMISING EMPLOYEE LOOKS LIKE.

WE ARE ALSO SEEING PRE-EXISTING BIAS IN THIS EXAMPLE: THE AI TOOL WAS TRAINED ON HISTORICAL DATA ABOUT PAST EMPLOYEES, WHO WERE PREDOMINANTLY MALE

[3] Amazon scraps secret AI recruiting tool that showed bias against women. Reuters (2018)

HERE'S ANOTHER EXAMPLE, LATER YET IN THE HIRING PROCESS, PERHAPS DURING A POST-INTERVIEW BACKGROUND CHECK BY A POTENTIAL EMPLOYER -

LATANYA SWEENEY, A COMPUTER SCIENCE PROFESSOR ON THE FACULTY AT HARVARD,

SHOWED THAT GOOGLING FOR AFRICAN-AMERICAN SOUNDING NAMES IS MORE LIKELY TO TRIGGER ADS SUGGESTIVE OF A CRIMINAL RECORD THAN GOOGLING FOR WHITE-SOUNDING NAMES,

EVEN CONTROLLING FOR WHETHER AN INDIVIDUAL IN FACT HAS A CRIMINAL RECORD! [4]

Kristen

THIS IS PRE-EXISTING BIAS AT PLAY -

MANIFESTING LONG-STANDING RACIAL PREJUDICES OF SOCIETY.

Latanya

[4] Racism is Poisoning Online Ad Delivery, Says Harvard Professor. MIT Technology Review (2013)

THE CASES PRESENTED HERE HAVE ONE THING IN COMMON: THEY SHOW THAT AI CAN REINFORCE AND EXACERBATE UNLAWFUL DISCRIMINATION AGAINST MINORITY AND HISTORICALLY DISADVANTAGED GROUPS.

OFTEN THIS IS CALLED OUT AS "BIAS IN AI".

SO, WHY ARE SOPHISTICATED SYSTEMS THAT AIM TO MAKE HIRING MORE EFFICIENT FAILING AT THIS, AND ARGUABLY MAKING THINGS WORSE?

OF COURSE, THE ISSUES OF BIAS IN EMPLOYMENT ARE NOT NEW. THEY EXHIBITED THEMSELVES IN THE ANALOG ERA AS WELL.

FOR EXAMPLE, IN THEIR WELL-KNOWN 2004 STUDY, MARIANNE BERTRAND AND SENDHIL MULLAINATHAN SENT FICTITIOUS RESUMES TO HELP-WANTED ADS IN BOSTON AND CHICAGO NEWSPAPERS. [5]

Are EMILY and GREG more employable than LAKISHA and JAMAL?

TO MANIPULATE PERCEIVED RACE, THEY RANDOMLY ASSIGNED AFRICAN-AMERICAN- OR WHITE-SOUNDING NAMES TO RESUMES.

WHITE NAMES RECEIVE 50 PERCENT MORE CALLBACKS FOR INTERVIEWS.

THIS CASE SHOWS THAT BIAS CAN BE DUE TO HUMAN DECISIONS.

[5] Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. Marianne Bertrand & Sendhil Mullainathan (2003)

LET'S REVISIT PRE-EXISTING BIAS THAT OFTEN EXHIBITS ITSELF IN THE DATA.

DATA IS AN IMAGE OF THE **WORLD**, ITS **MIRROR REFLECTION**.

WHEN WE THINK ABOUT BIAS IN THE DATA, WE INTERROGATE THIS REFLECTION.

ONE INTERPRETATION OF "BIAS IN THE DATA" IS THAT THE REFLECTION IS DISTORTED -

WE MAY SYSTEMATICALLY OVER-REPRESENT OR UNDER-REPRESENT PARTICULAR PARTS OF THE WORLD IN THE DATA,

OR OTHERWISE **DISTORT** THE READINGS.

RECALL THE FAILURE OF AMAZON'S RECRUITING AI TO IMPROVE WORKFORCE DIVERSITY.

THIS TOOL WAS TRAINED USING HISTORICAL DATA: RESUMES OF PEOPLE WHO WERE HIRED IN THE PAST.

THAT TRAINING WAS SUBJECT TO PRE-EXISTING BIAS.

IN THAT DATA, THERE WAS AN UNDER-REPRESENTATION OF WOMEN IN THE WORKFORCE, AND IN TECHNICAL ROLES.

A MORE SUBTLE POINT IS ABOUT DISTORTIONS.

WHEN WE CONSIDER FEATURES, LIKE AN INDIVIDUAL'S SCORE ON A STANDARDIZED TEST, DO WE TAKE THESE AT FACE VALUE?

OR DO WE ACCOUNT FOR DIFFERENCES IN ACCESS TO EDUCATIONAL OPPORTUNITY,

LIKE GOING TO A BETTER SCHOOL, OR HAVING ACCESS TO PAID TUTORING?

ANOTHER INTERPRETATION OF "BIAS IN THE DATA" IS THAT EVEN IF WE WERE ABLE TO REFLECT THE WORLD PERFECTLY IN THE DATA,

IT WOULD STILL BE A REFLECTION OF THE WORLD SUCH AS IT IS,

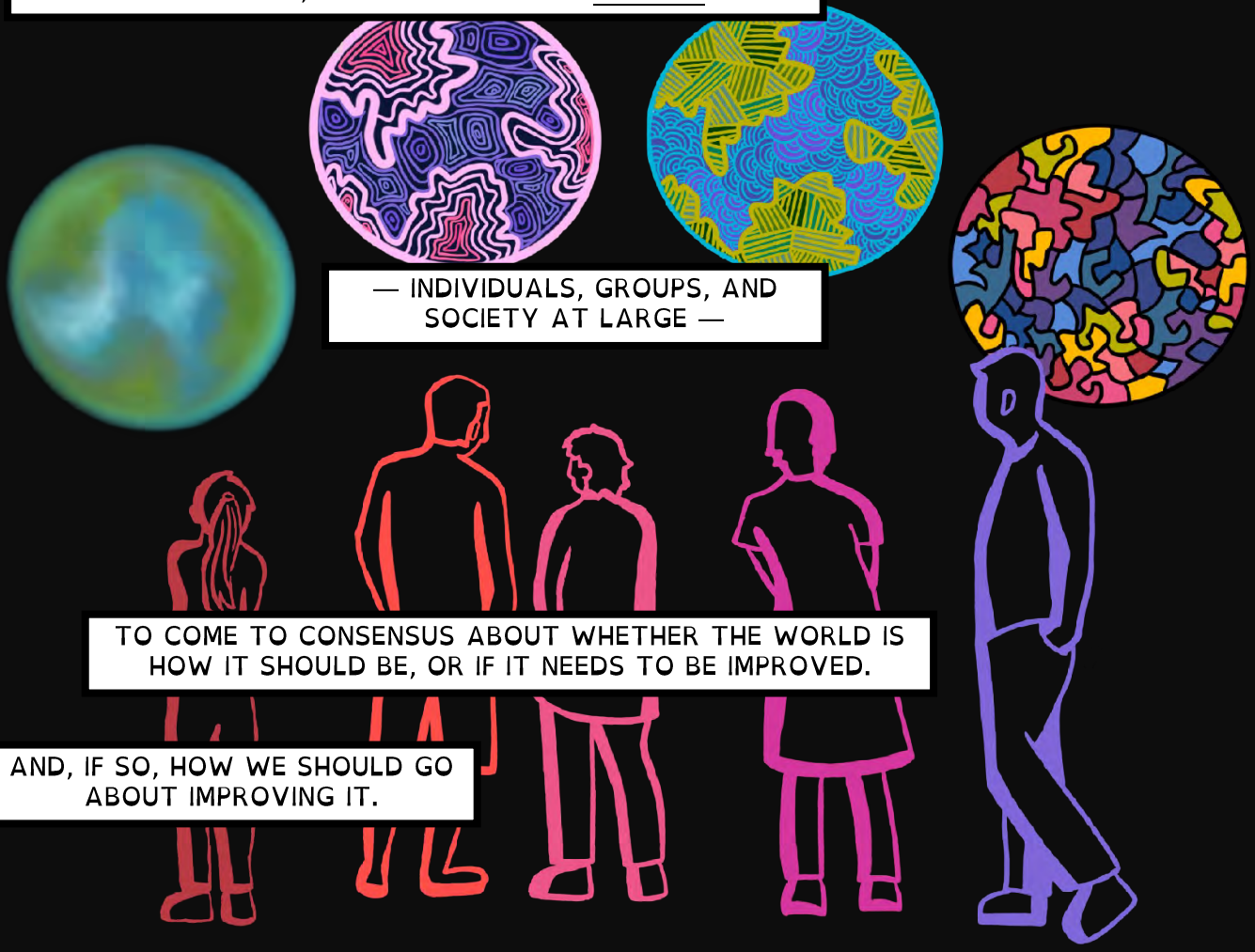
AND NOT NECESSARILY OF HOW IT COULD OR SHOULD BE.

IT IS IMPORTANT TO KEEP IN MIND THAT A REFLECTION CANNOT KNOW WHETHER IT IS DISTORTED.

DATA ALONE CANNOT TELL US WHETHER IT IS A DISTORTED REFLECTION OF A PERFECT WORLD, A PERFECT REFLECTION OF A DISTORTED WORLD,

OR IF THESE DISTORTIONS COMPOUND.

THE SECOND POINT IS THAT IT IS NOT UP TO DATA OR ALGORITHMS, BUT RATHER UP TO **PEOPLE**



— INDIVIDUALS, GROUPS, AND SOCIETY AT LARGE —

TO COME TO CONSENSUS ABOUT WHETHER THE WORLD IS HOW IT SHOULD BE, OR IF IT NEEDS TO BE IMPROVED.

AND, IF SO, HOW WE SHOULD GO ABOUT IMPROVING IT.

THE FINAL POINT HERE IS THAT CHANGING THE REFLECTION MAY NOT CHANGE THE WORLD.

IF THE REFLECTION ITSELF IS USED TO MAKE IMPORTANT DECISIONS -

FOR EXAMPLE, WHOM TO HIRE OR WHAT SALARY TO OFFER TO AN INDIVIDUAL BEING HIRED,

THEN COMPENSATING FOR THE DISTORTIONS IS WORTHWHILE.

BUT THE MIRROR METAPHOR ONLY TAKES US SO FAR.

WE HAVE TO WORK MUCH HARDER — USUALLY GOING FAR BEYOND TECHNOLOGICAL SOLUTIONS — TO MAKE LASTING CHANGE IN THE WORLD,

NOT MERELY BRUSH UP THE REFLECTION.

CIRCLING BACK NOW TO THE THREE-HEADED BIAS DRAGON.

WHEN SPEAKING ABOUT TACKLING BIAS IN AI, WE TEND TO FRAME THE PROBLEM AS FINDING A WAY TO SLAY THE BIAS-DRAGON.

PRE-EXISTING

TECHNICAL

EMERGENT

BUT THROUGH OUR DISCUSSION OF THE LINK BETWEEN HUMAN BIAS AND MACHINE BIAS,

WE FIND OURSELVES QUESTIONING THE VERY NATURE OF THIS TALE -

AT THE END OF THE DAY, MAYBE THE QUESTION ISN'T -

HOW TO SLAY THE DRAGON AND RESCUE THE PRINCESS?

THE QUESTION WE REALLY SHOULD BE ASKING OURSELVES IS -

WHAT DO WE DO ABOUT A SOCIETY THAT LOCKS UP PRINCESSES IN CASTLES, IN THE FIRST PLACE?

FIN.