The Elephant in AI

A Toolbox for Those in the AI Ecosystem to Identify and Mitigate Bias in Recruiting and Hiring Platforms

March 2021

University of Pennsylvania Law School
Policy Lab on AI and Implicit Bias

Cover by Bianca Nachmani

“Dr. Joy Buolomwini is a Rhodes Scholar, MIT researcher, poet, and scientist. Her work on coded bias sheds light on the threats of AI to human rights and reveals how facial recognition computer software works better when the person wears a white mask. Apart from the allusion to Frantz Fanon’s famous work “Black Skin, White Masks” on the construction of Black identity, the “White Masks” in our report cover borrows from the idea developed by Dr. Buolamwini. Her newly coined term “coded gaze” refers to the bias in coded algorithms. Her work at the MIT Media Lab’s “Gender Shades” Project uncovers racial and gender bias in AI systems and blows the whistle about the potential threats of unchecked AI. Using Dr. Buolamwini’s model, we too wanted to curate stories of a new generation of professionals experiencing bias via AI.”

— PROFESSOR RANGITA DE SILVA DE ALWIS

Students of the Spring 2021 Policy Lab

Mihir Awati	Otis Hill	Chandrasekhar Nukala
Kimerly Biesinger	Wooseok Ki	Yinran Pan
Amani Carter	Ashley Marcus	Sara Shayanian
Sonari-Nnamdi Chidi	Alex Mueller	Anna Sheu
Emily Friedman	Bianca Nachmani	Jessica Shieh
Germaine Grant	Govind Nagubandi	Ashley Suarez
Marleca Higgs	Maryam Nasir	

Teaching Assistants for AI & Bias Policy Lab:
Lindsay Holcomb, University of Pennsylvania Law School
Levi Marelus, University of Pennsylvania Law School

Report Design by Bianca Nachmani,
J.D. ‘21 with emphasis in law and design
With special help from Maryam Nasir

ABOUT US

The Univeristy of Pennsylvania Law School Policy Lab on AI and Implicit Bias incubates ideas for an intersectional approach to inclusive artificial intelligence.

Primarily through a series of multilateral conversations with international stakeholders, including leaders in technology, technologists, lawyers, researchers, and designers, we will seek to understand whether and how gender and intersectional bias, including implicit and unconscious biases are being baked into technological design and algorithms, and whether and how these biases are being reproduced in new technologies. Currently, there is gender and intersectional asymmetry in the AI workforce. Those designing, coding, engineering and programming AI technologies do not represent a diverse demographic. Our theoretical explorations included the human rights framework, gender equality theory, post- colonial theory, implicit bias, in group favoritism, and affinity bias to explore subtle barriers to equality that bleed into the design of AI technologies. The lab engaged in survey- based research and data collection on a new generation of algorithmic bias and the human rights tools that could address them. Our work will help tech leaders, designers, and technologists in their efforts to embed pluralism and inclusion into AI systems and the future of work.

LED BY

Dr. Rangita de Silva de Alwis, Senior Adjunct Professor, University of Pennsylvania Law School Nonresident Leader in Practice at Harvard Kennedy School 's Women and Public Policy Program Hillary Rodham Clinton Research Fellow on Gender Equity, Georgetown University (2020-2021)

IN COLLABORATION

Steve Crown, Vice President and Legal Counsel at Microsoft

Mitchell Baker, CEO of Mozilla Corporation

Sanjay Sarma, Fred Fort Flowers (1941) and Daniel Fort Flowers (1941) Professor of Mechanical Engineering at MIT and VP of Open Learning

CONTENTS

FOREWORD

In 1956, a group of scientists at Dartmouth College coined the term “Artificial Intelligence” to enable computers to mimic human intelligence. Six decades later, Stephen Hawking has predicted that AI could “end the human race,” and Elon Musk has called it our “biggest existential threat.” Yet we know that AI has profoundly important potential for good in averting mass scale disasters, predicting conflict and droughts, and diagnosing disease, to name a few.

While developers claim that AI can be more objective and reliable than human cognition, most often, algorithms are, in part, our opinions embedded in code. These neural networks use big data to analyze and reveal patterns and trends. The algorithm is trained to behave in a specific way by the data it is fed. Human decision-making is fraught with bias and subjectivity, which correlate to discriminatory behavior. These biases reproduced in algorithms can create not only unintended consequences but systemic and structural biases.

This Foreword explains our experimental work and the twin pillars of our Lab: the importance of story telling and the human rights framework as two important tools in mitigating bias.

The Elephant in AI Surveys

The Lab reports on Elephant in AI look at mitigating algorithmic bias in employment decisionmaking. Although overt forms of discrimination have been mitigated due to antidiscrimination laws, there are limits in the law to provide adequate remedies for those harmed by unintended consequences of discrimination and subtle bias.

In order to better understand what those biases might be the Lab conducted four sample studies. These studies are statistically insignificant and by no means conclusive. Our primary goal was to ask hard questions and to build on the paradigmatic Elephant in the Valley Report.

Five years ago, in 2015, in the Elephant in the Valley Report, researchers from Stanford University and Kleiner Perkins for the first time, collected data on women’s experiences in the tech field.

The Elephant in the Valley survey data revealed that 87 percent of the women reported receiving demeaning comments from male colleagues. Sixty six percent said they had been excluded from important social or networking events. Ninety percent of the women surveyed reported witnessing sexist behavior at their company. Sixty percent had been harassed. Comments indicated that at meetings, women were ignored in favor of male subordinates. However, despite the sexism, women were afraid to complain in fear of retaliation. Five years later, little has changed, and these fears remain a threat to employment security. In 2020, in a much-publicized case, Dr. Timnit Gebru, a preeminent Back woman engineer was forced out of Google for complaining about the lack of diversity in the company, the work she was supposed to address at Google.

A Narrative Theory: Apart from our Lab surveys, the Lab engaged in conversation with a diverse group of tech leaders from major corporations to further understand and interrogate case studies and lived experiences of bias and discrimination in AI and emerging technology.

Judith Butler's feminist theory examines how gender is performative and is constituted and constructed through repeated performance. This performative behavior can be amplified and magnified through AI. The only way we can address these stereotypes is through storytelling and alternative narratives. The women CEOs and other industry partners who collaborated with the Policy Lab on AI and Bias, engaged in story telling as a way to address the lacunae in women's narratives in technology.

Deborah Raji discussed the study by her colleagues Joy Buolamwini and Timnit Gebru study on Gender Shades – the study found that darker skinned females were the most misclassified group with an error rate of 34.7 percent. In contrast, lighter skinned males had a maximum error rate of 0.8 percent.

A Women's Human Rights Based Approach to Addressing Bias in AI:

Although 1.4 million new computer science jobs in the US will be available by 2021, less than 3 percent of those jobs are expected to be filled by women. Moreover, Silicon Valley loses more than 16 billion annually from the turnover of women who enter the tech field reflecting the homogenization of an organization.

Two examples show some of the weightiest challenges to gender equality in technology.

Two years ago, in a live person survey of 1,000 persons, half the respondents could name a male tech leader and only 4 percent could name a female tech leader and a quarter of the respondents named Siri and Alexa – who are virtual assistants as female tech leaders.

Four years ago, a google employee in 2017, circulated an internal email that suggested several qualities, which he thought were more commonly found in women, including high anxiety, explains why they were not thriving in the world of coding. Google fired him on the basis that they could not employ someone who would argue that his colleagues were unsuited for the job.

Although much has been written about the way in which AI implicates human rights, less has been written about applying human rights principles to address inequity in AI. Human rights provide an agreed set of universal norms and a shared framework around which we can engage. While the extant scholarship on human rights and AI have looked at the International Convention on Civil and Political Rights (ICCPR) and the International Convention on Economic Social and Cultural Rights (ICESCR), few have looked at the importance of the Convention on the Elimination of Discrimination against Women's (CEDAW) core articles in combating Bias in AI. Our Lab anchored its work on two articles of the CEDAW as important as standard- setting guidelines.

CEDAW's Article 5:

"modify the social and cultural patterns of conduct of men and women, with a view to achieving the elimination of prejudices and customary and all other practices which are based on the idea of the inferiority or the superiority of either of the sexes or on stereotyped roles for men and women."

CEDAW Article 10:

Article 10 States Parties shall take all appropriate measures to eliminate discrimination against women in order to ensure to them equal rights with men in the field of education and in particular to ensure, on a basis of equality of men and women: (a) The same conditions for career and vocational guidance, for access to studies and for the achievement of diplomas in educational establishments of all categories in rural as well as in urban areas;...

(c) The elimination of any stereotyped concept of the roles of men and women at all levels and in all forms of education by encouraging coeducation and other types of education which will help to achieve this aim and, in particular, by the revision of textbooks and school programmes and the adaptation of teaching methods;

(d) The same opportunities to benefit from scholarships and other study grants; (e) The same opportunities for access to programmes of continuing education, including adult and functional literacy programmes, particulary those aimed at reducing, at the earliest possible time, any gap in education existing between ***men and women....***

The application of CEDAW Article 5 and 10 are important new tools to address algorithmic bias.

In the final analysis, diversity is not about fulfilling quotas, or compliance-based diversity trainings, but rather shifting the culture through education (including human rights and digital humanities), a multi- disciplinary and plural approach to AI development, and inclusion in the highest levels of leadership.

— PROFESSOR RANGITA DE SILVA DE ALWIS

SPEAKERS

Speaker Excerpts

ON WOMEN LEADERS IN AI

Shelly Kapoor Collins

CEO, Shatter Fund

"We just can't afford to leave women further behind, again... **our economy depends on it**. There has never been a more important time to step up and redouble investment in women entrepeneurs."

"One of the hardest parts of my career has been breaking into the male-dominated VC world. When I speak with male counterparts, they often say their portfolios are diversified, but when you really look at those portfolios you see they are not. They are not looking through a gender lens. **Historically, companies with female founders performed 66% better than those with all-male founders.**"

"Women CEOs in tech are very rare. I sometimes wonder if we don't get attention because I'm a woman or because of the non-profit piece."

Mitchell Baker

CEO, Mozilla Corporation

ON INTERDISCIPLINARY APPROACHES TO AI

Safiya Noble

Author, *Algorithms of Oppression*

“I think there will be many more people who will continue to **engage critical race studies and Black feminism in their studies of digital media and technical systems** precisely because it broadens in key ways our ability to pursue questions about issues of power, control, benefit, etc., in ways that are not prioritized by other theoretical approaches.”

Wendy Chisholm

Principal Accesibility Architect, Microsoft

“One of the most exciting things for me is how do we **create anti-racist, anti-ablest design** processes? We bring together design and social work for trans-disciplinary, human-centered design processes.”

“I come to AI with a different perspective because **I am not a technical person. I spent 10 years at Google with two humanities degrees.** I studied renaissance literature, so I’m always thinking about how different disciplines can impact AI.”

Sarah Drinkwater

Director of Responsible Technology, Omidyar Network

ON THE POSSIBILITY OF ETHICAL AI

Mitchell Baker

CEO, Mozilla

“Can you make a system that’s totally ethical? In the big picture I don’t know how to make ethical tech. Bias is obvious, but **if we looked back would we invent the internet? Is it ethical? I don’t know.** Ethics in tech is not a decision of will I build or will I not, it’s a lot of constant evaluations.”.

Deborah Raji

Mozilla Fellow & Forbes 30 Under 30

“One of the most important lessons I’ve learned through this work is how data is inherently biased and includes a lot of decision-making. **Data science is an inherently stereotypical act** because it’s trying to make a whole person out of statistics.”

Rati Thanawala

Fellow, Harvard Kennedy School

“Executives don’t even spend the effort to solve the problems because they have an inherent belief that they’re not wrong. But what people don’t realize is that **making changes in a company’s morals can have a positive business impact**.”

ON HARMFUL BIAS IN AI

Deborah Raji

Mozilla Fellow & Forbes 30 Under 30

"We were involved in a project called Gender Shades, auditing publicly deployed facial recognition products for their performance. We discovered that a lot of these products **don't work very well for darker skinned women.**"

"They are using our money to be biased against us."

Gitanjali Swamy

Managing Partner, IoTAsk

Safiya Noble

Author of Algorithms of Oppression

"When I discovered that we were so grossly misrepresented in search engines and their results, I felt that this would be an important topic to write about. I was disgusted, to be honest, that **pornography was the primary representation of Black girls in large commercial search engines.**"

"Think about Netflix recommendation systems for example. There might be a kid who is **outed to his parents because Netflix recommendations are only showing LGBTQ content.** These recommendations communicate to parents pretty clearly that their child is interested in LGBTQ content. This can be really dangerous for young people who aren't in safe homes."

Jenni Olson

Filmmaker, LGBT Activist

Sarah Drinkwater

Director of Responsible Technology, Omidyar Network.

"Yes, humans often have prejudices that lead to discriminatory decisions, and we often have no way of knowing when and why people are biased. With machine learning we have the potential to make less biased decisions. But **algorithms trained with biased data pick up and replicate these biases** and develop new ones."

"If you're poor in America there are a lot of decisions being made about you by algorithms. They affect how much money you'll get for food, rent assistance, healthcare, and more. These algorithms affect people in a way they don't opt into. **The profit model is not aligned with the user.** These engineers are building tools for people and not thinking about who those tools will affect, and there is the possibility that that incentive is misaligned with that population."

Deborah Raji

Mozilla Fellow & Forbes 30 Under 30

ON AI IN HIRING

Wendy Chisholm

Principal Accesibility Architect, Microsoft

"How do we represent ourselves to technology? How does someone with a disability establish herself to job matching and career matching algorithms. **How do we represent what we've done; what we're capable of.** People identify not as a disabled person with these limitations but as a human with these capabilities. How can a hiring algorithm pick up on that?"

"If you're hiring someone for a management position and you feed your algorithm data from the last 30 years, the data will be skewed, and **the projected ideal candidate will be someone male, white, and in his 40s or 50s.** I am a woman in my early 30s, so I would be filtered out immediately, even if I'm suitable for that position.

Sandra Wachter

Professor, Oxford Internet Institute

"We are always more diverse than AI thinks we are."

Sarah Drinkwater

Director of Responsible Technology, Omidyar Network.

ON TECH COMPANIES AND DIVERSITY IN ENGINEERING

Rati Thanawala

Fellow, Harvard Kennedy School.

"Women and minorities must get to positions of influence. It doesn't mean C-suite necessarily, but it means a person in operations, in research, etc. **You're never going to get the system to change without people in positions to notice the red flags** and identify them as serious problems, not just as exceptions."

"There is definitely a lack of diversity in the engineering ranks at these companies. When I was at Google, **there were almost no machine learning engineers who were not white men.** Humans have blinders, and these blinders make it difficult to access perspectives other than their own."

Sarah Drinkwater

Director of Responsible Technology, Omidyar Network

Deborah Raji

Mozilla Fellow & Forbes 30 Under 30

"At these tech companies, a lot of the challenges they face were the result of different actors who weren't documenting what they were doing. No one was labeling the provenance of the data they were using or which data they were using. A lot of times **it would take months and months for engineers to recognize that there were problems.** So slowing down product development to allow for more testing is one way that they can develop more ethical, careful AI."

“There are intentional and calculated types of bias and also accidental biases. It already means a lot to go after people who discriminate intentionally. But we have to think more about accidental bias because that’s where we see most bias in AI. **This is systemic, historical bias creeping into AI because of bad data and biases of the engineers creating these algorithms.** This is where our focus belongs.”

Gitanjali Swamy

Managing Partner, IoTask

Safiya Noble

Author of Algorithms of Oppression.

“I often say that **those who know so little about society have no business designing and deploying their technologies on society.** I think the colorblind and gender-blind ideologies that bolster a false notion that only the best and brightest are working in large tech corridors around the world—Silicon Valley being the prime example of this - precludes more nuanced and complex ways of thinking.”

ON REGULATING AI

Safiya Noble

Author of Algorithms of Oppression

“When I first started my academic research in this area almost a decade ago, most people, many of whom were in positions of authority over my academic career, in fact, **told me that it was technologically impossible for socio-technical or computational systems to racially discriminate.**”

“Some people suggest that the tech companies should regulate themselves, but **isn’t that a bit like having the fox guard the hen house?** You need both carrots and sticks. Most importantly you need to have long-term social changes and social norms with tech companies being held accountable.

Gitanjali Swamy

Managing Partner, IoTask

“Regulation can cause problems if requirements are not well defined from the outset. Some people in the AI community feel that you can’t always give explanations because **not even the developers of the systems actually understand how they work.**”

“The U.S. believes in a more soft-touch, self-regulatory approach. Their current policies focus more on education of researchers and voluntary codes of practices for the private sector. This might be the result of **their belief that too much regulation can have a negative effect on research, innovation, and economic growth.** The EU is more inclined to create hard laws that are enforceable.”

Sandra Wachter

Professor, Oxford Internet Institute

“Recently, **there has been this incredible kind of awakening.** Finally, tech people are reading this work and engaging with it. Whether they’re focused on trust and safety or policy, they’re starting to think about applying these ethical principles in practice.”

Sarah Drinkwater

Director of Responsible Technology, Omidyar Network.

Speaker Spotlights

MEHRNOOSH SAMEKI

Profile by Lindsay Holcomb

On his first day in office, President Joe Biden repealed former President Donald Trump's executive order on immigration, known colloquially as the "Muslim Ban," which prevented individuals from primarily Muslim and African countries from immigrating to the U.S. The ban caused extraordinary hardship both for Muslims in the U.S. who were separated from their loved ones, and for Muslims living abroad who sought to come to the U.S. for educational and professional opportunities. As a result of the previous administration's immigration policies, the U.S. has experienced a significant drain on its talent pool, particularly in STEM fields, which have for decades been bolstered by immigrants, including many from Muslim-majority countries. According to Eric Rosenblum, Managing Partner at Tsingyuan Ventures, who spoke with the AI & Bias Lab last month, technical universities in Canada and the United Kingdom have seen significant upticks in enrollment from non-native students during the Trump era as engineers from the Middle East and Africa have been repelled by American hostility towards immigrants. Historically, one of the primary feeder countries for STEM students at U.S. universities, particularly female STEM students, is Iran, which boasts far more female engineers than the U.S. In fact, in Iran nearly 70 percent of university graduates in STEM are women, and many come to the U.S. to earn graduate degrees and take on high profile technical and managerial roles at American technology companies. One such engineer is Dr. Mehrnoosh Sameki, who graduated from Sharif University, Iran's premier technical university, with a degree in Computer Engineering and came to the U.S. to pursue a PhD in Computer Science. Today in her job at Microsoft, Dr. Sameki works to eradicate a much more subtle form of prejudice than the overt discrimination of the previous administration's immigration policies: algorithmic bias.

Dr. Sameki has grown accustomed to being one of the only women in many of the engineering spaces she has inhabited since coming to the U.S., so achieving greater diversity, equity, and inclusion in the tech space has become an important commitment for her. To that end, Dr. Sameki is a member of the non-profit Persian Women in Tech, which was established in Silicon Valley in 2015. The non-profit's mission is to "connect, mentor, and empower Persian women in technology globally," and it boasts over 20,000 members, the vast majority of whom possess some sort of technical or entrepreneurial background.

The organization's work also had personal significance for Dr. Sameki who immigrated to the U.S. because she felt that first rate opportunities in science and technology simply did not exist in Iran due to the economic sanctions imposed on the country. Lacking opportunities to make connections with large, multinational technology companies like Microsoft, which do not have offices in Tehran, Dr. Sameki felt that her only choice was to leave. By connecting Persian women in technology globally, Persian Women in Tech opens new frontiers for young engineers like Dr. Sameki to cultivate professional relationships no matter where they are in the world, mitigating historical inequities which have given the upper hand in tech careers to those physically located near Silicon Valley.

This commitment to inclusivity has also translated into Dr. Sameki's work at Microsoft where she leads the product efforts behind open-source offerings for InterpretML, Fairlearn, and Error Analysis, which aim make users aware of potential biases in their machine learning models, identify specific unfairness issues, and mitigate those issues. As Dr. Sameki explains, "AI is only as unfair as the data put into it," and as a result, it can perpetuate historical inequities. Ultimately, Dr. Sameki hopes that her efforts will lead to more transparent and accountable algorithms and encourage AI stakeholders to be more wary of the ways in which their machine learning software may be exacerbating latent societal biases and treating some demographics unfairly.

WENDY CHISHOLM

Profile by Lindsay Holcomb

The phrase, “Stairs make the building inaccessible, not the wheelchair” has become something of a mantra for Wendy Chisholm, Principal Accessibility Architect at Microsoft. Chisholm has spent more than two decades building accessibility into technological tools and processes and encouraging a trans-disciplinary approach to human-centered design.

Since 2018, Chisholm managed the selection process for Microsoft’s AI for Accessibility program – a $25 million grant program to accelerate AI innovations that are developed by or for people with disabilities. Her efforts have spawned projects including Object Recognition for Blind Image Training (ORBIT), which created a public data set using photos and videos submitted by blind and low vision people to facilitate personalized object recognition, and SeeingAI, an iOS application which uses a device’s camera to identify and audibly describe people, objects, and images for blind and low vision people. These innovations have been described by their users as truly life-changing, allowing them to more safely and assuredly interact with the world around them.

Among her numerous accolades, she has independently consulted for companies including Microsoft, Google, and Adobe to integrate universal design and helped the University of Washington’s Access Computing project increase the number of people with disabilities in computing fields. As Chisholm explains, centering disability in conversations about AI can help remediate some of AI’s harms and make AI more inclusive for the more than one billion people worldwide with a disability.

Accessibility is an issue of immense personal importance to Chisholm, who has written publicly about her experience surviving trauma and living with PTSD. Her personal experiences have made her think differently about all sorts of aspects of the fast-paced, high profile world of large tech companies. “As a woman with disabilities, I cannot work 80-hour weeks,” Chisholm explained, adding that when she goes to multi-day industry conferences she builds in time for the naps and meditations that allow her to be her best. “It’s important to remember that people identify not as a disabled person with these limitations, but as a human with these capabilities.” Carefully listening to how people with disabilities represent themselves is crucial to ensuring that biases and partial inferences do not form the bases of algorithm-based understandings of their lived experiences.

Diversity is a human right that is part of our universal human rights framework, and the right to participate in public life is a bedrock guarantee enshrined in the Convention on the Rights of Persons with Disabilities (CRPD). Chisholm’s work has been both inspired and informed by these efforts, and she often invokes the call to action that galvanized the disability rights community over a decade-long process in drafting the CRPD: “Nothing about us without us.”

DEBORAH RAJI

Profile courtesy of MIT Review

The spark that sent Inioluwa Deborah Raji down a path of artificial-intelligence research came from a firsthand realization that she remembers as "horrible."

Raji was interning at the machine--learning startup Clarifai after her third year of college, working on a computer vision model that would help clients flag inappropriate images as "not safe for work." The trouble was, it flagged photos of people of color at a much higher rate than those of white people. The imbalance, she discovered, was a consequence of the training data: the model was learning to recognize NSFW imagery from porn and safe imagery from stock photos—but porn, it turns out, is much more diverse. That diversity was causing the model to automatically associate dark skin with salacious content.

Though Raji told Clarifai about the problem, the company continued using the model. "It was very difficult at that time to really get people to do anything about it," she recalls. "The sentiment was 'It's so hard to get any data. How can we think about diversity in data?'"
The incident pushed Raji to investigate further, looking at mainstream data sets for training computer vision. Again and again, she found jarring demographic imbalances. Many data sets of faces lacked dark-skinned ones, for example, leading to face recognition systems that couldn't accurately differentiate between such faces. Police departments and law enforcement agencies were then using these same systems in the belief that they could help identify suspects.

"That was the first thing that really shocked me about the industry. There are a lot of machine-learning models currently being deployed and affecting millions and millions of people," she says, "and there was no sense of accountability."

In 2016, MIT researcher Joy Buolamwini (one of MIT Technology Review's 35 Innovators Under 35 in 2018) gave a TEDx talk about how commercial face recognition systems failed to detect her face unless she donned a white mask. To Raji, Buolamwini was the perfect role model: a black female researcher like herself who had successfully articulated the same problem she had identified. She pulled together all her code and the results of her analyses and sent Buolamwini an unsolicited email. The two quickly struck up a collaboration.

At the time, Buolamwini was already working on a project for her master's thesis, called

Gender Shades. The idea was simple yet radical: to create a data set that could be used to evaluate commercial face recognition systems for gender and racial bias. It wasn't that companies selling these systems didn't have internal auditing processes, but the testing data they used was as demographically imbalanced as the training data the systems learned from. As a result, the systems could perform with over 95% accuracy during the audit but have only 60% accuracy for minority groups once deployed in the real world. By contrast, Buolamwini's data set would have images of faces with an even distribution of skin color and gender, making it a more comprehensive way to evaluate how well a system recognizes people from different demographic groups.

Raji joined in the technical work, helping to prepare the data for Buolamwini's audits. The results were shocking: among the companies tested—Microsoft, IBM, and Megvii (the company best known for making the software Face++)—the worst identified the gender of dark-skinned women 34.4% less accurately than that of light-skinned men. The other two didn't do much better. The findings made a headline in the New York Times and forced the companies to do something about the bias in their systems.

Raji has since worked on several other projects that have helped set standards for algorithmic accountability. After her time at the Media Lab, she joined Google as a research mentee to help the company make its AI development process more transparent. Whereas traditional software engineers have well-established practices for documenting the decisions they make while building a product, machine-learning engineers at the time did not. This made it easier for them to introduce errors or bias along the way, and harder to check such mistakes retroactively.

Along with a team led by senior research scientist Margaret Mitchell, Raji developed a documentation framework for machine-learning teams to use, drawing upon her experience at Clarifai to make sure it would be easy to adhere to. Google rolled out the framework in 2019 and built it into Google Cloud for its clients to use. A number of other companies, including OpenAI and natural-language processing firm Hugging Face, have since adopted similar practices.

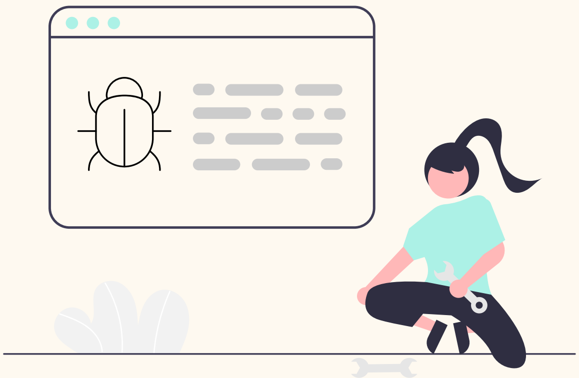
It hasn't always been easy. At Google, she saw how much time and effort it took to change the way things were done. She worries that the financial cost of eliminating a problem like AI bias deters companies from doing it. It's one reason she has moved back out of industry to continue her work at the nonprofit research institute AI Now. External auditing, she believes, can still hold companies accountable in ways that internal auditing can't.

Testimony of Manish Rhaghavan, PhD. to New York City Council

SPEAKER TO AI BIAS POLICY LAB

My name is Manish Raghavan. I am a researcher at Cornell University studying the societal impacts of algorithmic decision-making, particularly in the context of hiring. I have extensively studied the types of automated employment decision tools being discussed today, and my testimony is largely based on this research.

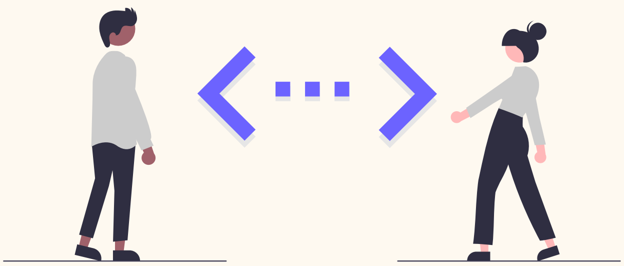
In this testimony, I offer my recommendations regarding Int. 1894, which seeks to regulate algorithmic tools deployed for candidate evaluation. I appreciate the Council's attention on this important topic. Automated employment decision tools are increasing in prevalence, often with little to no public transparency into their inner workings. In my view, this bill is a step in the right direction. In its current form, it carries some vital provisions to ensure that automated hiring tools are carefully scrutinized for potential discrimination. At the same time, it's important to recognize the limitations of this bill (and indeed, any attempt to regulate these tools through prospective auditing). In this testimony, I will detail two such limitations: 1. Current interpretations of anti-discrimination law do not preclude all discriminatory behavior that algorithms can exhibit. 2. Audits have limited power to detect discrimination in terms of undisclosed attributes, such as sexual orientation or disability status. Before diving deeper into these points, it's important to note that hiring tools can perpetuate discrimination even in the absence of explicit bad actors. Due to historical patterns of inequity, algorithms can behave in discriminatory ways simply due to negligence or insufficient attentiveness to these issues. It's crucial that we implement guardrails that protect us from these more subtle, insidious forms of discrimination.



Current interpretations of anti-discrimination law do not preclude discriminatory behavior. Vendors of automated employment decision tools, to the extent that they consider issues of bias at all, typically think of anti-discrimination law in terms of the EEOC's 4/5 rule. The 4/5 rule requires that applicants from different protected groups be selected at roughly the same rate—that is, if half of the candidates evaluated are women, then approximately half of the candidates selected by the tool should be women. A violation of the 4/5 rule do not necessarily constitute discrimination, but it can be the basis to open a discrimination suit. In the absence of specific requirements, it is natural that bias audits will focus on ensuring that tool in question satisfies the 4/5 rule. In my view, this is insufficient, and inconsistent with standards in industrial-organizational psychology.

A particularly important metric to consider is validity, which measures how good a tool is at correctly identifying high- vs. low-performing candidates. How is validity related to bias? One key way in which algorithmic tools can discriminate is via differential validity, which occurs when a tool is better at evaluating members of one group than another. For example, if the tool is very good at identifying the top-performing white candidates and not very good at identifying the top-performing African-American candidates, this would be an instance of differential validity. Even if an assessment satisfies the 4/5 rule, meaning it recommends candidates from all racial groups at roughly equal rates, the top-performing African-American candidates would be more likely to be screened out by the assessment than their white counterparts. Differential validity has been repeatedly found in practical applications of data-driven decisionmaking,[3] and it's important to ensure that employment decision tools don't perpetuate this form of discrimination. Assessments that exhibit differential validity are not explicitly illegal, according to current interpretations of the law. However, simply requiring an auditor to report on measures of differential validity may induce vendors of automated employment decision tools to ensure that their products work well for everyone, not just those who have been well-represented in historical data. In my view, testing whether a tool performs well across the entire population should be an integral part of any bias audit, and to this end, I believe this bill should explicitly require differential validity testing. Audits have limited power to detect discrimination in terms of attributes like sexual orientation or disability status. Audits can only be performed with respect to protected attributes on which vendors maintain data. If a vendor doesn't collect data about, say, applicants' sexual orientation, it is impossible for an auditor to know whether a tool produces disparities along these attributes. Nor is it necessarily desirable that vendors maintain this sort of sensitive data; applicants may not feel comfortable divulging this information.

Thus, an audit cannot identify all forms of illegal discrimination, and as such, it's important to be clear on the goals of such an audit. The current language of Int. 1894 refers to compliance with "any . . . applicable law relating to discrimination in employment." In practice, this will not be possible. We should acknowledge the narrow scope of what is possible through audits, and what forms of discrimination cannot be detected through these means. Recommendations. While the above challenges are in a sense inherent to the problem of auditing for bias, there are concrete steps we can take to begin to address them. 1. Set specific standards for what measures should be included in an audit. 2. Require auditors to report on metrics of differential validity. 3. Use caution in interpreting the results of audits. An audit can only test for specific discriminatory behaviors; it cannot certify that a tool is free of bias.

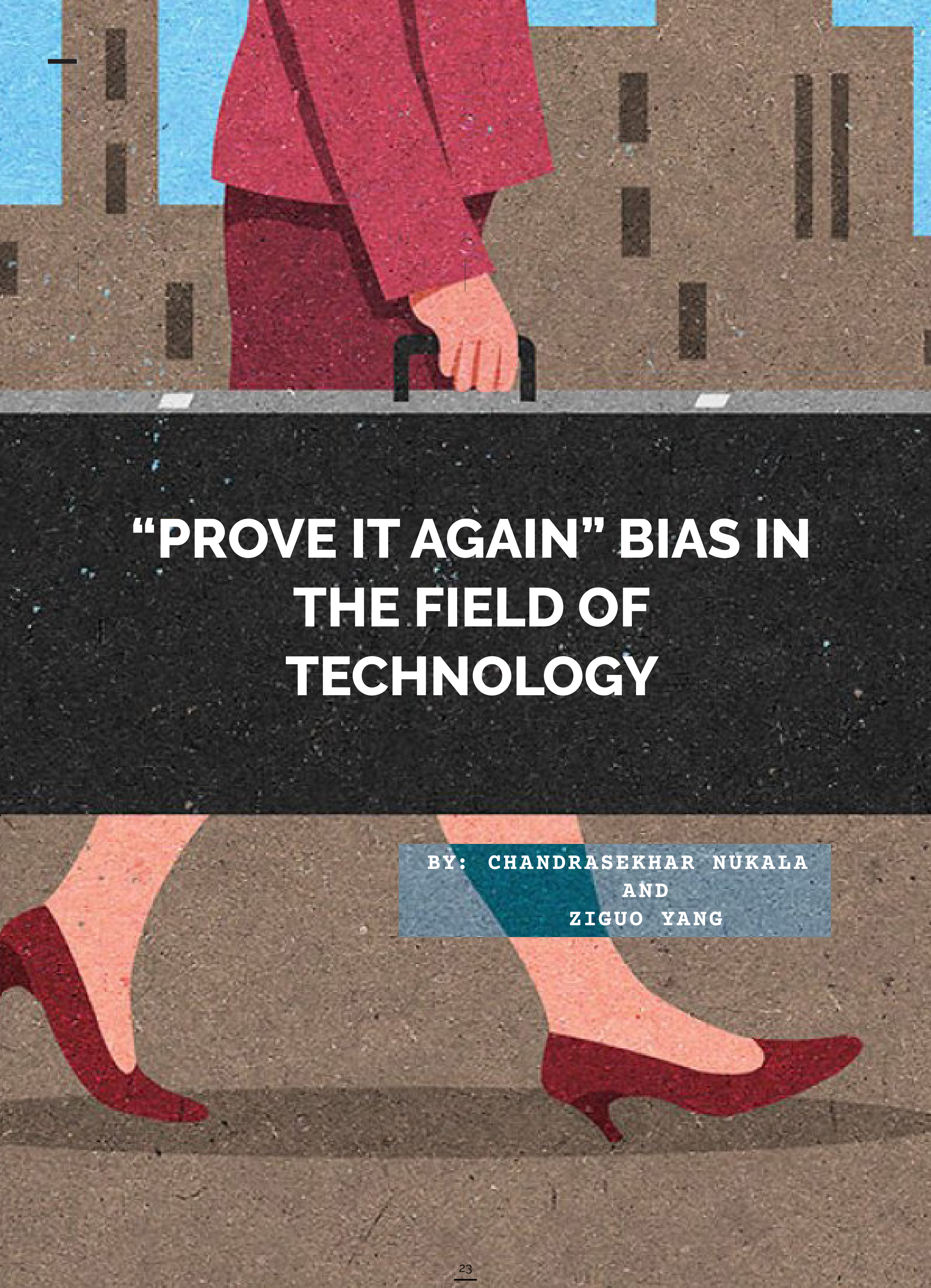


SURVEY IN SILICON VALLEY

What is the „Prove It Again“ Bias?

The “Prove it again!” bias, is the concept that in a male dominated industry, men are presumed to be competent based on their potential, while women, especially women of color often have to prove their worth over and over again. For example, men may be given the benefit of the doubt, but not women. In addition, women’s mistakes may be highlighted and constitute a blemish in her record, while men’s mistakes are ignored or soon forgotten.

“One of the most common examples of ‘prove it again bias” is the double standard that men are judged on their potential, while women are judged strictly on what they already have accomplished,” argues Professor Joan Williams. The prove it again bias is compounded when it comes to women of color in the workplace.Can AI systems designed for job performance be trained to mitigate this bias or will they run the risk of reproduce it?



"PROVE IT AGAIN" BIAS IN THE FIELD OF TECHNOLOGY

BY: CHANDRASEKHAR NUKALA AND ZIGUO YANG

Introduction

Silicon Valley's diversity problem has been well documented since the landmark "Elephant in the Valley" reports. Recent news from Snap Inc's revelations on its diversity shortcomings show that there is very little progress.

Many of the companies in the Valley have switched to people analytics to highlight the use of big data and algorithms from hiring to performance management. The use of Big Data in a world full of biased datasets and gender data gaps amplify and accelerate harms caused to women and people of color.

This paper tries to answer two important questions: Do women and people of color have to prove their competence over and over again? Are these biases being replicated into newer AI based platforms? The answer to both these questions is a resounding ***YES***. Prove It Again bias is acutely felt by women, is worse for women of color and is worst for women over the age of 35.

Professor Williams and Rachel Dempsey in their 2018 book, What Works for Women at Work, say that "Men are often judged on their potential, but women are judged on their achievements". Based on the women I have spoken to in my career as well as the women surveyed as part of this project, women saw "potential" is used as a sword to cut them and their accomplishments down again, again and again. For women, the term potential has become more negative as they progressed in their career.

The report is based on a survey of 47 individuals employed in various technology companies in the Silicon Valley regarding their perceptions and attitudes of "Prove It Again!" bias and its impact on social platforms like LinkedIn. The respondents predominantly were women and people of color. Around one-fifths of the respondents left comments, the tone of the comments shows an interest and reaction to the topic of Prove it Again! bias. The survey asked respondents if they had faced Prove it again bias and the mitigations that they took to work around the bias.

The author created an online questionnaire, that consisted of five parts, background questions, Likert scale questions to identify prove it again bias, effects of prove it again bias, mitigations pursued by the individuals and a qualitative question. The questionnaire was distributed through the author's social network and internet communities based in the Silicon Valley from February 19, 2021 to March 18, 2021. Background information and Likert scale analysis is located in the appendix.

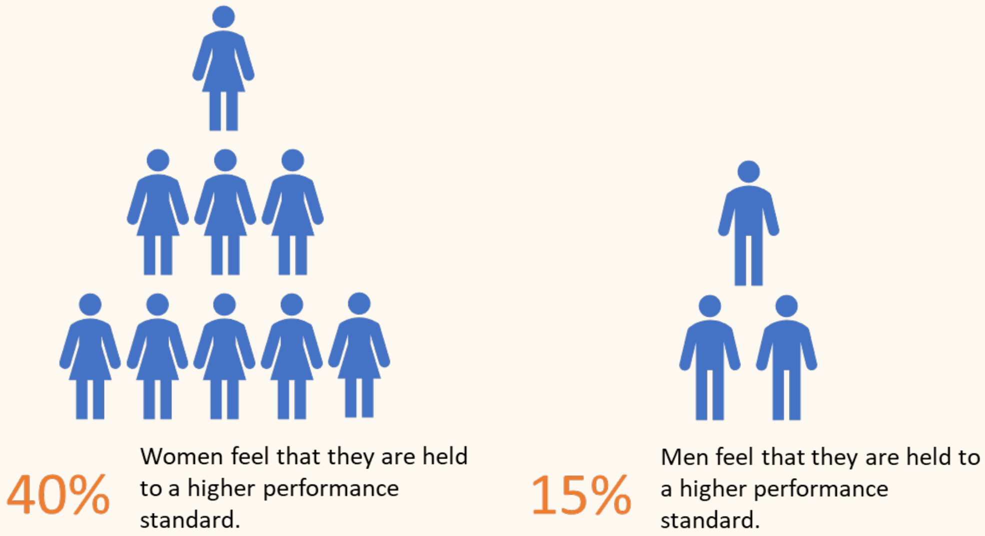
The lab has partnered with the President of TiE Boston, Anu Chitrapu and the Private Capital Research Institute at Harvard Business School's Dr. Gitanjali Swamy on a survey on "Prove it Again Bias. "

53% of women vs. 15% of men reported that they have to consistently prove themselves to get the same levels of rewards as their peers.

" Being a predominantly software (woman) engineer/ software (woman) executive in a hardware company was always given the least amount of time and budget, had to bail out hardware engineers/executives all men and when company fell on hard times all my team were terminated and our jobs were sent to India. Company product quality declined dramatically, number of women in organization dropped and many terminated engineers left the industry. "

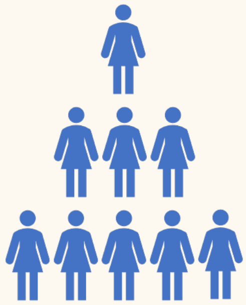
-A woman respondent

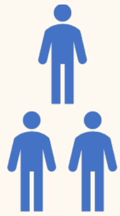
Studies dating back to 1970s have documented that Women and people of color often need to be more competent than white men in order to be seen as equally competent.
The survey that we conducted across multiple silicon valley working women reiterated the same with 40% of women saying they were held to higher performance standard than their peers while only 15% men felt the same.



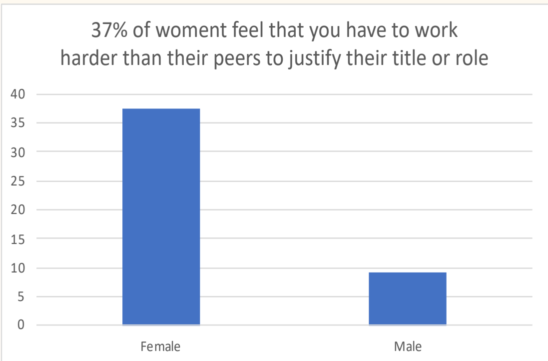
What was interesting to note was that more than 53% women felt that they have to consistently prove themselves to get the same rewards as their peers while the percentage of men remained same at 15%.

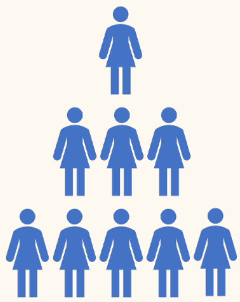
There was not much of a difference in the response between rewards and recognition as in many instances rewards are tied to recognition and vice-versa.

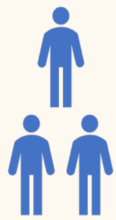




There was not much of a difference in the response between rewards and recognition as in many instances rewards are tied to recognition and vice-versa.







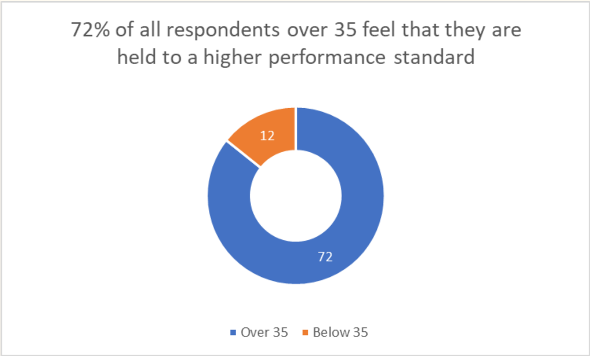
When it came to working harder, only 37% women felt that they have to work harder than their peers. One of the reasons for the lower percentage might be that the younger women and some of the women over 35 are used to working harder so hardship has become the norm that there are very little distinguishing factors in their view between working hard to working harder.

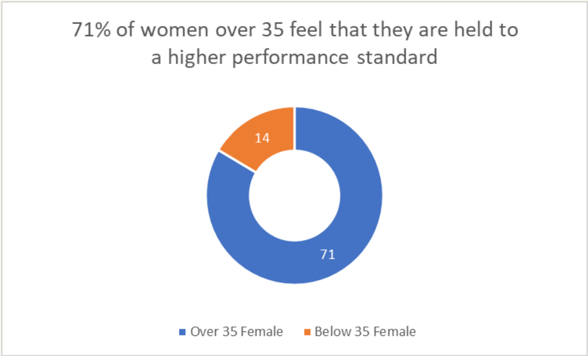
The prove-it-again phenomenon is a complex bias that is fueled by double standards (Includes leniency bias), in-group favoritism and confirmation bias. The survey highlights that, early in one's career it is very hard to spot prove-it-again bias as opposed to the individual proving themselves in an official environment.

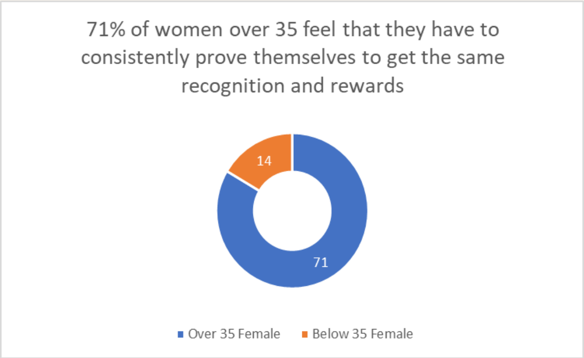
*“The “prove it again” bias, in many ways, is a belief grounded in comparing oneself with another person. However, there are times where I need to “prove it again” to myself that I am capable of achieving what I set my mind to.
I think a little bit of “prove it again” bias is beneficial to push me past my limit or ceiling. But, there is also a fine line, that when crossed, can be detrimental to ones mental health.”*

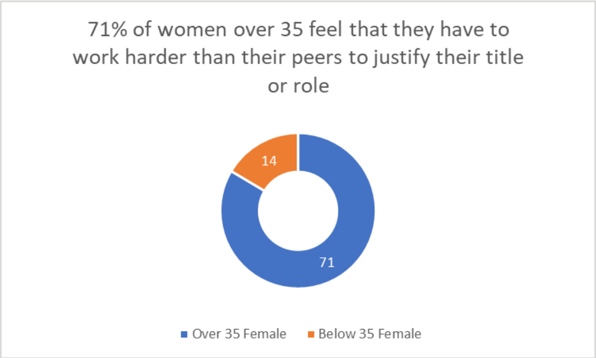
-A woman under age 35

This was also seen in the survey results.









The prove-it-again bias is triggered by gender, race, disability and more importantly age. This bias along with in-group favoritism leads to women losing out on many opportunities (They are not part of the “good old boys club”)

" As an older woman, my organizational and social skills are acknowledged, even overestimated, and promoted. I am bypassed when it comes to more technical opportunities, even when my resume clearly puts me in the lead. My technical ability is often ignored, especially in group situations, and especially by men in my age group. Because of my demographics, I am not comfortable with and often not included in casual networking, and am often out of the loop. "

There is a prevailing assumption that women are not fit for engineering.

" "I notice a few of my male colleagues have a tendency to rephrase something I say in a group discussion. "

Many Women noted that their credentials were discounted.

"There is a ground hog day feeling of always having to start from the ground floor at everything with no recognition for my credentials."

Not only are their credentials discounted, but their successes are also attributed to others in the organization.

"The elephant is always in the room. It is not very easy to mitigate such bias. After building a high-productive team, and delivering the best results and i was moved to take care of other groups and the person who stepped in my role was promoted in the very next cycle (6 months)."

Women leaders performance is taken for granted, their budgets cut and assigned impossible goals.

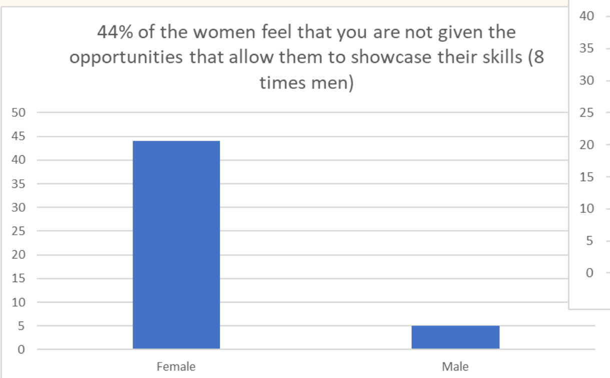
"

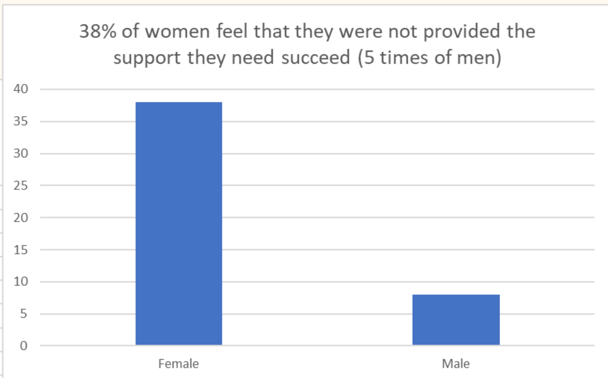
"Often held the closest relationships with customers and yet sales people who didn't even understand the products were given the credit and compensation.
Opened new markets and designed features to obtain new customers and credit was always given to men in the company.
Being a predominantly software(woman) engineer/software (woman) executive in a hardware company was always given the least amount of time and budget, had to bail out hardware engineers/executives all men and when company fell on hard times all my team were terminated and our jobs were sent to India. Company product quality declined dramatically, number of women in organization dropped and many terminated engineers left the industry."

"

WThere is also a sinister impact of the unconscious biases - assigning work and providing support to enable success. As professor Williams in her Harvard business review article states "Some assignments can set you up for promotion — this is the glamour work. Other assignments are necessary but unsung — this is the office housework. Research shows that women and people of color are much more likely to get housework-type assignments". Assignment gaps is prevalent in Silicon Valley and the tech industry.

44% of women responded saying that they did not feel they were given opportunities to showcase their skills as compared to 5% of men.



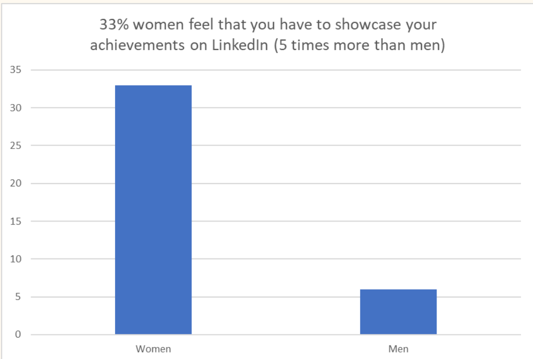


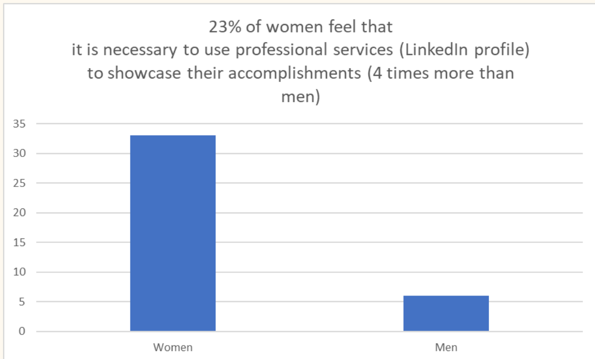
Furthermore, In the few instances when they had the opportunity, 38% of women felt that they were not provided the support to succeed as compared to 8% of men.
For women in the Silicon Valley on the career ladder, Prove-It-Again bias makes the climb very difficult if not impossible. While this kind of bias is discouraging, many of the professional women are using ways and working through or around the bias.

Mitigations

INDIVIDUAL MITIGATIONS

Professor Joan Williams in her book "What works for women at work" provides many mitigations for women and people of color to work around the bias. Some of the strategies recommended by her are: Trump the stereotype, Get over yourself, Know your limits, Address the bias -with kid gloves and Play a specialized or technical role.
Women are making their own mitigations that align with Professor William's recommendations. The survey reveals that - Women were five times more likely to "showcase their achievements on LinkedIn (and other social media)" and were four times more likely to use professional services to showcase their accomplishments.





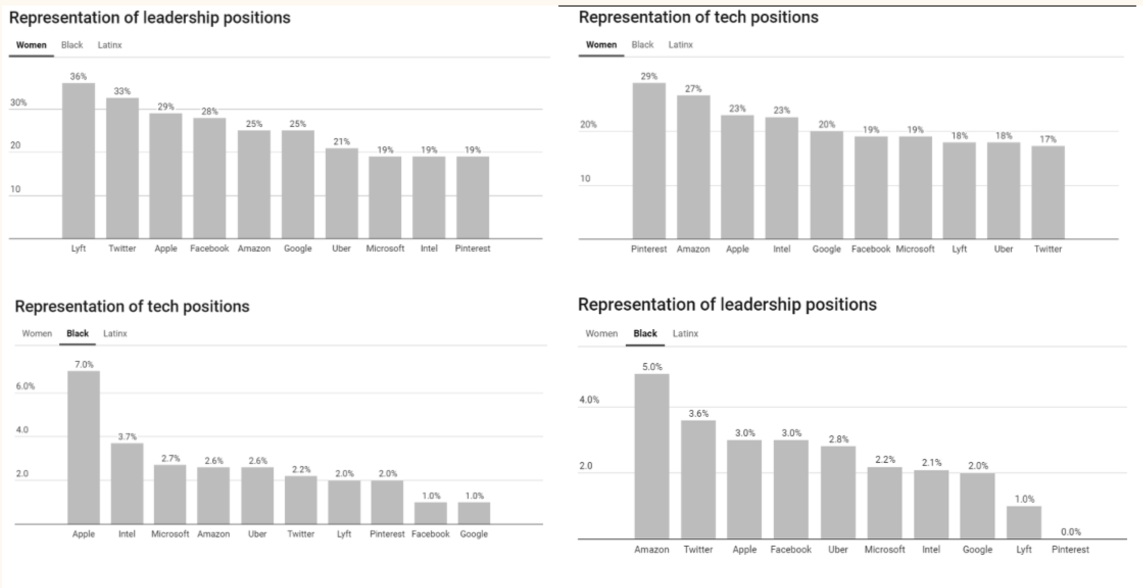
REGULATORY MITIGATIONS

The problem with just these individual mitigations is that while this solves the problem at an individual level the systemic problems persist. The systemic problems bleed in the datasets being used to train algorithms that lead to algorithms of mass discriminations (An example of how the algorithms discriminated against women is Amazon's hiring AI tool. Link: https://www.aclu.org/blog/womens-rights/womens-rights-workplace/why-amazons-automated-hiring-tool-discriminated-against The tool downgraded resumes that included the word "women's" — as in "women's rugby team.")
The focus of future research and recommendations should be the companies and institutions. One of the most important question that one needs to ask is the about the role of government and regulation in explicit and implicit biases leading discrimination whether the biases are held by humans or algorithms.

Disclosure-Based Regulations

DIVERSITY DISCLOSURES

The first step towards any recommendations or regulation is collection of data. The Equal Employment Opportunity Commission is looking at determining the current workforce composition, but this is very rudimentary (Current EEO-1 component survey requires companies to provide employment data categorized by race or ethnicity, gender, and job category). The EEO-1 survey has laid bare the diversity truths across the Silicon Valley and the tech industry as shown in the graphs below.



ALGORITHMIC USE AND AUDIT DISCLOSURES

Silicon Valley is at the leading edge of AI innovation and uses algorithms in all aspects of their business but there is very little transparency in the use of algorithms - do they use algorithms to hire, promote or employee evaluation? Do they test the datasets, models and algorithms for biases and discrimination? Again, FTC should require companies to disclose the use of the algorithms and the corresponding risk assessments.

FIDUCIARY DUTY AND INCIDENT DISCLOSURES

Finally, when individuals report incidents of bias or discrimination – these incidents are investigated by the employer's HR agents with the singular goal of reducing the company's risk. There is an information asymmetry that makes the reporting of the incident more of an antiquated check-box item than a true inquiry. The government should mandate that the individual HR agents investigating the reports of any workplace incident should have fiduciary duty to the employees reporting the incident. OSHA should require that companies disclose the various incidents including bias and discrimination complaints.

Corporate Mitigation

Silicon Valley has largely focused on being compliant with the law. Companies follow the framework outlined in the United States Sentencing Guidelines for corporate compliance even though the individual steps provide little improvement or in many cases are harmful in Diversity and Inclusion programs. For example, there have been many studies that highlight that diversity training does not work but the sentencing guidelines state that "Training" is the hallmark of a well-designed compliance program.
There are many mitigations that corporations can make to mitigate the Prove it again bias. The paper lists a few of the important one below.

PROJECT ASSIGNMENT TO EMPLOYEES

Companies should classify all the projects into two categories – glamour work and office housework. These projects should be allocated fairly across all groups. There are many mechanisms for fair allocation of the projects – one of the most common way of allocation is to allocate equal number of glamour and office housekeeping projects to individuals/teams. This enables the company to evaluate the individuals/teams on all dimensions -performance on glamour and office housekeeping projects.

AI FOR EMPLOYEE EVALUATION

The first recommendation is to use AI or algorithms for employee evaluation. This will enable the company to define and document the attributes for success. The attributes will have to be specific, measurable, and time-bound. These attributes will apply to all the employees not just one group. Another advantage of using this method is that it removes subjective measurements like "potential" which has come to mean various biases. Another method to enable use of subjective measures is to track the false positives and false negatives with these subjective measures. The evaluators who use these subjective measures should be evaluated using the false positives/negatives.

SECOND LOOK

Another mechanism to mitigate the prove it again bias is the use of second look. Individuals and teams that do not fit the mold or are outliers should be given a second look when allocating the projects and evaluations for promotions. MIT Sloan has used this method to create a hiring algorithm to improve hiring quality and diversity in companies (https://mitsloan.mit.edu/ideas-made-to-matter/exploration-based-algorithms-can-improve-hiring-quality-and-diversity). The algorithm assigns an "exploration bonus" to candidates whose quality the firm knows the least about (given the firm's existing data) and surfaces their resumes to human evaluation - A second look.

The second look method is used in various facets but not employee evaluation. Another example of second look is in the Venture Capital industry. If the founding team had unique characteristics, the VCs would give the team and their pitch a second look. The second look enables the VCs to overcome their own biases and blind spots.

Conclusion

Prove it again bias and other implicit biases make is harder, if not impossible, for women and people of color to succeed in Silicon Valley.

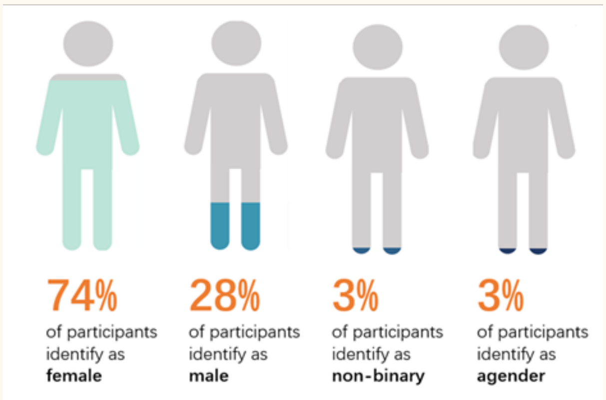
A lot of current research into "Prove it again" bias is focused on identifying biases and individual mitigations. For broad systemic change, research for mitigations should focus on corporations and regulatory agencies. This paper attempts to get the discussion started in these two areas.

Corporate mitigations need to focus not only on the operational aspects but also compliance aspects like "Tone at the top" and executive buy-in. Regulatory mitigations start with disclosures and need to expand to align with the equal rights laws. These steps will help bring about the systemic changes that are long overdue.

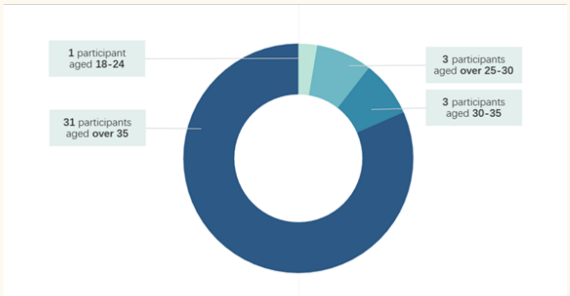
Appendix

Survey background

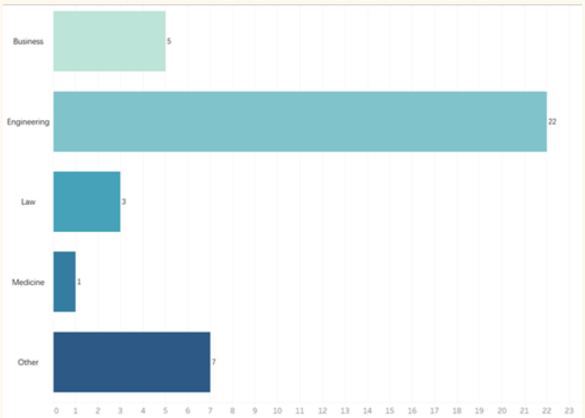
Gender



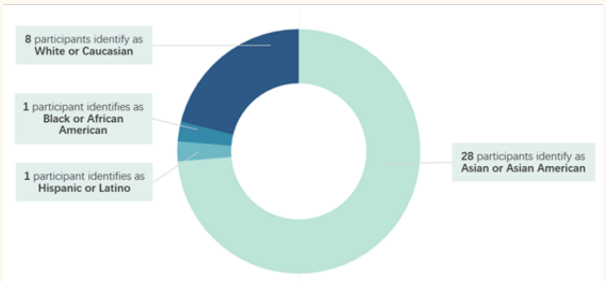
Age



Field



Racial Identity



Role

Likert Scale Analysis

Likert Scale was used to understand the questionnaire data and the range of intensity of the respondent feelings. In order to determine the Likert Scale reliability, Cronbach alpha was used as a measure.

Cronbach alpha values were calculated using the ANOVA method. The Cronbach alpha across the data set as well as groups of Over 35 and Male-Female show high values as shown below. The high Cronbach alpha values reiterate the reliability of the data.

Data	Cronbach Alpha
Overall	0.8772
Over 35	0.8985
Male-Female	0.8476

Women average scored 3.52 on the question "How much, if at all, do you feel that you have to consistently prove yourself to get the same level of rewards as your peers?" while Men on an average scored 3.14 hence Women strongly feel that they consistently have to prove themselves compared to Men to get the same level of rewards.

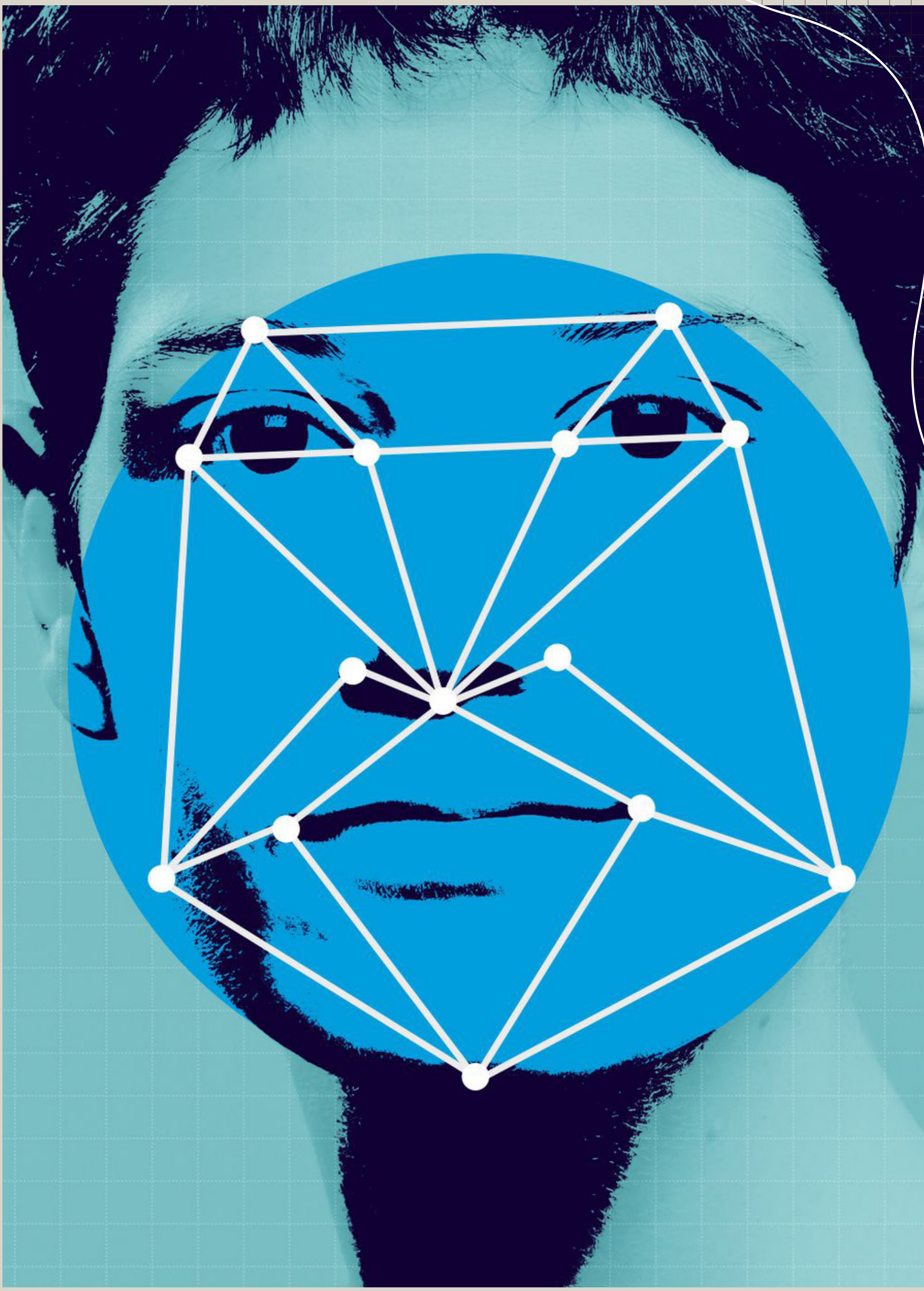
Women on average scored 3.347 on the question "How much, if at all, do you feel that you have to consistently prove yourself to get the same level of recognition as your peers?" while Men average score was 3.143 hence Women felt strongly that they have to provide themselves to get the same level of recognition as peers when compared to Men.

Women on average scored 3.043 on the question "How much, if at all, do you worry that you have to work harder than your peers to justify your title or role?" while Men average score was 2.714 hence Women had to work harder than their peers to justify the title or role when compared to Men.

Bibliography

1. Joan Williams, Rachel Dempsey, and Anne-Marie Slaughter, What Works for Women at Work: Four Patterns Working Women Need to Know
2. Joan Williams, The 5 Biases Pushing Women Out of STEM, https://hbr.org/2015/03/the-5-biases-pushing-women-out-of-stem
3. Deborah Rhode, Women and Leadership, Oxford Press
4. Caroline Criado Perez, Invisible Women: Data Bias in a World Designed for Men
5. Joan C. Williams and Marina Multhaup, For Women and Minorities to Get Ahead, Managers Must Assign Work Fairly, https://hbr.org/2018/03/for-women-and-minorities-to-get-ahead-managers-must-assign-work-fairly
6. Joan Williams, BIAS INTERRUPTERS, https://biasinterrupters.org/toolkits/individualtools/
7. Rachel Goodman, Why Amazon's Automated Hiring Tool Discriminated Against Women, https://www.aclu.org/blog/womens-rights/womens-rights-workplace/why-amazons-automated-hiring-tool-discriminated-against
8. Rani Molla, How Facebook compares to other tech companies in diversity, https://www.vox.com/2018/4/11/17225574/facebook-tech-diversity-women
9. Brian Eastwood, Exploration-based algorithms can improve hiring quality and diversity, https://mitsloan.mit.edu/ideas-made-to-matter/exploration-based-algorithms-can-improve-hiring-quality-and-diversity
10. Taylor Fang, Perspectives from female founders, https://medium.com/tsingyuan-ventures/perspectives-from-female-founders-3749336eaf88
11. Chandra Nukala, Can AI provide a 'Second Look?', https://www.law.upenn.edu/live/news/10981-ai-and-implicit-bias-policy-lab-asking-hard

COMPARISON SURVEY CHINA



BIAS AND HIRING ALGORITHMS IN CHINA

ZIGUO YANG

INTRODUCTION

This report is based on an empirical study of 68 Chinese lawyers and 93 Chinese computer engineers regarding their perceptions of online hiring platforms and attitudes towards bias underlying the hiring algorithms.

The data, collected through the online questionnaire, demonstrate that female professionals are more likely to be recommended with jobs that are below their qualifications and they are, in general, more sensitive to bias incurred by hiring algorithms than their male counterparts. Moreover, Chinese computer engineers, regardless of gender, tend to be concerned about bias related to age and gender, while Chinese lawyers, especially female lawyers, are more likely to worry that employers using hiring algorithms may not consider them due to their marital status and work experience.

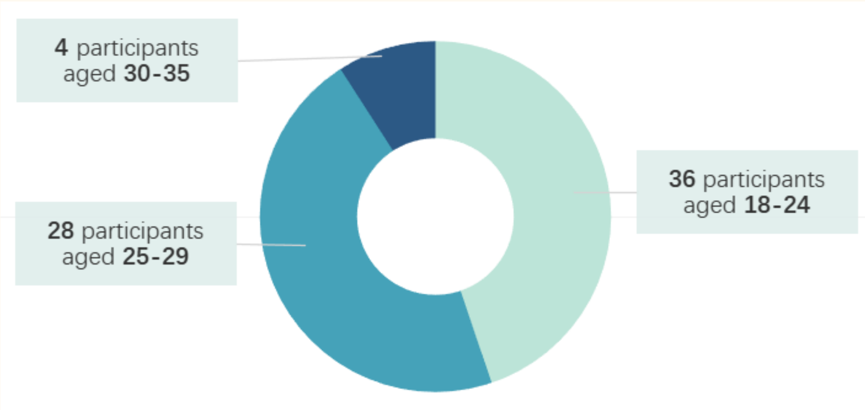
Following the empirical analysis, this report also provides an experiment designed to explore bias against female candidates in one of China's mainstream hiring websites. The result discloses that, for candidates with entirely identical backgrounds, male candidates tend to receive more invitations from the employers, and the positions offered to male candidates provide a higher pay on average. Although the contributors to such discrepancy between male and female candidates remain unclear, this result does imply that the hiring algorithms may not fulfill the promise of eliminating human bias in hiring process. This report, as noted by Collett and Dillon (2019, p. 10), is aimed at demonstrating challenges posed by AI toward gender equality and applying feminist theories to the process of hiring using algorithms.

EMPIRICAL ANALYSIS

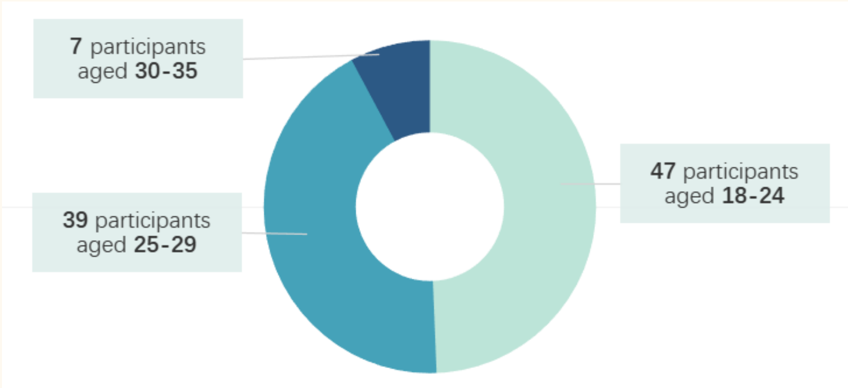
Methodology

Without publicly available data and models, it is difficult to comprehensively characterize the practice of hiring platforms. As illustrated by the empirical study focused on the assessment vendors offering pre-employment assessments, consultation can serve as an efficient approach to identify potential bias incurred by the hiring algorithms (Raghavan et al, 2019, p. 5). In order to simplify the process and explore different dimensions of bias incurred by hiring algorithms, the author established an online questionnaire on wjwlkj.wjx.cn, which consists of three parts, namely descriptive questions, Likert scale questions, and a qualitative question. The questionnaire was distributed through the author's social network and internet communities of Chinese lawyers and computer engineers. From February 16, 2021 to March 16, 2021, 202 answers were collected, and 161 were effective.

DESCRIPTIVE ANALYSIS

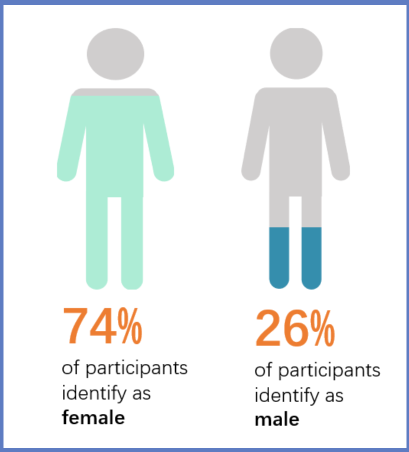


Graph 1: Age distribution of respondent lawyers

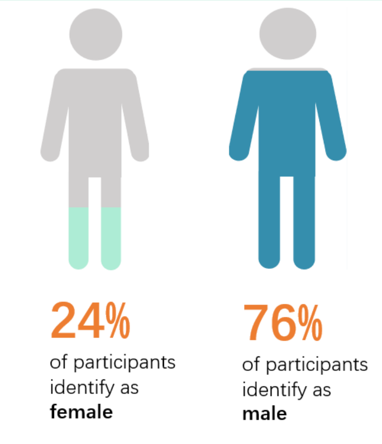


Graph 2: Age distribution of respondent computer engineers

Based on the data collected, 52% of the participants are in the age group of 18-24 and 42% of the participants are aged 25-29. Participants falling within these two age groups are usually people who are starting and growing their career. These groups of people, also known as Generation Z and Millennials, are also the most internet-dependent generations who are most likely to seek for jobs through online hiring platforms.



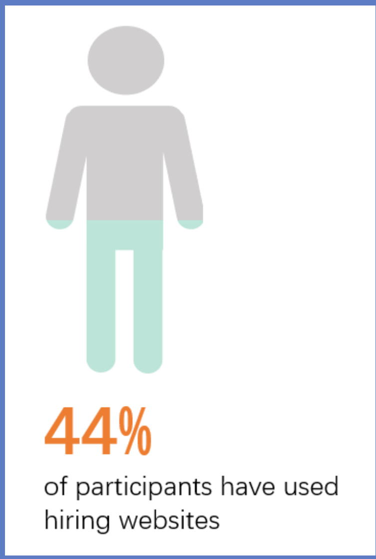
Graph 3: Gender Distribution of respondent lawyers



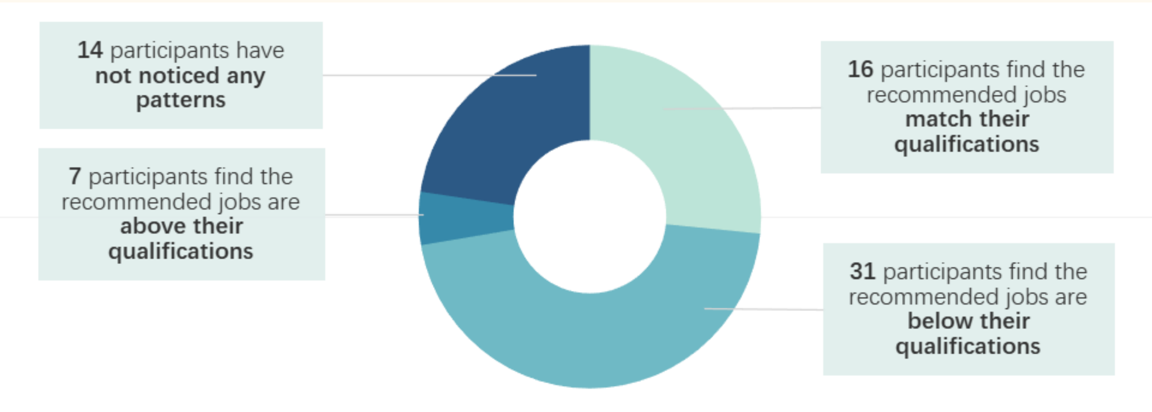
Graph 4: Gender Distribution of respondent lawyers

The gender ratios of participants, however, are uneven in both groups. The fact that the female computer engineers account for less than 25% of the group reflects the dearth of women in STEM professions (Raso et al, 2018, p. 43). Furthermore, the uneven gender distribution may affect the overall results of the Likert scale questions, especially those relevant to gender. To eliminate such influence, this report provides comparison not only between lawyers and computer engineers, but also between female professionals and their male counterparts.

Graph 5: Proportion of lawyers using hiring platforms



Graph 6: Proportion of computer engineers using hiring platforms

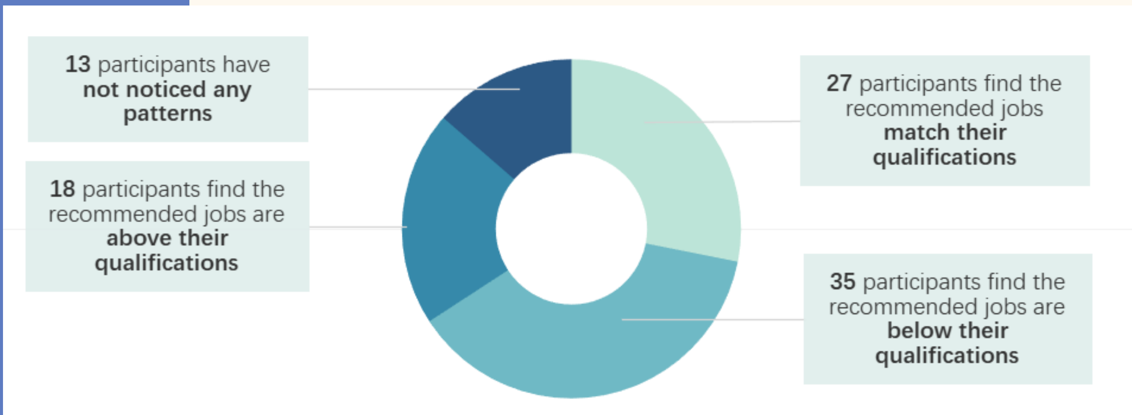


Graph 7: Lawyers' answers to "Do you feel that these hiring platforms recommend jobs that match your skills and expertise?"

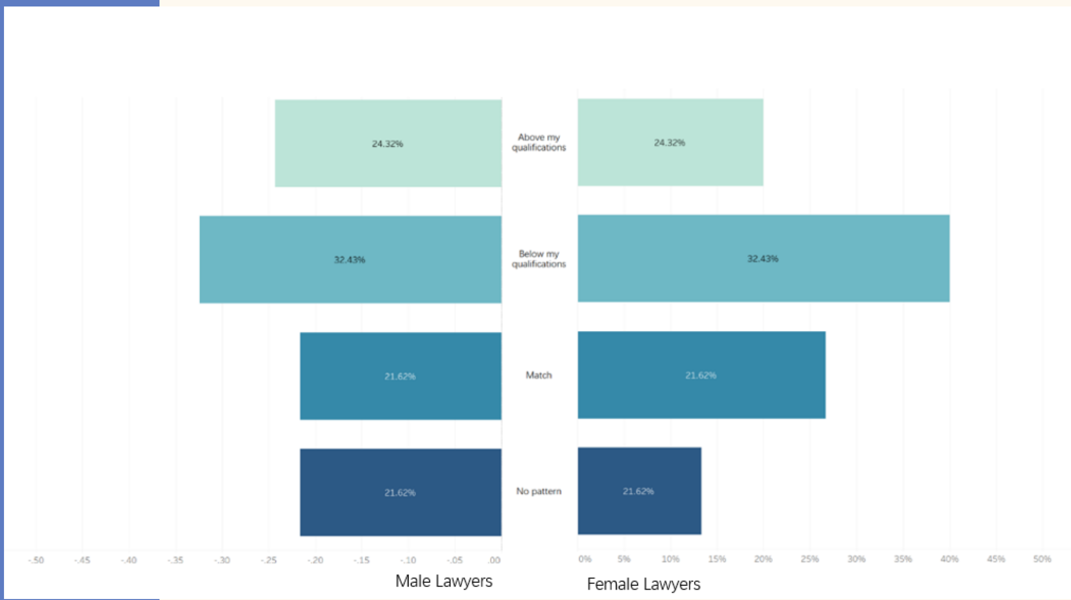
The data disclose that, compared with Chinese computer engineers, Chinese lawyers are less likely to use online hiring platforms. The fact that domestic law firms, together with China offices of international law firms, seldom pose job openings on hiring websites may partially account for this phenomenon. Moreover, compared with the United States, hiring platforms are not extensively relied upon by Chinese jobseekers[1], which may also explain the relatively low proportions of lawyers and computer engineers using the hiring platforms.

---

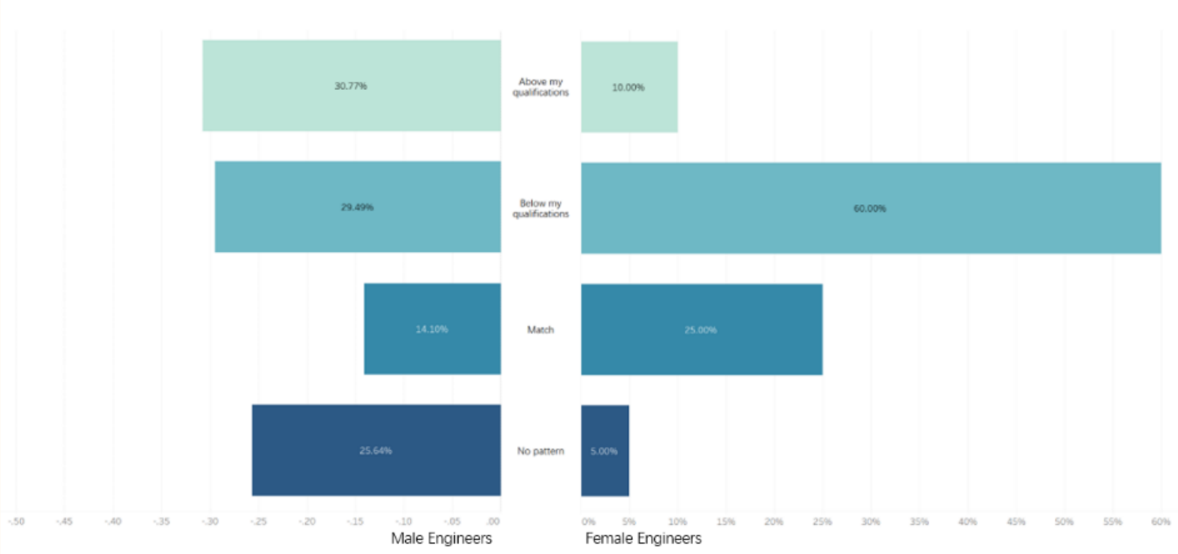
1 According to user statistics provided by Linkedin, there are more than 171 million Linkedin users from the United States, accounting for 52% of the entire population, while there are merely 5.397 million Linkedin users from China, accounting for 0.4% of China's entire population. See Statista (2021) Leading Countries Based on Linkedin Audience Size as of 2021 [Online]. Available at: https://www.statista.com/statistics/272783/linkedins-membership-worldwide-by-country/ (Accessed 25 March 2021).



Graph 8: Computer engineers' answers to "Do you feel that these hiring platforms recommend jobs that match your skills and expertise?"



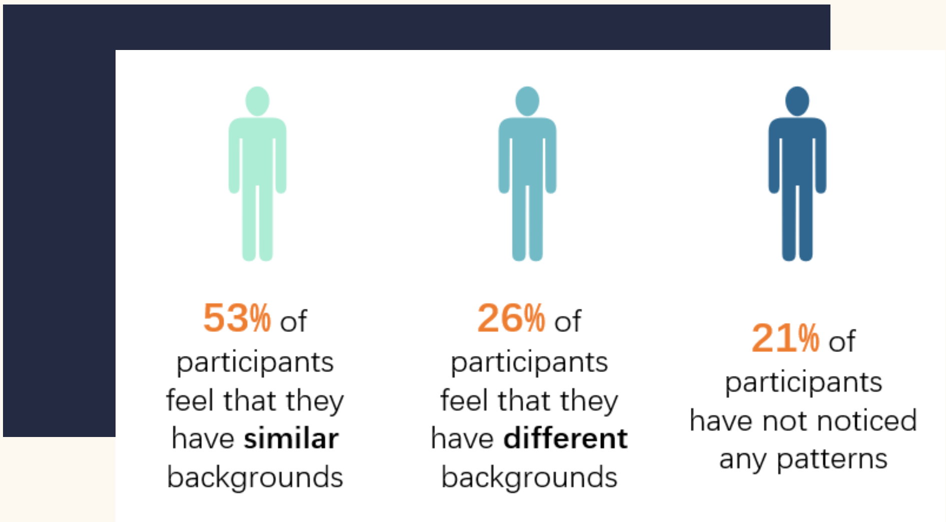
Graph 9: Male lawyer v. Female lawyers, "Do you feel that these hiring platforms recommend jobs that match your skills and expertise?"



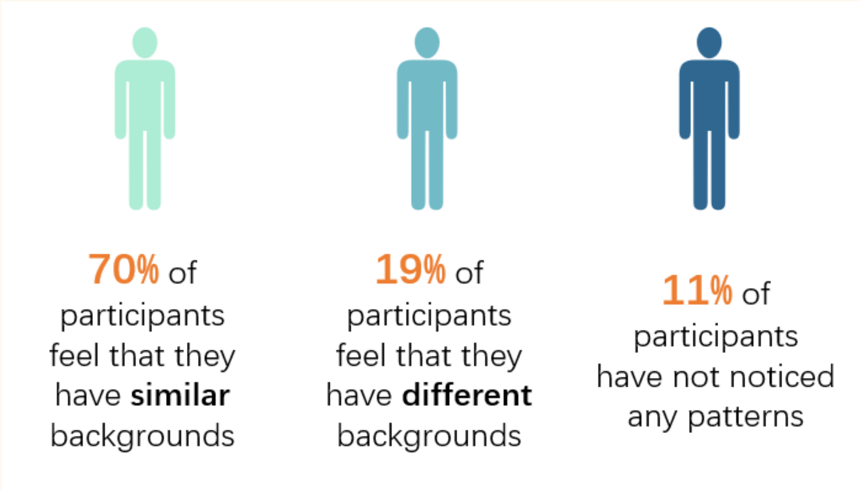
Graph 10: Male engineers v. Female engineers, “Do you feel that these hiring platforms recommend jobs that match your skills and expertise?”

From the perspective of professions, more than 45% of Chinese lawyers and 38% of Chinese computer engineers find the jobs recommended by hiring platforms are below their qualifications, and about 24% Chinese lawyers and 29% Chinese computer engineers find the recommended jobs match their qualifications, which appears to demonstrate that the hiring algorithms are selecting candidates slightly overqualified for the positions. Such mechanism can promise the employers more efficient use of recruitment budgets (Bogen, 2019).

However, when the data are compared through the lens of gender, they may tell a different story. As illustrated in the graphs above, regardless of professions, females are more likely than their male counterparts to be recommended with jobs that are below their qualifications. And such phenomenon is more remarkable with regard to female computer engineers than female lawyers. Correspondingly, male lawyers and computer engineers tend to encounter more jobs that are above their qualifications or match their qualifications and are less sensitive to the patterns of recommendation. This mechanism, thereby, may incur the problem that females are more likely to get stuck in jobs they are overqualified for.



Graph 11: Lawyers' answers to "Do you feel that these hiring platforms recommend connections with other professionals who have similar backgrounds to you?"



Graph 12: Computer engineers' answers to "Do you feel that these hiring platforms recommend connections with other professionals who have similar backgrounds to you?"

Most of the respondents tend to feel that the hiring platforms are recommending connections with other professionals who have similar backgrounds to them. And computer engineers seem to be more sensitive to such patterns compared with lawyers.

LIKERT SCALE ANALYSIS

The Likert scale questions can be divided into three dimensions, including the gender-related dimension (gender, age, marital status), the qualification dimension (education, work experience), and race. To begin with, the reliability and validity of the data are examined.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.966	.968	3

Table 1: Reliability test of the gender-related dimension, Lawyers

KMO aand Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy	.724
Bartlett's Test of Sphericity	
Approx. Chi-Square	267.205
df	3
Sig.	.000

Table 2: Validity test of the gender-related dimension, Lawyers

Since the Cronbach's Alpha is larger than 0.9, and the KMO is approaching 0.8, the data collected from the lawyers are reliable, and the gender-related dimension can adequately capture the characteristics of the respondents.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.822	.883	2

Table 3: Reliability test of the qualification dimension, Lawyers

KMO aand Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy	.750
Bartlett's Test of Sphericity	
Approx. Chi-Square	64.371
df	1
Sig.	.000

Table 4: Validity test of the qualification dimension, Lawyers

Similarly, the data collected from the lawyers are reliable, and the dimension can adequately capture the characteristics of the respondents.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.962	.963	3

Table 5: Reliability test of the gender-related dimension, Engineers

KMO aand Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy	.769
Bartlett's Test of Sphericity	
Approx. Chi-Square	352.242
df	3
Sig.	.000

Table 6: Validity test of the gender-related dimension, Engineers

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.924	.946	2

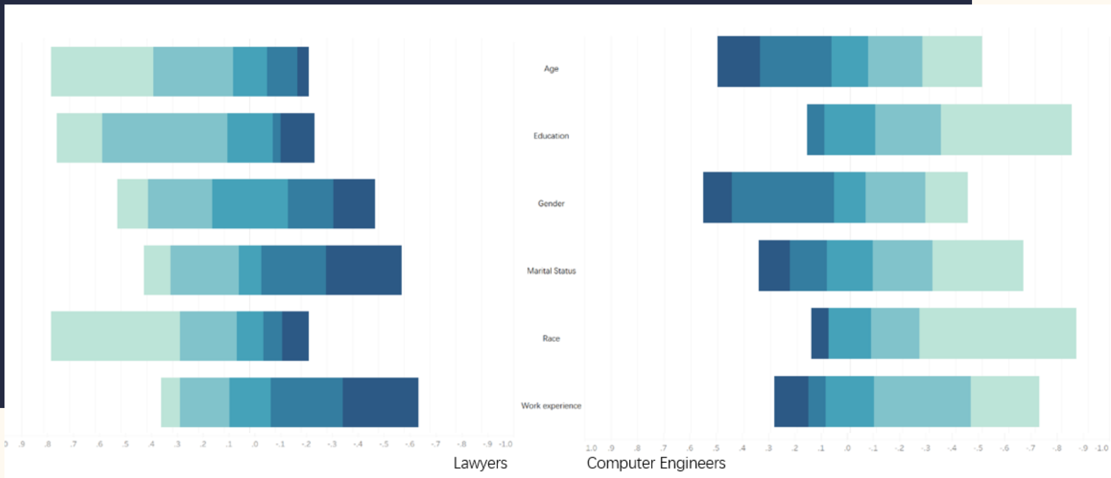
Table 7: Reliability test of the qualification dimension, Engineers

KMO aand Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy	.500
Bartlett's Test of Sphericity	
Approx. Chi-Square	147.543
df	1
Sig.	.000

Table 8: Validity test of the qualification dimension, Engineers

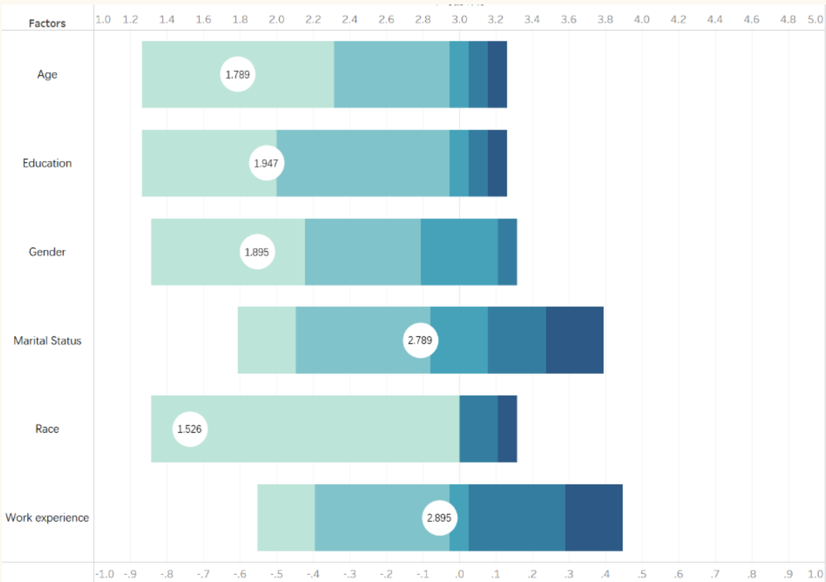
Therefore, all data collected through the Likert scale questions are reliable, and the three dimensions can adequately depict the sample.



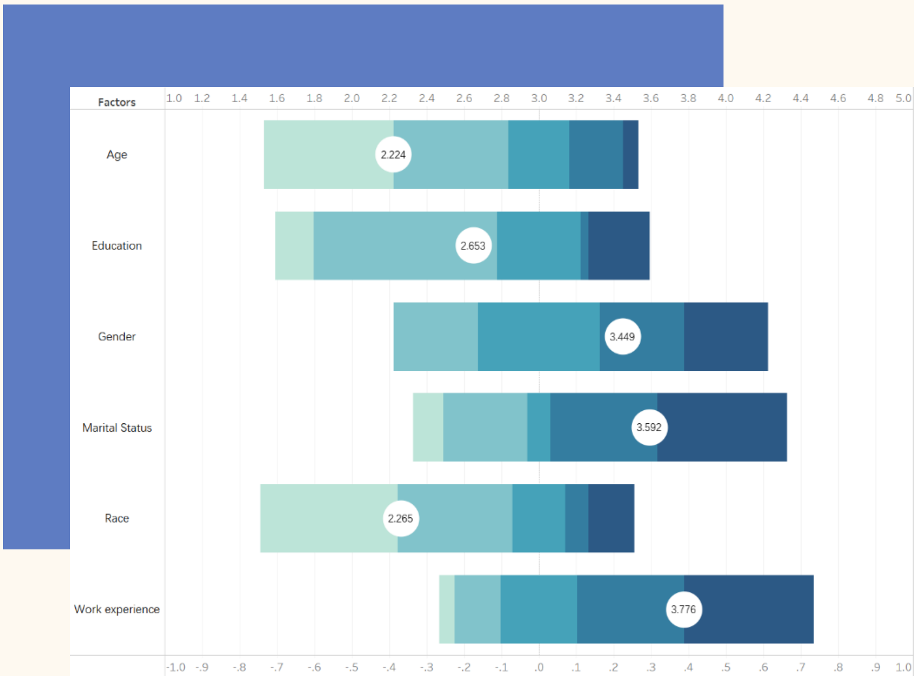
Graph 13: Comparison of the answers to the Likert scale questions between lawyers and engineers

The above graph demonstrates the frequencies each option appears in the answers, indicating how much the respondents worry that the employers using hiring algorithms might not consider them because of their gender, age, marital status, race, education, and work experience. According to the graph, computer engineers are more likely than lawyers to worry that the employers might not consider them because of their age, while the lawyers are more concerned about their gender, marital status, race, education, and work experience.

Chinese lawyers are most worried about, based on the graph, whether their work experience and marital status will have negative effects on the online hiring process. Meanwhile, Chinese computer engineers are more concerned about gender and age. Race and education, however, are less concerned about by the participants.

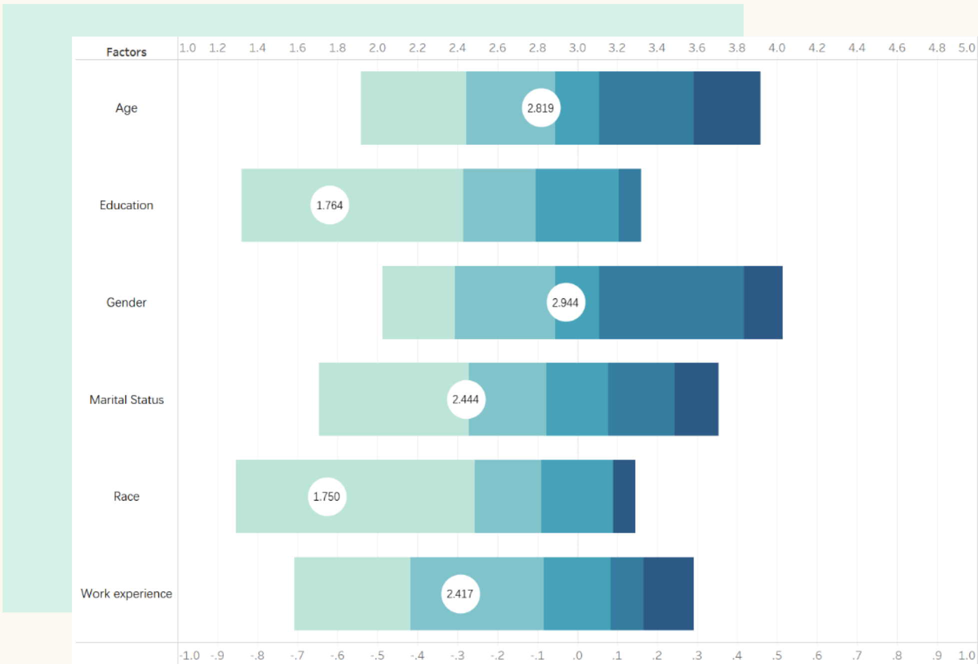


Graph 14: Male lawyers' answers to the Likert scale questions

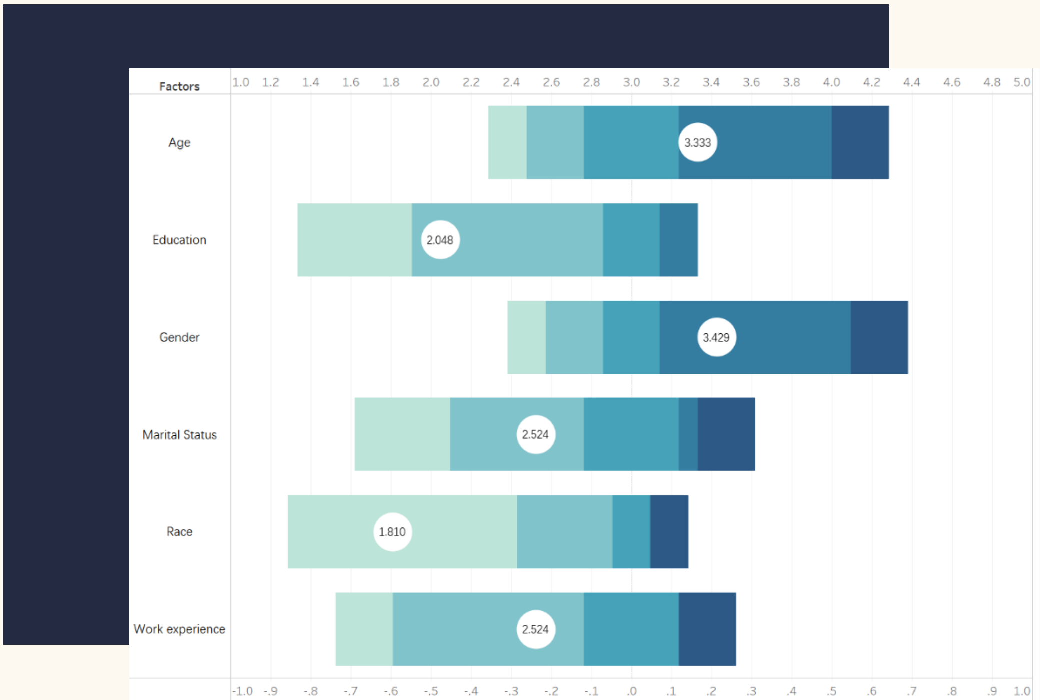


Graph 15: Female lawyers' answers to the Likert scale questions

The graphs illustrate that female lawyers in China tend to have a higher degree of worry, especially with regard to their marital status, work experience, and gender identity. Male lawyers in China, however, have merely slight concern about implicit gender bias underlying the hiring algorithms.



Graph 16: Male engineers' answers to the Likert scale questions



Graph 17: Female engineers' answers to the Likert scale questions

It is noteworthy that both male and female computer engineers tend to worry that the employers using hiring algorithms may not consider them due to their age and gender, which is different from that of lawyers. In the answers to the qualitative question asking the respondents to comment on bias they have encountered in the process of using hiring platforms, several male computer engineers commented that, while emphasis is frequently placed on gender bias against women in the field of computer engineering, men are also biased against. They complained that they have been rejected by the employers simply because the employers wanted to hire more female computer engineers "in order to balance the gender ratio" or "to improve the atmosphere of the office".

To conclude, female lawyers and computer engineers tend to have a higher degree of worry than their male counterparts pertaining to bias introduced by hiring algorithms. Among the three dimensions, the gender-related dimension, involving gender, age, and marital status, represents the major concern of female participants.

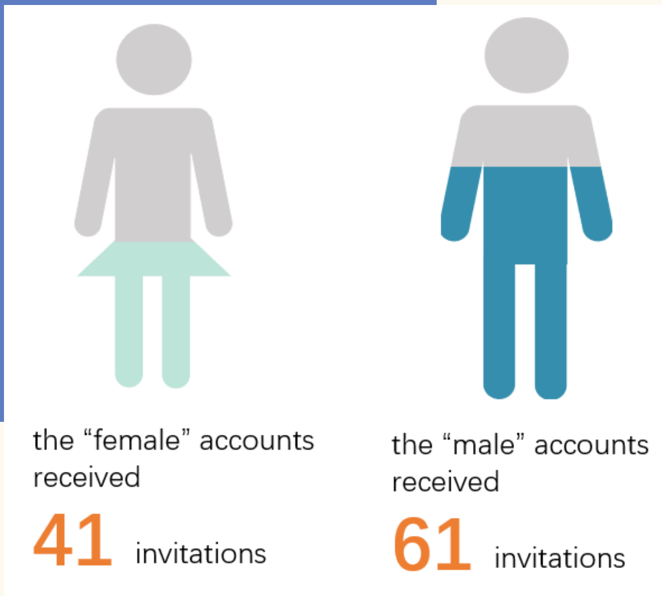
DOES THE MAINSTREAM HIRING PLATFORM CONTRIBUTE BIAS?

Pursuant to the case study conducted by Edelman and Luca, unintended consequences may be triggered by a seemingly routine mechanism of the algorithms (2014, p. 7). The author designed an experiment to test for the implicit gender bias against candidates in one of China's mainstream hiring platforms, BOSS Zhipin.

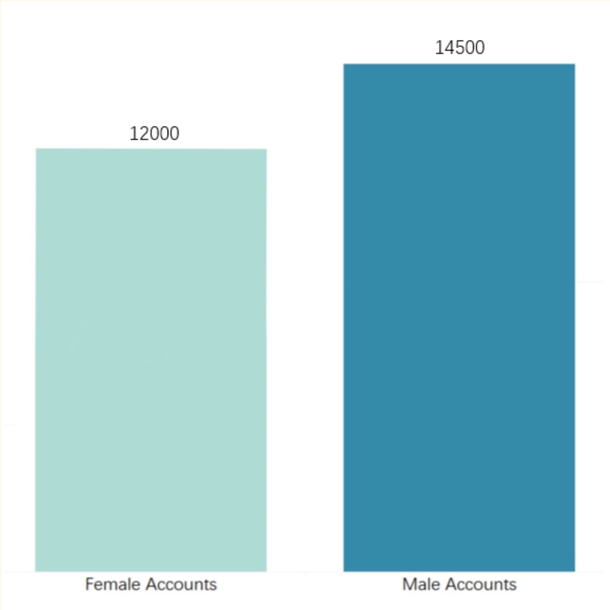
Similar with Linkedin, a candidate on BOSS Zhipin is asked to upload detailed personal information, including name, gender, birthday, education background, and work experience before searching for job openings. The candidate is also encouraged to upload an official resume, which will be automatically sent to the employer upon the candidate's consent to the employer's invitation. The predictive technologies, realized through algorithms, will recommend proper candidates for the employers and simultaneously recommend the positions for the candidates. During this stage, only limited information is available for the employers, including the candidate's name, gender, age, and education background, on the basis of which the employers can make the decision whether to invite the candidate or not. Only after the candidate's consent to the invitation, can both parties start a dialogue and communicate more detailed information.

The experiment adopted a control variable, the gender of the candidate, to test for bias against female or male jobseekers. The author uploaded identical information except for gender to the hiring platform, creating three "male" accounts and three "female" accounts, and evaluated the hiring algorithms through analyzing the number of invitations received by different accounts and the pay for each position.

The experiment lasted for five consecutive days, and the six accounts received in aggregate 102 invitations from the employers.



Graph 18: Invitations received by “male” and “female” accounts



Graph 19: Monthly pay for the positions

As a result, the “female” accounts received 41 invitations from the employers, and the average monthly pay for the positions offered was RMB 12,000. Meanwhile, the “male” accounts received 61 invitations, and the average monthly pay was RMB 14,500. Admittedly, the result cannot serve as confirmative evidence that the hiring algorithms adopted by BOSS Zhipin are discriminative against female due to the limited tenure and scale of the experiment. However, the result should alert the hiring platform, which is responsible for providing unbiased pre-employment assessments, that it shall consider regular audit of the algorithms and adopt proper approaches to de-biasing (Raghavan et al, 2019, p. 9). Furthermore, the result does provide insight into future study combatting bias incurred by AI. For example, according to Bogen (2019), the hiring algorithms may drift toward bias at different stage of hiring. Since the present experiment merely went through the stage of shaping the candidate pool, more future studies can be undertaken to explore how the algorithms can effectuate bias at each stage of hiring.

CONCLUSION

Based on the empirical study of more than 160 Chinese lawyers and computer engineers, this report discloses that female lawyers and computer engineers are more likely to feel that they are recommended with jobs below their qualifications and are more sensitive to bias underlying the hiring algorithms. A possible explanation for this phenomenon is that, since females are more likely to experience similar bias in the hiring process, they tend to be more familiar with, as well as pay more attention to the manifestations of unequal treatment. The experiment may also provide support for the notion that despite being designed to combat subjective biases, AI may instead amplify bias by default (Bogen, 2019). Therefore, the situation in China may provide useful insight into bending hiring algorithms toward equity in the global contexts.

References

[1] Bogen, M. (2019). All the Ways Hiring Algorithms Can Introduce Bias [Online]. Available at: https://hbr.org/2019/05/all-the-ways-hiring-algorithms-can-introduce-bias (Accessed: 25 March 2021).

[2] Collett, C. and Dillon, S. (2019). 'AI and Gender: Four Proposals for Future Research', Cambridge: The Leverhulme Centre for the Future of Intelligence.

[3] Edelman, B., and Luca, M. (2014). 'Digital Discrimination: The Case of Airbnb.com', Harvard Business School Working Paper, No. 14-054 [Online]. Available at: https://www.hbs.edu/faculty/Pages/item.aspx?num=46073 (Accessed: 25 March 2021).

[4] Raghavan, M., Barocasy, S., Kleinberg, J., and Levy, K. (2019). 'Mitigating Bias in Algorithmic Hiring: Evaluating Claims and Practices', Cornell University, arXiv:1906.09208 [Online]. Available at: https://arxiv.org/pdf/1906.09208.pdf (Accessed: 26 March 2021).

[5] Raso, F. A., Hilligoss, H., Krishnamurthy, V., Bavitz, C., and Kim, L. (2018). 'Artificial Intelligence & Human Rights: Opportunities & Risks', Berkman Klein Center, Research Publication No. 2018-6 [Online]. Available at: https://cyber.harvard.edu/publication/2018/artificial-intelligence-human-rights (Accessed: 24 March 2021).

EMERGING PROFESSIONALS SURVEY

Stereotype Threat and AI-Based Hiring Platforms

A PILOT STUDY BY UNIVERSITY OF PENNSYLVANIA LAW STUDENTS, AI BIAS POLICY LAB

Methodology

Between February and March of 2021, members of the AI & Bias Policy Lab interviewed 131 individuals ranging in age between 18 and 40. While each respondent was affiliated with the University of Pennsylvania in some capacity, results were received from nations across the world including India, China, Germany, Colombia, Kenya, and within the United States. Of the respondents, approximately 70 percent were female, two thirds were people of color, and six percent self-identified as differently abled.

In terms of material, respondents were presented with a 22-question, online survey that probed their accounts and perceptions of algorithmic bias on popular hiring and recruiting platforms (such as LinkedIn, ZipRecruiter, Indeed, etc.) within the past year. 96 percent of respondents had used one or more of the most popular hiring platforms (such as LinkedIn, ZipRecruiter, Indeed, etc.) within the past year.

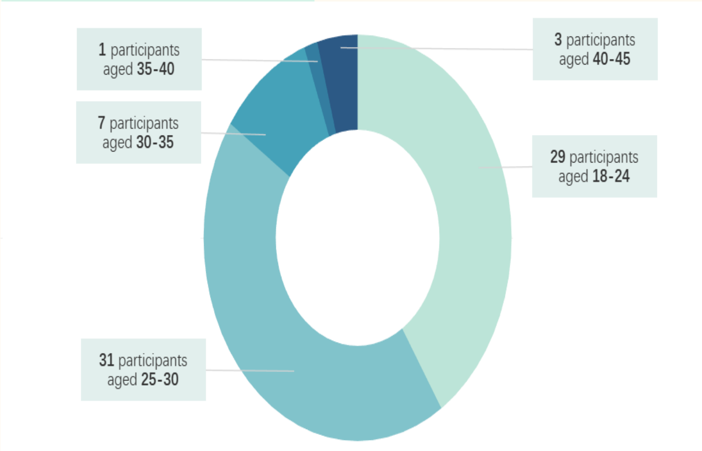
The results of the survey are presented in the section that follows. While these results failed to materialize particular instances of bias, our analysis has revealed an overwhelmingly pervasive and deeply concerning risk for stereotype threat within the platforms.

Stereotype threat refers to a situational predicament where one feels they are being negatively stereotyped based on a social identity such as age, religion, gender, or race. The perception of this judgement has been said to jeopardize an individual's identity and actions to follow and is particularly salient to our survey. In scenarios such as recruiting and hiring, the risk of stereotype threat becomes heightened as individuals believe their abilities are being measured and that the stereotype in question is pertinent to their ability to complete the task.

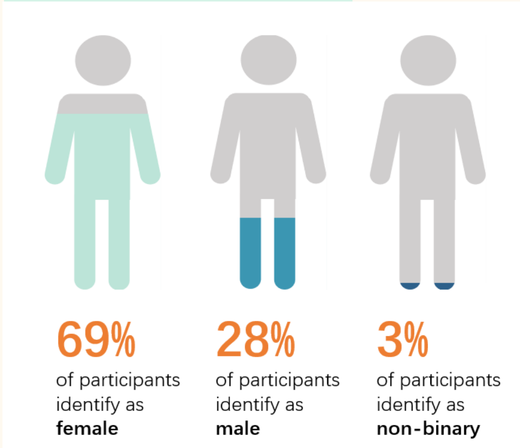
Demographics

The majority of respondents are studying law, are under 30 years old, and skew female, a grouping that may provide some color on results. However, the individuals do have diversity of background and the cohort largely represent a target demographic for online recruitment platforms.

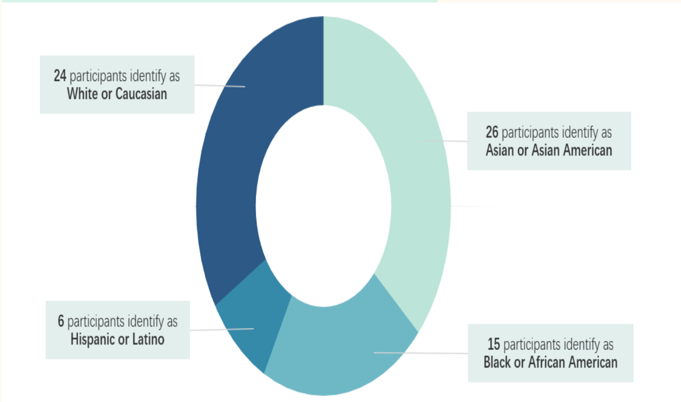
Q1 What is your age?



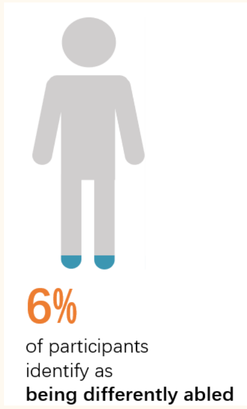
Q2 What is your gender identity?



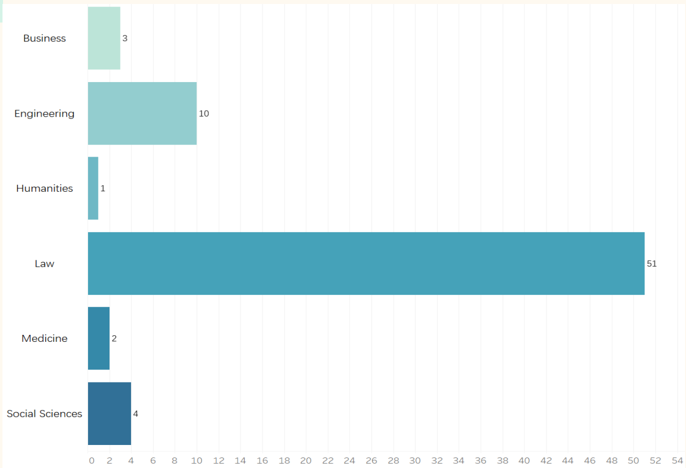
Q3 What is your racial identity?



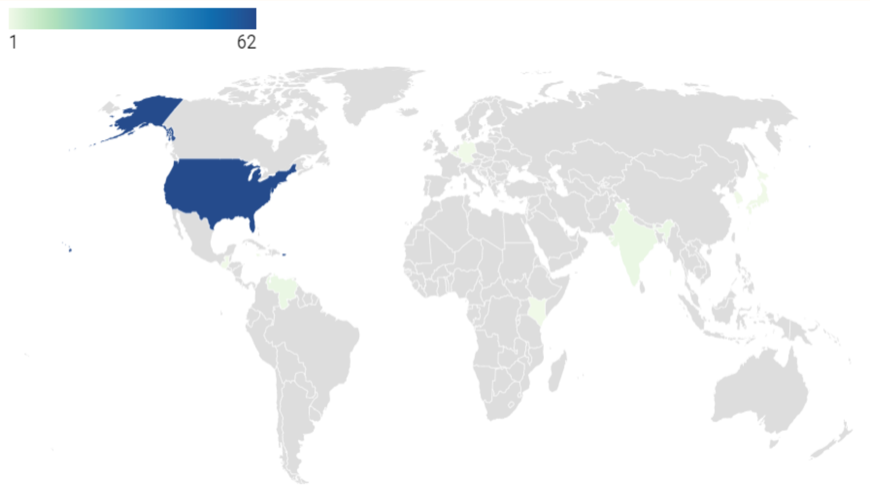
Q4 Do you identify as being differently abled?



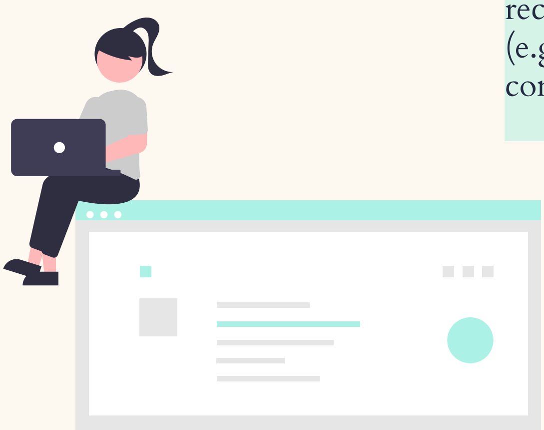
Q5 What field are you currently pursuing?



Q6 What is your home country?

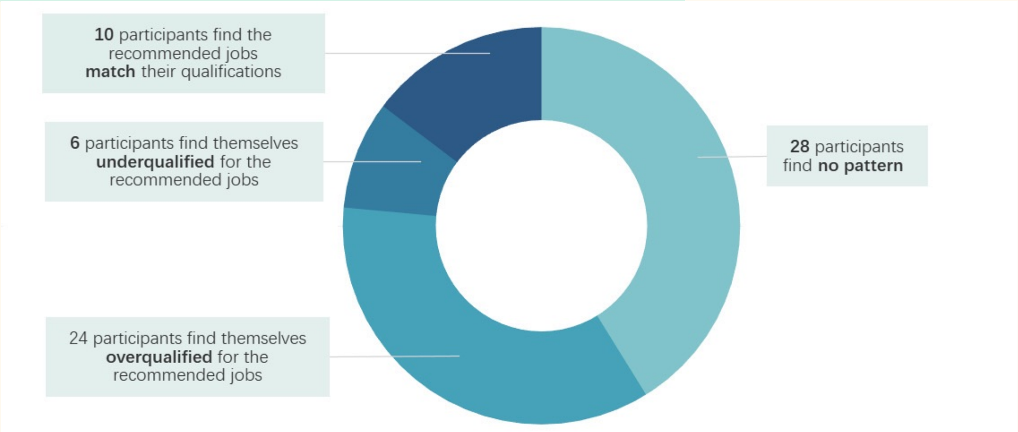


Survey Results



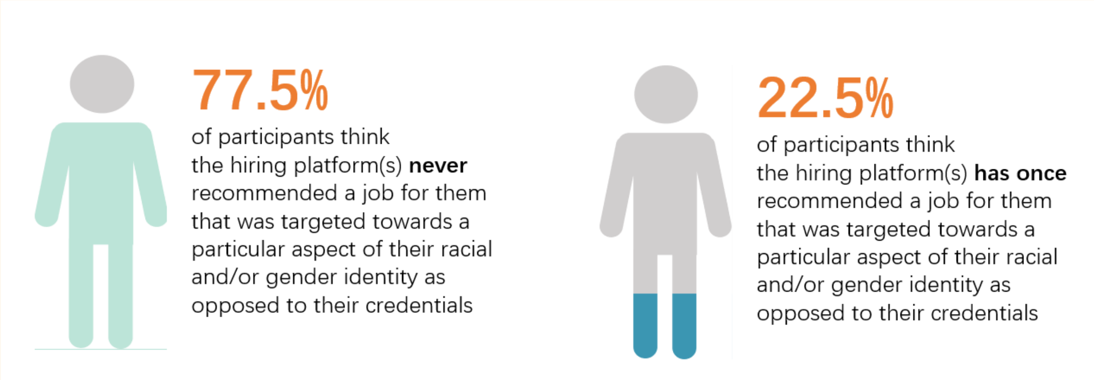
Q7 Have you used any hiring / recruiting websites in the past year? (e.g. LinkedIn, Indeed, Monster.com, ZipRecruiter, etc.)?

Q8 Do you feel that the hiring platform(s) that you use recommend jobs that match your skills and expertise?



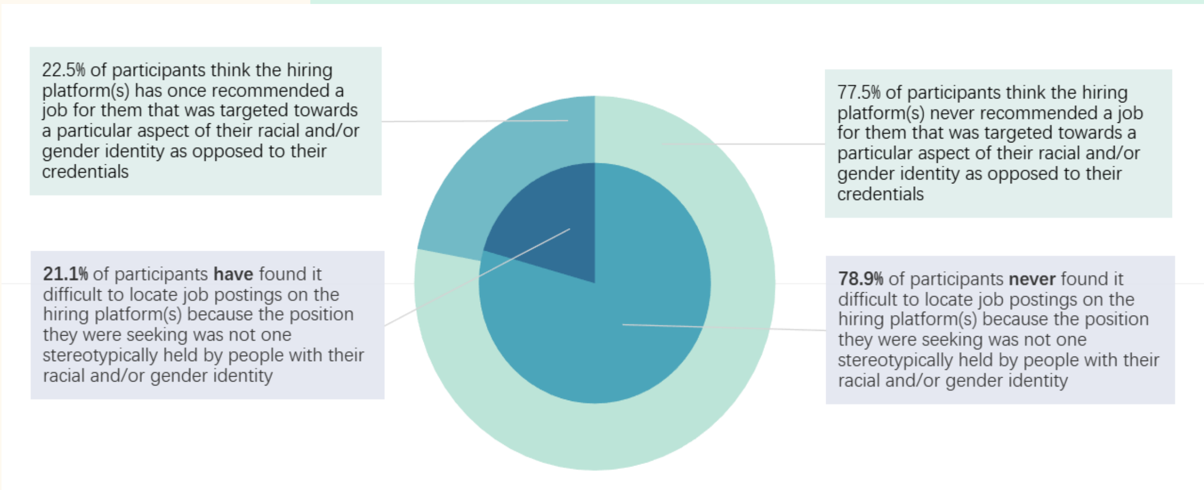
While the majority of those surveyed do not notice a pattern in job matching, those who have noticed a pattern largely feel they are overqualified for the positions recommended. This sentiment was also present in the survey of Chinese professionals, and may also be marketed as an "efficiency" in online recruiting. Unfortunately our data cannot speak more to this phenomenon in both studies.

Q9 Has the hiring platform(s) that you use ever recommended a job for you that you felt was targeted towards a particular aspect of your racial and/or gender identity as opposed to your credentials? (e.g. female medical school student getting recommended positions in nursing / home health care as opposed to doctors positions in hospitals)?



Most participants did not experience racial or gender bias through their searches nor in their recommendations, with over 77% never experiencing these biases. However, this leaves roughly 1 in 5 users with a noticeable experience of bias. Although the number of recommendations each user receives can be high, a single experience of noticeable bias might hold lasting effects.

Q10 Have you ever found it difficult to locate job postings, or receive job recommendations, on the hiring platform(s) that you use because the position you were seeking was not one stereotypically held by people with your racial and/or gender identity? (e.g. Asian female undergraduate student looking for jobs in the NFL but not getting job recommendations in the sports field).



Answer to this question is very similar to the previous one.

Q11 Do you feel that the hiring platform(s) that you use recommend connections with other professionals who have similar backgrounds to you in terms of credentials?

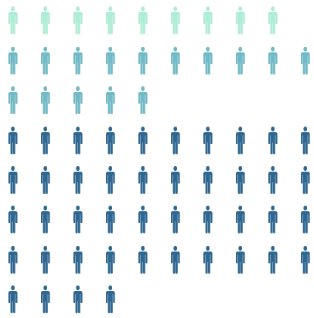
46 participants think that most of the time the connections recommended have similar credentials

8 participants think that most of the time the connections recommended are less credentialed

3 participants think that most of the time the connections recommended are more credentialed

11 participants have not noticed any patterns

Q12 Do you feel that the hiring platform(s) that you use recommends connections with other professionals who have similar backgrounds to you in terms of racial identity?

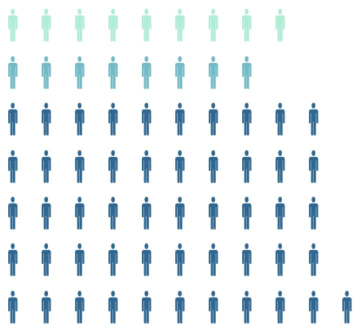


9 participants think that most of the time the connections recommended have a **similar** racial or ethnic background

15 participants think that most of the time the connections recommended have a **different** racial or ethnic background

44 participants have not noticed any patterns

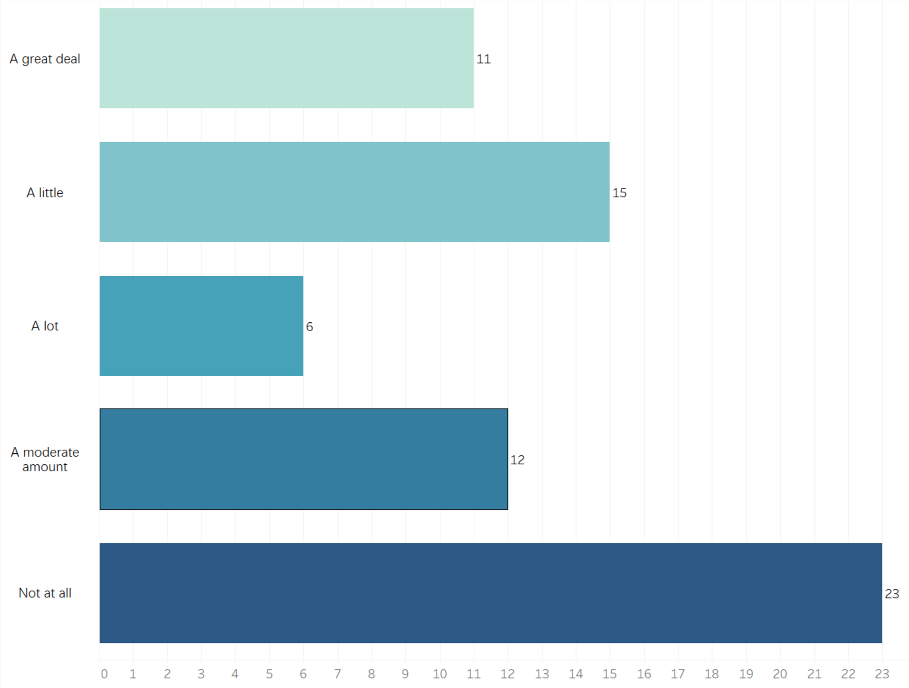
Q13 Do you feel that the hiring platform(s) that you use recommends connections with other professionals who have similar backgrounds to you in terms of gender identity?



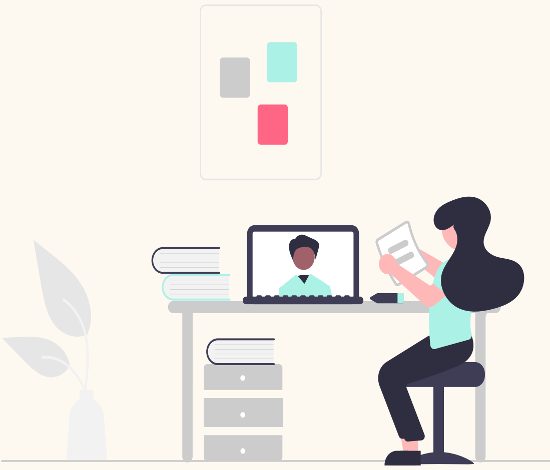
9 participants think that most of the time the connections recommended have the **same** gender identity

8 participants think that most of the time the connections recommended **do not** have the same gender identity

51 participants have not noticed any patterns

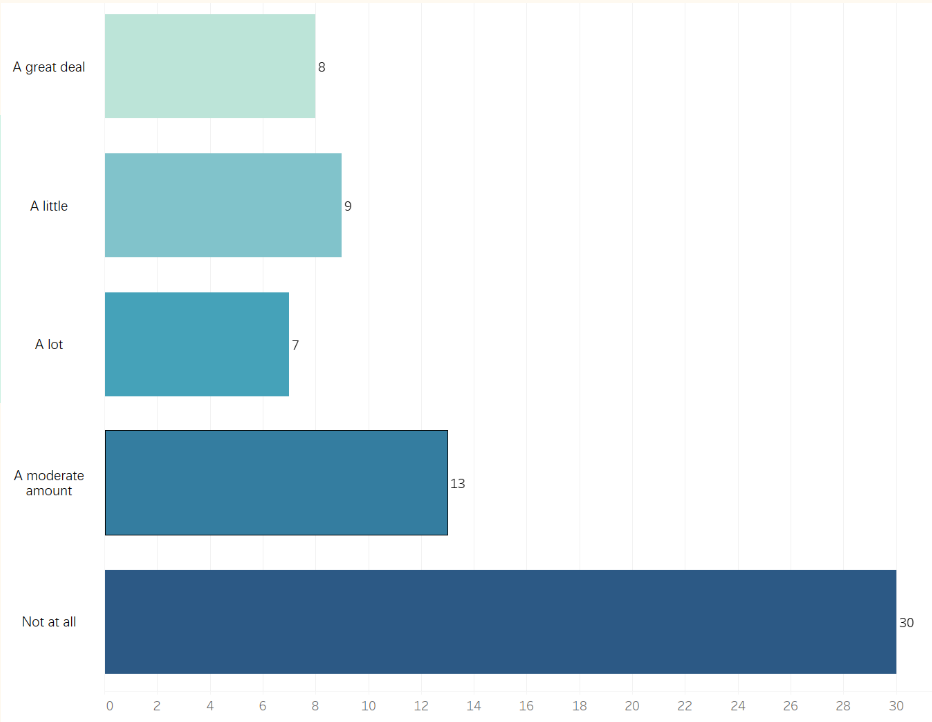


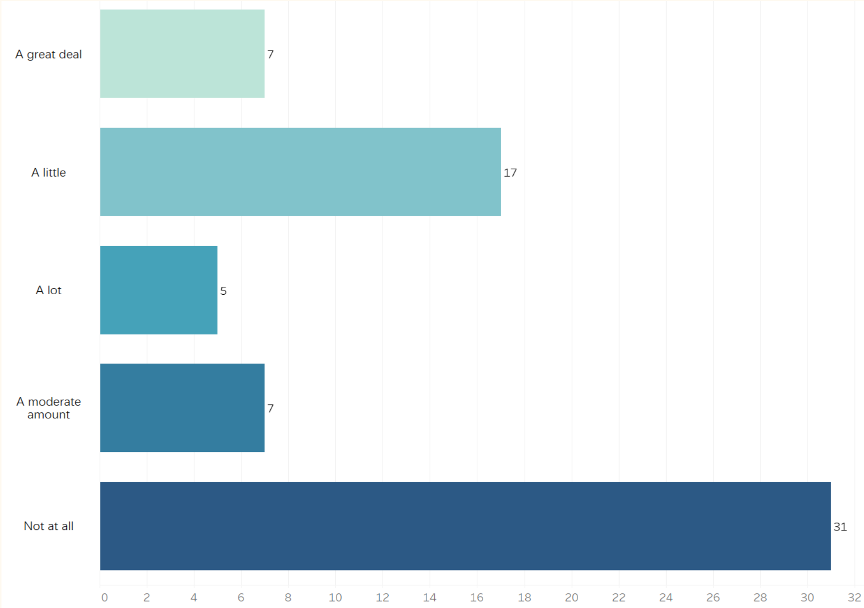
Q14 How much, if at all, do you worry that employers or managers using AI-based recruiting tools might not see your profile or consider you for a position because of the address listed on your resume or applicant profile?



The largest single groups of respondents are not worried that AI-based hiring systems will exclude their resume. However, the majority of the group thinks AI-based systems will either consider their geography as a factor (65%+) or their name as a factor (55%+).

Q15 How much, if at all, do you worry that employers or managers using AI-based hiring platforms might not consider you for a position because of your name?

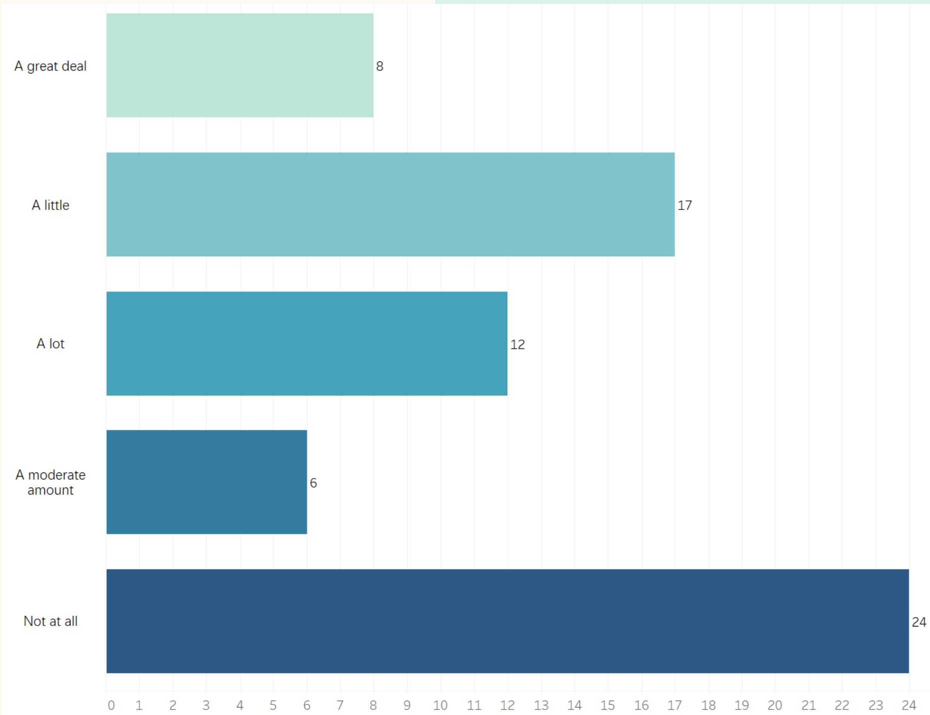




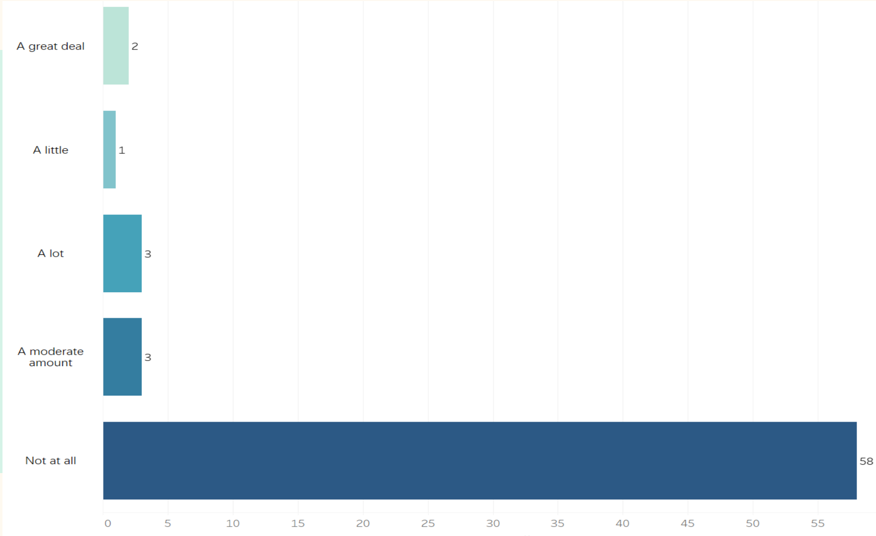
Q16 How much, if at all, do you worry that employers or managers using AI-based hiring platforms might not consider you for a position because of your gender identity?

Most respondents feel that gender identity (~53%) and racial identity (~64%) are factors contributing to AI-based hiring. The threat of stereotypes affect their perception of AI-based hiring.

Q17 How much, if at all, do you worry that employers or managers using AI-based hiring platforms might not consider you for a position because of your racial identity?

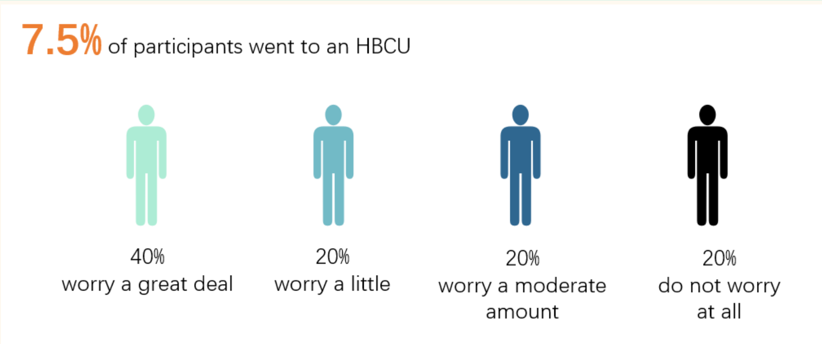


Q18 How much, if at all, do you worry that employers or managers using AI-based hiring platforms might not consider you for a position because you are differently abled or because features of your resume or applicant profile might indicate that you are differently abled?

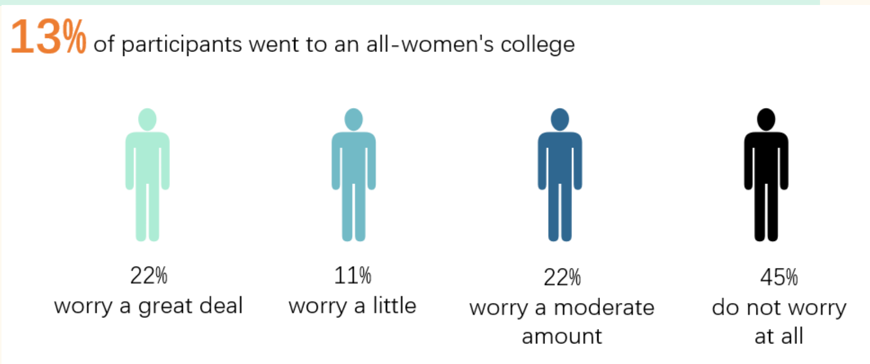


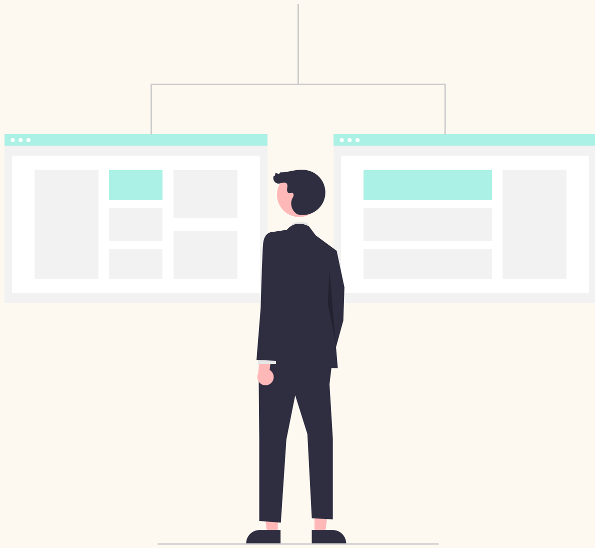
A small percentage of respondents attended a college that has an association with a racial identity (HBCU) or a gender identity (women's college). The majority of those who attended worry that AI-based hiring systems will exclude their candidacy based on their college's association irrespective of their own.

Q19 If you went to an HBCU, how much, if at all, do you worry that employers or managers using AI-based hiring platforms might not consider you for a position because of the education listed on your resume or applicant profile?



Q20 If you went to an all-women's college, how much, if at all, do you worry that employers or managers using AI-based hiring platforms might not consider you for a position because of the education listed on your resume or applicant profile?

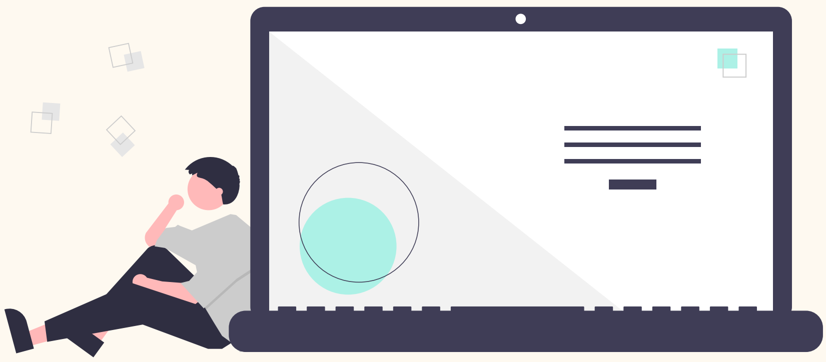




Q21 Have you observed any sort of bias in the hiring / recruiting process on hiring / recruiting sites? (e.g. a search for Stephanie Williams brings up results for Stephen Williams).

12 participants **have observed bias** in the hiring / recruiting process on hiring / recruiting sites

55 participants **have not observed bias** in the hiring / recruiting process on hiring / recruiting sites

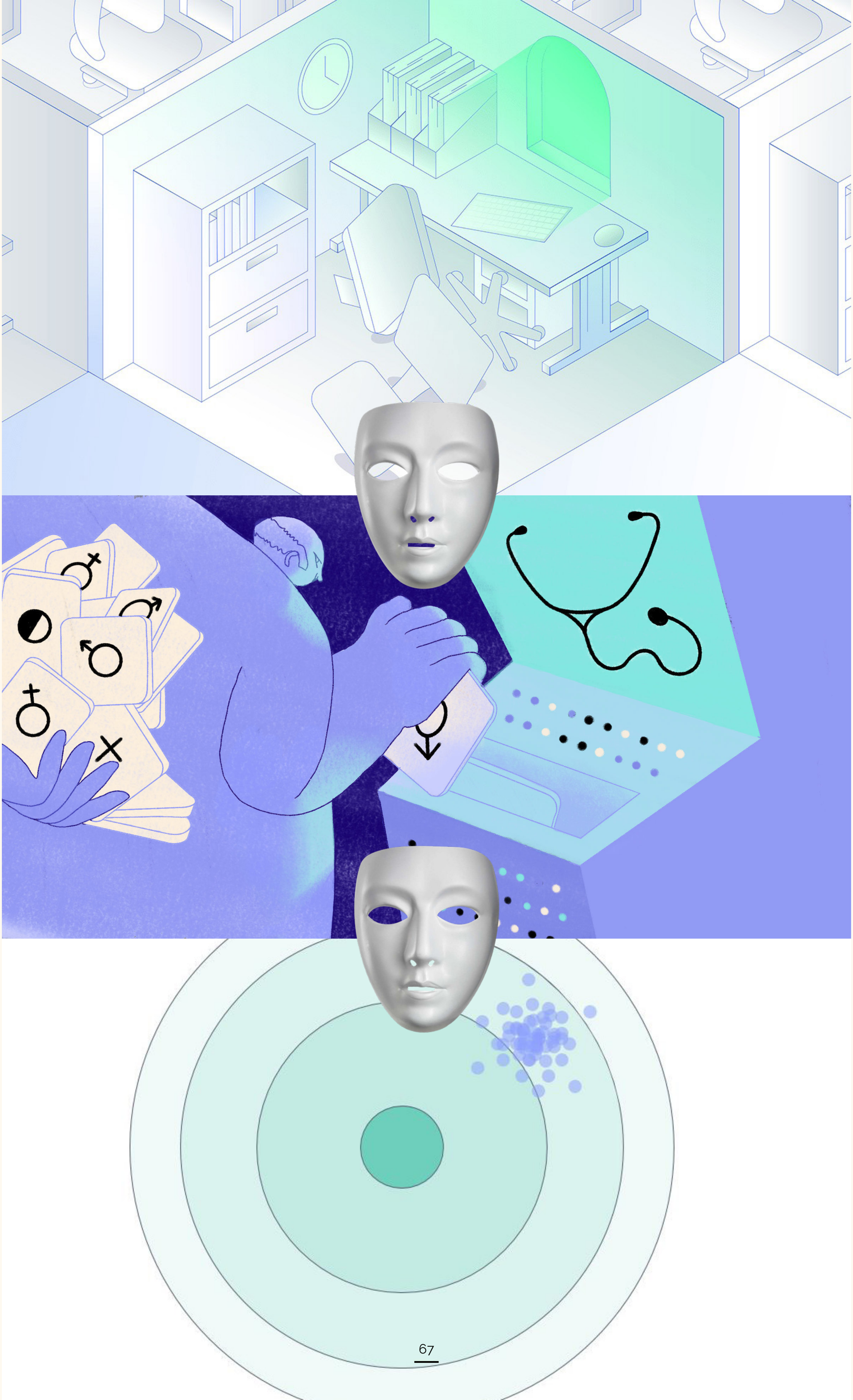


Q22 If yes, has your observation of bias on any of these hiring / recruiting platforms discourage you from using platform(s)?

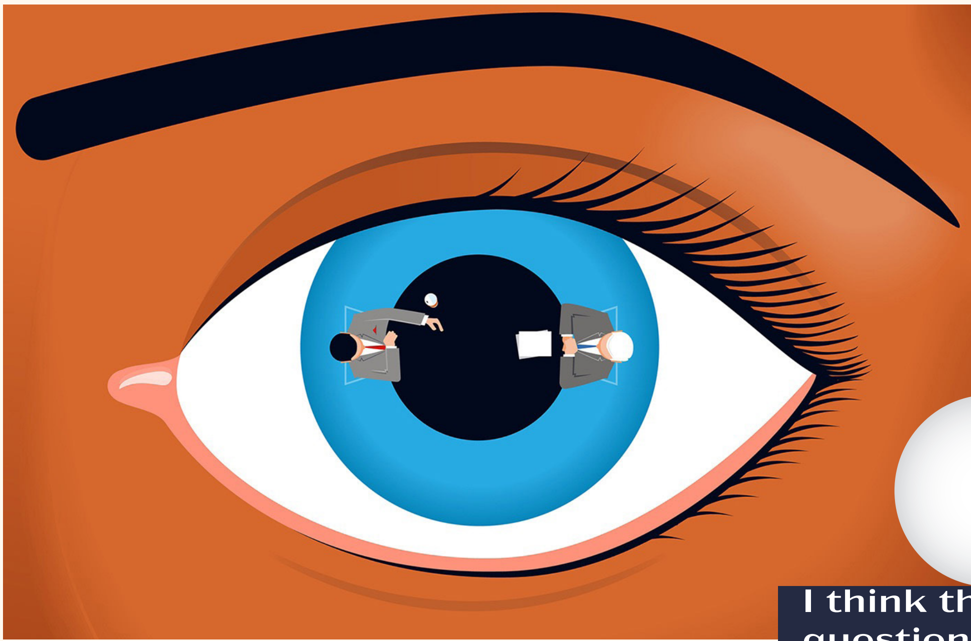
12 participants who have observed bias

7 participants **are** discouraged from using the platform(s)

5 participants **are not** discouraged from using the platform(s)



Respondent Narratives



I think the race question specifically should removed as a requirement on job applications, I hate this question and always put down other and write the word human.

PERSONAL QUESTIONS SUCH AS WHAT ARE MY PLANS IN TERMS OF EXPANDING MY FAMILY/HAVING BABY.

ALSO BEEN TOLD THE BEST PERSON FOR A PARTICULAR JOB POSITION IS FOR SOMEONE WHO DOESN'T HAVE KIDS.

OFFERED ABOUT $30 K LESS FOR THE SAME JOB POSITION AS MALE COUNTERPART.

When I was looking for jobs I would get recommendations for waitress, cart girl (at golf course), and other female-centered service jobs.

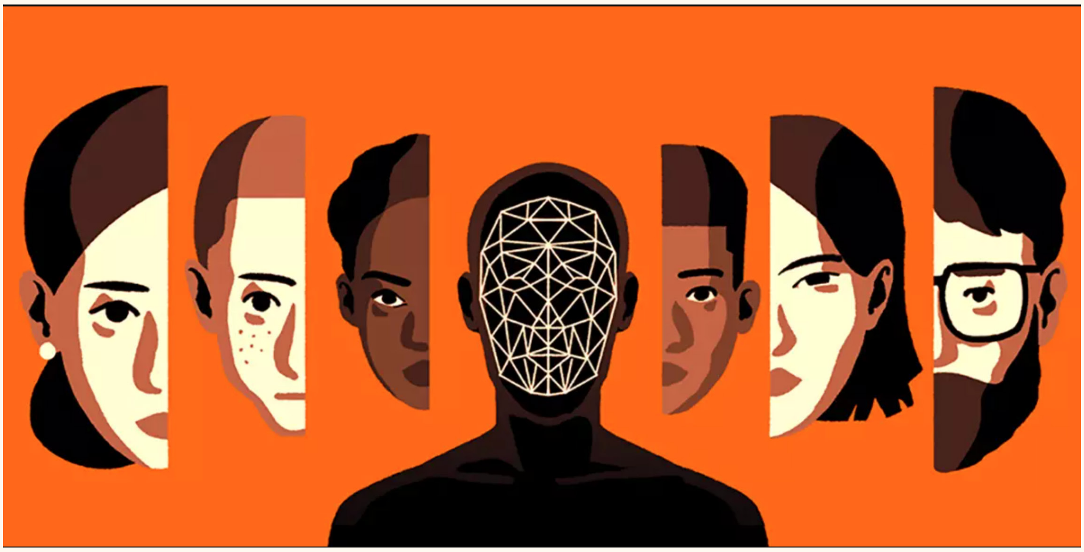
I do often worry that based upon my racial identity and gender identity, when viewing my LinkedIn, employers might not consider my application. I avoid attaching the link to my LinkedIn to minimize any biases an employer may have.

I’m not sure how the Indeed.com recommendation algorithm works, but it seems to suggest job openings that contain keywords matching those on my resume. These positions, for the most part, are relevant to my credentials and desired career path. However, I have no way of determining how job applications are processed/filtered/sorted once submitted.

I HAVEN’T HAD A TAD OF EXPERIENCE WITH AI BASED RECRUITING SERVICES BUT I CAN’T HELP BUT WORRY THAT THE ODDS SEEM STACKED AGAINST ME GOING INTO THE RECRUITING WORLD DUE TO MY GENDER IDENTITY, RACE, AND BEING DIFFERENTLY ABLED.

“RACIAL BIAS IS IN BETWEEN THE LINES.”

My concern is a lack of optimization for clubs/organizations traditionally associated with the Black community. For example, leadership in a resume is desirable but most of my leadership positions were in organizations specifically designed for Black students. Similarly, the sports that I played were not nearly as common for white folks as they were for Black folks. If the AI’s concept of a “good” resume is built using references to white resumes, folks who played golf or field hockey would have an advantage over those of us who preferred track or basketball.



Utilizing AI-based hiring/recruiting platforms in our contemporary time period is a multi-faceted experience. Many companies assert that they are focused on diversity, equity, and inclusion, and even more so after the "racial reckoning" of Summer 2020. In that regards, I have increasingly felt that companies are engaging in direct outreach to underrepresented candidates and are eager to see them apply and support them successfully through the process. At the same time, I understand that I come from a position of privilege as a law student with an Ivy League, T-14 law degree. This intersection - of both "elite" academic qualification and underrepresentation in the legal field has led to challenges in terms of recognizing and finding mentors who look like me, but at the same time has opened doors at companies and with mentors who are receptive to supporting those that may not look the same and are diverse. The multiple facets of this highlight the importance of hiring managers and C-level executives that will explicitly and frequently push for accountability in a company's DEI practices. I can only feel that my qualifications and experience are recognized, if companies actually recognize them and that requires stakeholder investment.

Discussion

The Role of Stereotype Threat

By Lindsay Holcomb

Overwhelmingly, the results of the survey seem to indicate the presence of stereotype threat, a psychological phenomenon defined as "a situational predicament in which individuals are at risk, by dint of their actions or behaviors, of confirming negative stereotypes about their group."[1] Stereotype threat tends to contribute to low performance or anxiety about low performance in historically marginalized groups who are aware of stereotypes impugning their intellectual ability. For example, women who are conscious of negative stereotypes questioning their ability to succeed in math and science, may fear confirming this stereotype in situations where they are required to display these skills.[2] Older adults who are conscious of negative stereotypes challenging their memory may become anxious or upset in situations where they are required to demonstrate their ability to recollect certain facts or ideas.[3] The fear of stereotype confirmation is most heightened in situations where individuals feel that they are being evaluated and group differences in performance are magnified.[4] In that sense, the hiring and recruiting process provide fertile ground for stereotype threat to manifest.

The strand of stereotype threat most pertinent to the hiring and recruiting process is what is known as "Own-Reputation Threat," which psychologists describe as "the fear of stereotypic characterization in the eyes of others – the fear of being judged or treated poorly by others because they may see one as negatively stereotypical."[5] This type of stereotype threat occurs where an applicant believes that there is an audience to her representation of her abilities and that that audience is judging her based on whether she belongs in, or represents, the stereotyped group. It is presumed that the audience is a member of an outgroup – a male, in the case of a female applicant, or a white person in the case of a black applicant – such that poor performance would reinforce, in another's mind, the negative stereotypes he or she might harbor about that group.

Consider, for example, the experiences shared by some of the black respondents to the survey. Of the respondents that attended a historically black college or university, 40 percent worried "a great deal" that employers or managers using AI-based hiring platforms might not consider them for a position because of the education listed on their resume or applicant profile. 80 percent of HBCU alumni reported that they worried about this at least "a little bit." This is a distressing result given that alumni status at an HBCU is a point of pride for many individuals – something that is central to their identity – but it is also more visibly linked to a stereotyped group than alumni status from a non-HBCU institution. The unfortunate output of such stereotype threat is that applicants may feel concerned or fearful about representing themselves authentically to recruiters and hiring managers. As one respondent commented, "I do often worry that based upon my racial identity and gender identity, when viewing my LinkedIn, employers might not consider my application. I avoid attaching the link to my LinkedIn to

1 MICHAEL INZLICHT & TONI SCHMADER, STEREOTYPE THREAT: THEORY, PROCESS, AND APPLICATION 5 (2011).

2 C. LOGEL ET AL., THREATENING GENDER AND RACE: DIFFERENT MANIFESTATIONS OF STEREOTYPE THREAT, ch. 10 (2011).

3 A.L. CHASTEEN ET AL., AGING AND STEREOTYPE THREAT: DEVELOPMENT, PROCESS, AND INTERVENTIONS, ch. 13 (2011).

4 INZLICHT & SCHMADER supra note 1.

5 Id. at 75.

minimize any biases an employer may have." Concern with how outgroup members might interpret one's abilities based on one's institutional affiliations can yield troubling behavioral outcomes ranging from failing to attach one's LinkedIn profile to a job application to failing to apply to the job altogether.

Stereotype threat also has the pernicious effect of making job applicants from stereotyped groups feel that they are judged based on their identity rather than based on their credentials. This experience is summed up powerfully in the words of former First Lady Michelle Obama who wrote of her time at Princeton: "No matter how liberal and open-minded some of my white professors and classmates try to be toward me...it often seems as if, to them, I will always be Black first and a student second."[6] Many of the survey's respondents seemed to feel similarly; nearly a quarter of reported believing that the hiring platform they used recommended a job for them that was targeted towards a pa particular aspect of their racial and/or gender identity as opposed to their credentials. As one respondent commented, "When I was looking for jobs, I would get recommendations for waitress, cart girl (at golf course), and other female-centered service jobs." Another respondent felt that she was viewed by hiring algorithms as a mother before a law school candidate. "I've received a lot of personal questions such as what are my plans in terms of expanding my family/ having baby," she explained. This gave her the sense that, "The best person for a particular job position is for someone who doesn't have kids." Such perceptions of "otherness" may make job applicants from stereotyped become more closely attuned to subtle cues about who belongs and who does not.[7] Where organizations are particularly focused on evaluations – testing candidates' performance and abilities as the sole metrics of success – stereotype threat can produce underperformance, potentially pushing candidates out of a particular career path they might have otherwise enjoyed.

Before even landing the interview, however, stereotype threat might affect the job opportunities with which candidates from stereotyped groups are presented. Consider, for example, the experiences shared on the survey by recent job seekers. A fifth of respondents reported finding it difficult to locate job postings on the hiring platform because the position they were seeking was not one stereotypically held by people with their racial and/or gender identity. This is particularly distressing when one considers that there are more than 11 million jobs posted on the major AI-based hiring platforms at any given moment.[8] If applicants feel that their opportunities are being narrowed as a result of their demographic factors, their aspirations may be threatened. Countless studies have shown this to be the case outside of the hiring realm. For example, a 1984 study found that exposure to television commercials depicting women in traditional roles led women to emphasize homemaking roles over professional achievement in describing their future lives.[9] And a 2005 study found that women who viewed such commercials were less likely to choose a leadership role in a subsequent task.[10] Thus, if candidates are only shown job postings that are stereotypically correlated with their race or gender, they may be less willing to embrace challenging positions and carve a new path. Rather, purely as a result of algorithms reproducing the biases of the offline world, they may self-handicap in the face of negative stereotypes.[11] Ultimately, eliminating identity-based discrepancies in job listings is crucial to reducing stereotype threat among eligible applicants from historically marginalized groups.

6 Michelle Robinson 2 (1985).

7 INZLICHT & SCHMADER supra note 1 at 92.

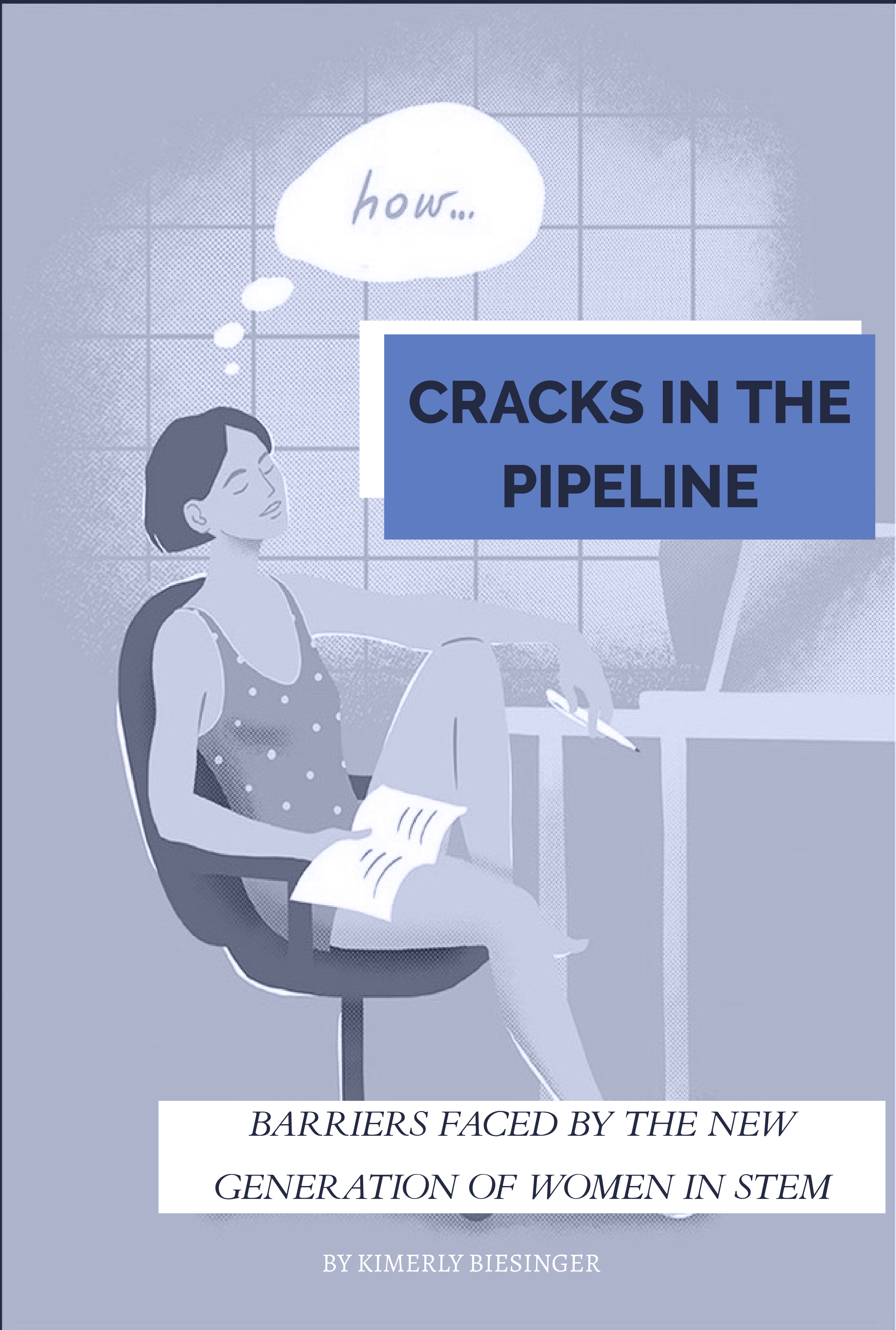
8 Craig Smith, LinkedIn Jobs Statistics (2021) DMR (Mar. 16, 2021) https://expandedramblings.com/index.php/linkedin-job-statistics/.

9 F.L. Brown et al., TV Commercials and Achievement Scripts for Women 10 SEX ROLES 513 (1984).

10 P.G. Davies et al., Clearing the Air: Identity Safety Moderates the Effects of Stereotype Threat on Women's Leadership Aspirations 88 J. OF PERSONALITY & SOC. PSYCH. 276 (2005).

11 INZLICHT & SCHMADER supra note 1 at 176

NEXT GEN SURVEY



CRACKS IN THE PIPELINE

BARRIERS FACED BY THE NEW GENERATION OF WOMEN IN STEM

BY KIMERLY BIESINGER

A survey of the next generation of women in STEM fields (students and young professionals under the age of 30) (n=30) suggests that the same biases that have kept women out of STEM fields in the past may still present barriers for young women today.

These biases disproportionally impact women with intersectional identities. In addition, the COVID-19 pandemic is uniquely affecting women in STEM fields.

Further research will provide further insight into the full extent of the impact of COVID-19 on women in STEM.

Introduction

Women have historically been underrepresented in science, technology, engineering, and mathematic fields due in part to deep-rooted biases within the tech industry that tend to push women out of their roles after they start their careers.[1] Women uniquely experience several major biases in the STEM industry including the following: prove-it-again bias where a woman is required to prove her professional worth repeatedly, tightrope bias where a woman is required to meet a delicate balance in behavior between masculine and feminine traits, and maternal wall bias where women who have or may have children are seen as less professionally competent. [2] Women with intersectional identities, particularly from ethnic minority groups, experience these biases to an even greater degree.[3]

Although efforts have been made to strengthen the pipeline, the same biases that have pushed women out of STEM fields in the past may still present barriers for young women today. Even with more women entering STEM fields, biases against these women act as cracks in the pipeline which dramatically weaken the number of women who stay in STEM careers. Until pressures against women in STEM fields are relieved, efforts to increase the number of women entering STEM industries can only be minimally effective.

The current COVID-19 pandemic has had unique and severe impacts upon women globally, and particularly upon women working in the medical field. Although post-pandemic rebuild can be an opportunity to build back better, it is likely that the pandemic has had a notable adverse impact upon women in STEM. Further research in the future will be able to provide insight into the full extent of the impact of COVID-19 upon women in STEM. Although this survey did not include a statistically significant sample, it does provide insight into some factors that may be influencing women in STEM which can help direct future research.

1 Williams, J. C., Phillips, K. W., & Hall, E. V. (2016). Tools for change: Boosting the retention of women in the stem pipeline. Journal of Research in Gender Studies, 6(1), 11. doi:10.22381/jrgs6120161

2 Tools for individuals: Bias interrupters. (2018, September 14). Retrieved April 02, 2021, from https://biasinterrupters.org/toolkits/individualtools/; Williams, J. C. (2015, March 24). The 5 Biases pushing women out of STEM. Retrieved from https://hbr.org/2015/03/the-5-biases-pushing-women-out-of-stem

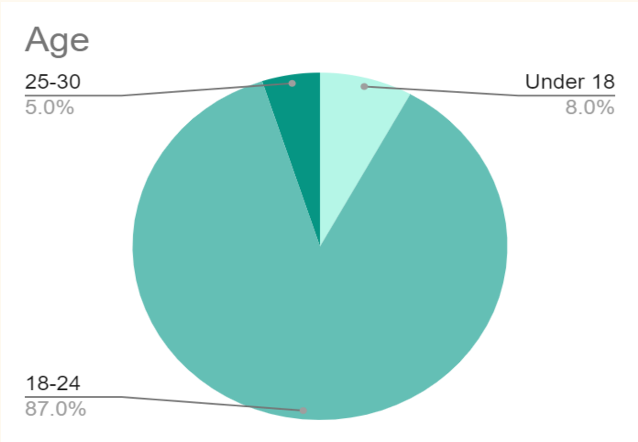
3 Williams, et al. supra note 1

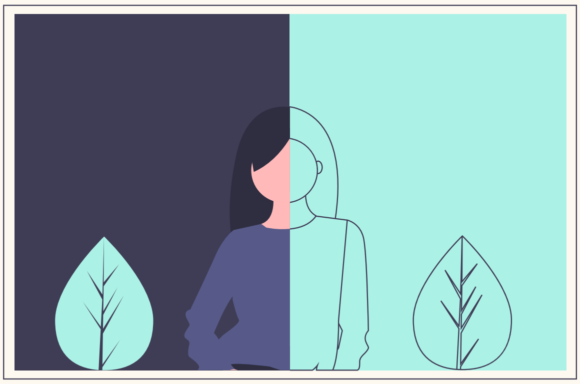
Methodology

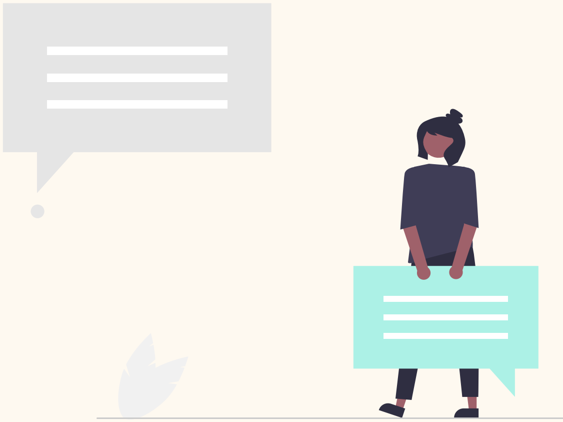
A small pool of women aged 18-30 who are students or young professionals in STEM fields were surveyed (n=28) in order to develop a high-level understanding of what biases are still being experienced by women beginning STEM careers today. All respondents except for one (from Kenya) were from the United States. The survey was broken up into five sections: demographics, experiences with bias, COVID-19 impacts, origins of STEM interest, and biases against women with intersectional identities. The section on experiences with bias contained questions based on the specified biases pushing women out of STEM. The COVID-19 section contained questions intended to understand some of the impacts being experienced by women in STEM due to the COVID-19 pandemic. The section on origins of STEM interest contained a question intended to understand what factors are most influential to sparking the interest of women and girls into STEM fields.

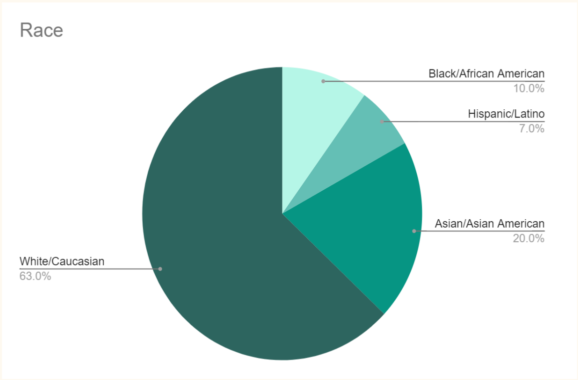
Of the survey respondents, 29% identified as racial minorities; only those individuals completed the section on intersectional identities. This section contained questions intended to understand how race influences the biases already experienced by women in STEM.

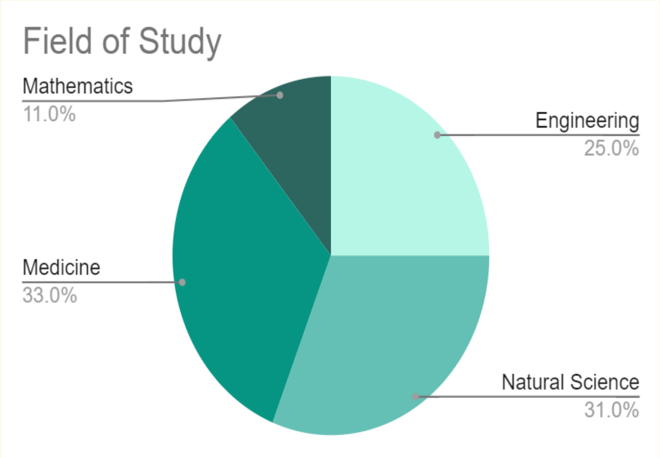
Demographic Data

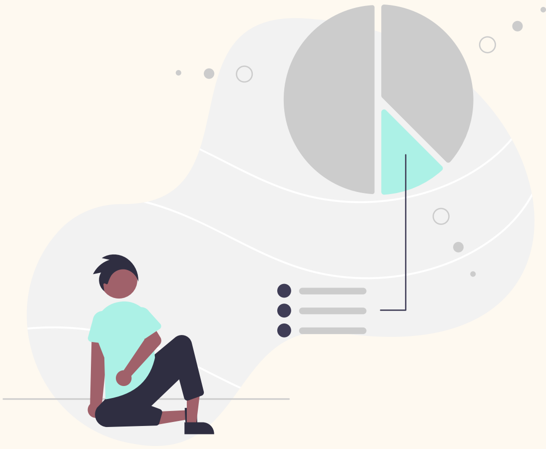








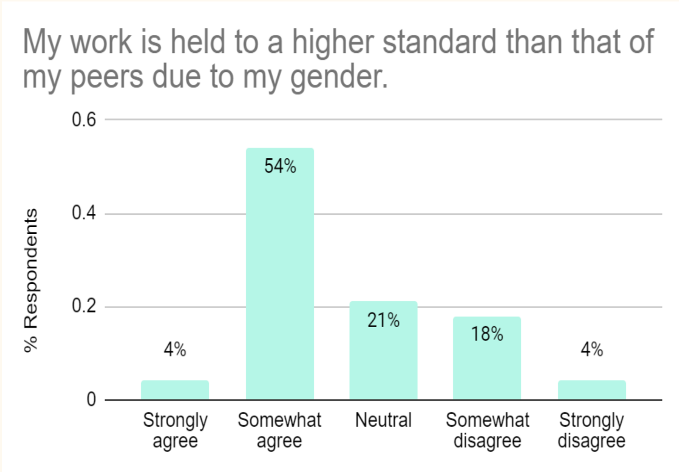




Findings

EXPERIENCE WITH BIAS

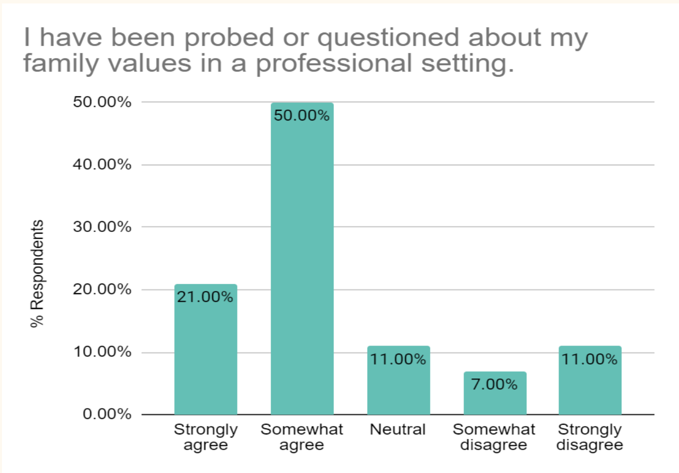
01



INSIGHT

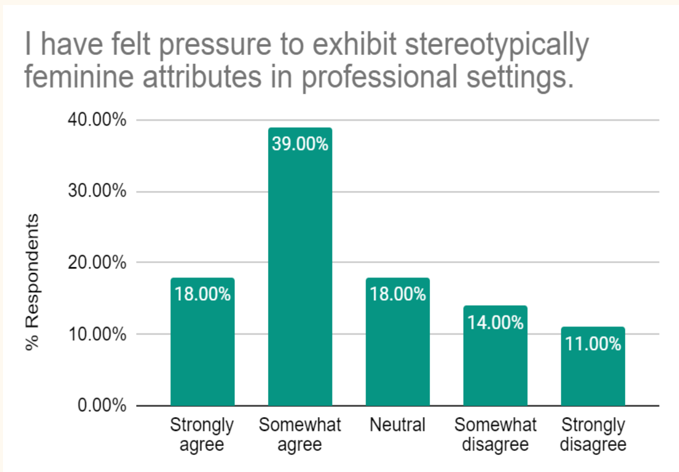
Prove-it-Again bias is a bias commonly experienced by women in STEM where a woman is required to work harder and prove herself more repeatedly than her male counterparts. She is held to a higher standard.

02

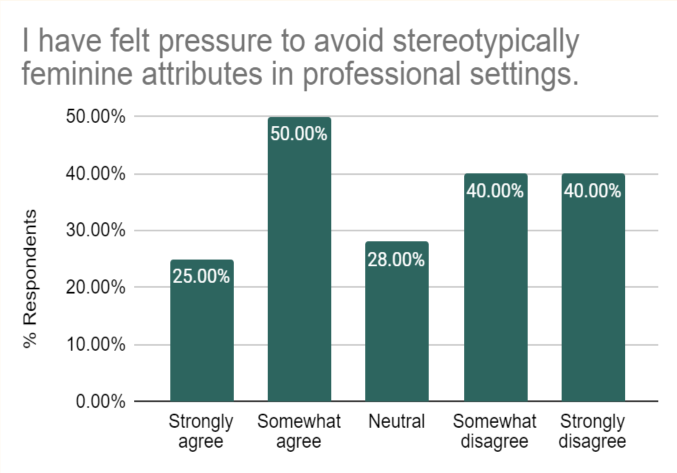


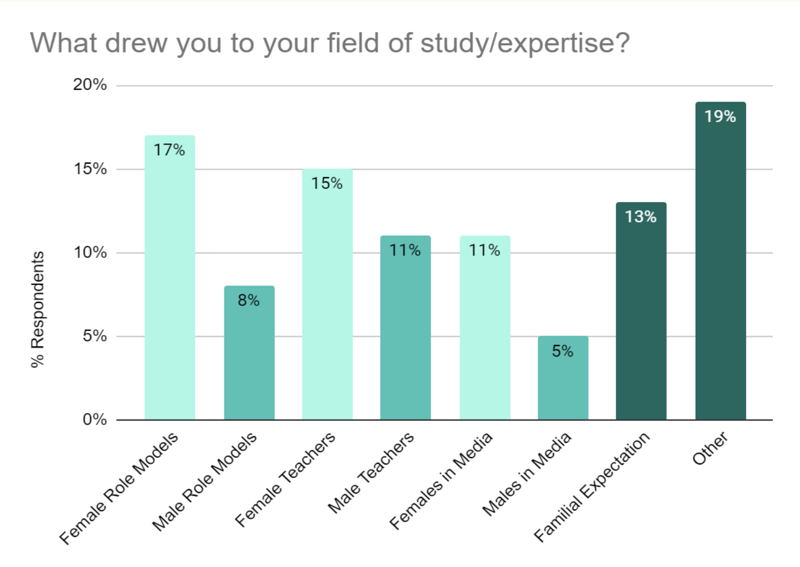
Maternal Wall bias is experienced by women whether she has children or not. Women are probed about familial values and status in interviews, and are treated as though they cannot be successful both in the workplace and the family.

03



Women in STEM are subject to Tightrope Bias where although they are expected to act in stereotypically feminine ways, they are simultaneously required to exhibit stereotypically masculine attributes to be taken seriously.





To grow the pipeline of women into STEM, it is useful to understand what factors spark women's interest in STEM. It appears that seeing other women in STEM fields may have stronger effect on young women than simply seeing men in those same fields. Of the responses labeled as "other," respondents mentioned personal interest, previous experience, and parental guidance.

Only 29% of survey respondents have a woman in their family who works in their field.

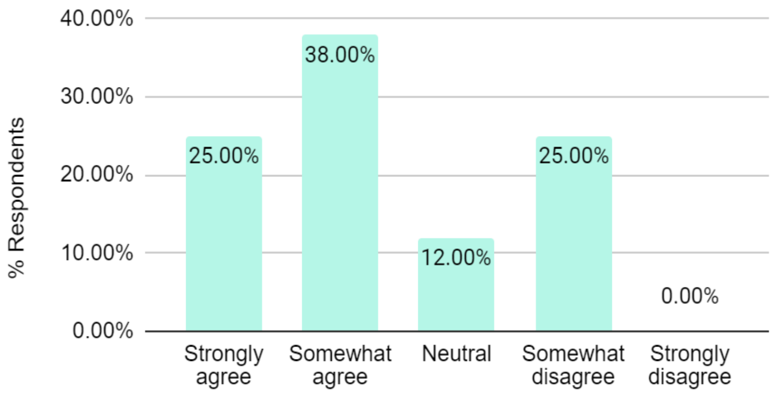
39% of survey respondents report having been mistaken for lower-level employees or students.

BIAS EXERIENCED BY WOMEN WITH INTERSECTIONAL IDENTITIES

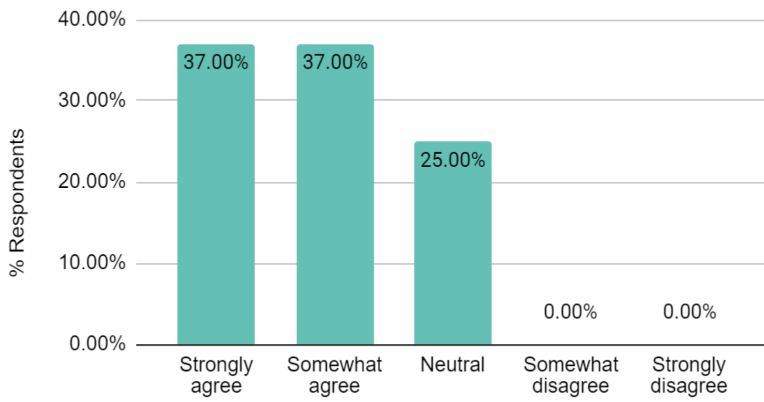
Of all survey respondents, 29% identified as racial minorities. Those individuals responded to these questions regarding their multiple identities

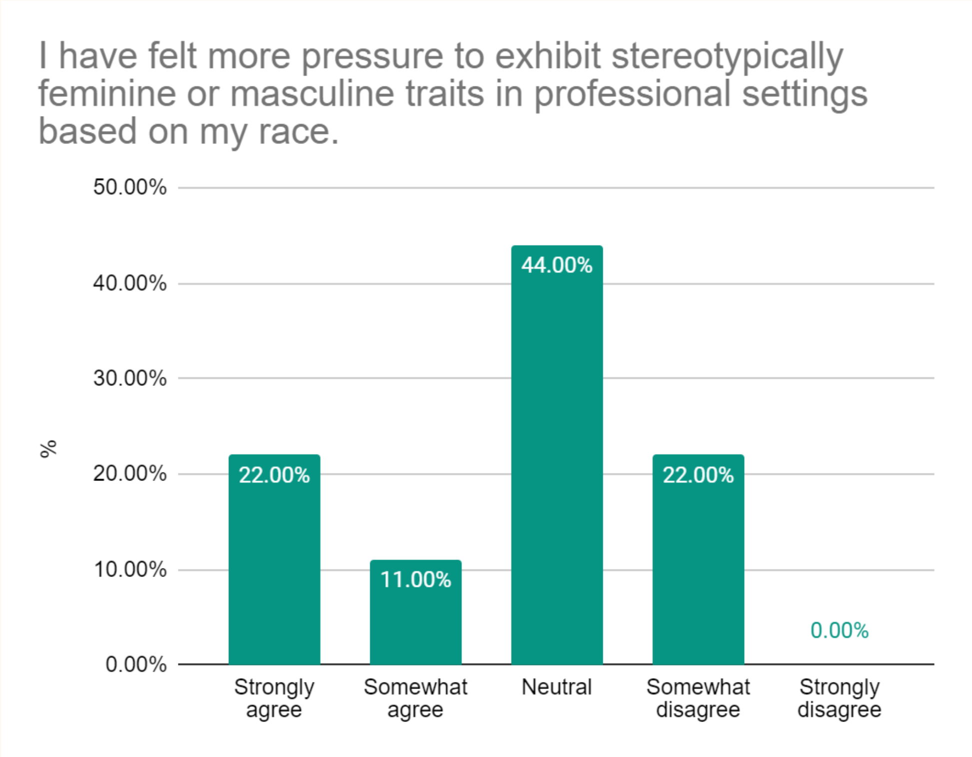
Although racial discrimination is not just a women's issue, women who identify with racial minority groups often experienced bias more acutely than men of their same race. Biases against women are often experienced more strongly by minority women.

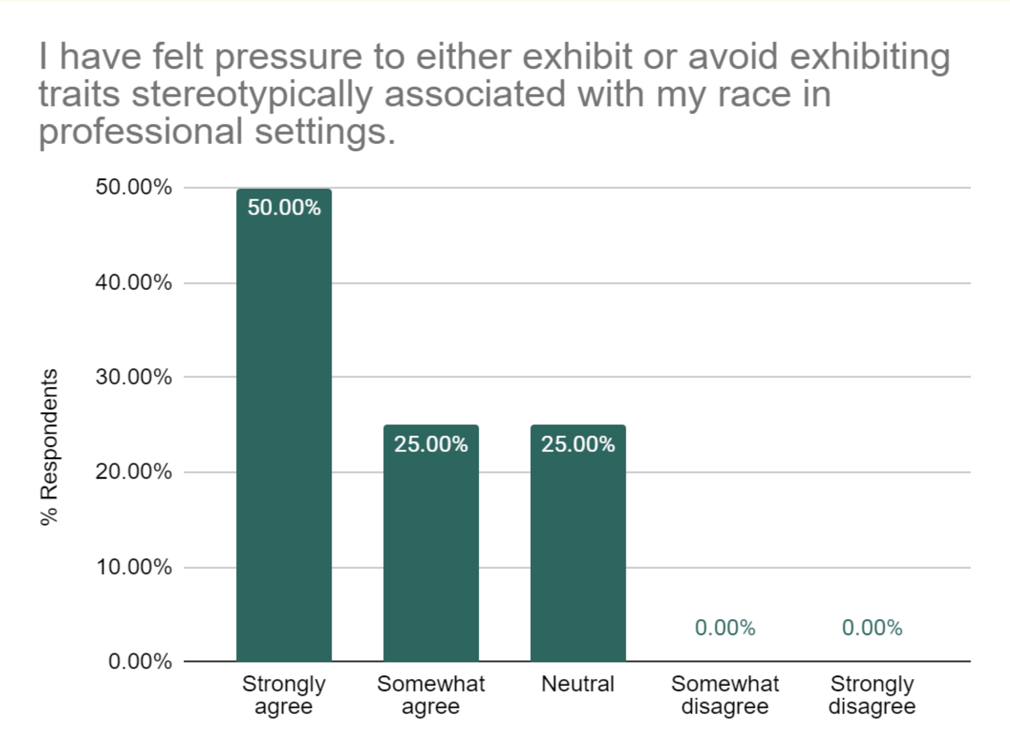
I have experienced or feared professional discrimination due to my race.



I have felt that my work must be held to a higher standard than that of my peers due to my race.







Pressure to act in certain ways based on stereotypes is often stronger for women who identify with racial minority groups. For example, black women may feel pressure to act in a more apathetic manner to avoid the "angry black woman" stereotype.

THE EFFECTS OF COVID-19

25% of survey respondents reported experiencing a shift in career priorities, ambitions, or aspirations due to the COVID-19 pandemic.

"I've been forced to relax my ambition due to anxiety around a lack of control."

"Had to take leave from nursing school, pursued alternate careers in medicine."

"I think because so much more is now online and at home, we have much higher standards for our work under larger time constraints. This idea that we can always work and always meet deadlines although we are struggling to stay afloat. Maintaining a healthy separation of work and home has been utterly impossible."

"My internship was cancelled so I pivoted from machine learning to software engineering."

18% of respondents report having experienced changes in their home relating to women in STEM careers due to the COVID-19 pandemic.

"Probably a lack of trust for medical professionals in general in my family"

"My Chinese mother has been denied promotional opportunities throughout all of COVID. She has worked for the state for almost 20 years. She is stressed about applying for new jobs, and I have been around to proofread her material, but she is being marginalized time and time again. She's a civil engineer."

"It seems that mental health is deteriorating due to COVID and for me as a woman, when I am struggling, I feel like I am stereotyped for being emotional or overly sensitive when I am feeling overwhelmed. Male colleagues do not seem to appreciate or understand."

"My family seems to really respect my career choice (biomedical research) especially now since a family member passed from COVID."

EVIDENCE-BASED PRACTICE GUIDE

EVIDENCE BASED PRACTICE GUIDE

TOOLBOX FOR MITIGATING BIAS

By Kimerly Biesinger, Chandra Nukala

CHALLENGE	SOLUTION
Individuals with certain protected attributes experience disparate hiring outcomes.	Consider protected attributes when building models but produce models that do not take protected attributes as input.

Outcome disparities are prevalent against protected groups due to human bias as well as bias built into AI hiring systems. Although normalizing models by race, gender, or other attributes could help mitigate these disparities, using information about these protected attributes while making hiring decisions should be avoided. Hiring models trained with data sets including these attributes can make equitable decisions without requiring the input of a candidate's protected identities.

CHALLENGE	SOLUTION
Use of AI hiring tools leads to increased bias that goes unnoticed due to lack of transparency into decision-making factors.	The exact pieces of information used to select or non-select a candidate should be made available to both candidates and third-party auditors whether decisions are made by an algorithm or by a human hirer.

The use of AI in hiring can be useful for mitigating bias, but also allows many biases to go unnoticed in the complex decision-making pathways used by AI. AI tools in particular make connections differently than human decision makers, and may use irrelevant attributes or inferences to make hiring decisions. Often these other attributes stand as a sort of proxy for protected attributes that should not be used in hiring decisions. Transparency in these decision networks will allow for candidates and auditors to clearly understand what biases are present in hiring systems.

CHALLENGE	SOLUTION
Hiring tests and processes are often created in such a way that better ranks members of certain groups regardless of actual ability.	Test hiring tests and processes on all groups prior to implementation to prevent exclusive hiring practices. Audit these processes regularly for bias.

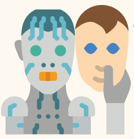
Differential validity takes place when the test results of certain hiring tests are more valid for some groups than other groups. These effects are often invisible but introduce a barrier to members of protected groups who are excluded from performing well and being hired. Hiring processes, whether executed by AI or human hirers, that have been tested on candidates of all identities are less likely to contain biases.

CHALLENGE	SOLUTION
Diversity policies and practices are not sufficiently followed.	Self-audits, third-party audits, and data collection should be conducted regularly surrounding all hiring practices and policies.

The establishment of hiring policies is not enough to ensure compliance and fairness in hiring. Without effective measurement and auditing, investments into these causes are less effective, and hirers are not held accountable. Regular data collection and auditing ensures that systems and practices are achieving their purpose. Audits should include review of candidate diversity to recognize bias in candidate sourcing, ethnic diversity of candidates by hiring stage to recognize stages that contain more bias, and relationship between interviewer/screener identity to the proportion of diversity in candidates hired or progressed to monitor and mitigate stereotype threat.

CHALLENGE	SOLUTION
Lack of direct accountability for implementation, adherence, and continued improvement of hiring practices inhibits progress of diverse hiring practices.	Appoint a committee or employee responsible for overseeing the implementation and measurement of equitable hiring strategies and their progress.

Without responsibility for equitable hiring efforts explicitly assigned to an individual or group, accountability for the effectiveness of these practices is easily lost. A designated entity can develop expertise necessary for further development and ensures compliance and measurement of change. Companies who hire a full-time diversity staff member see substantial increase of chances for minority individuals to be hired into management positions.



CHALLENGE	SOLUTION(S)
A candidate's name and photo can be used, even indirectly, to discriminate against individuals with certain protected identities.	Remove name and photo of candidate from data being considered for at least the first-round hiring processes.

Both human hirers and AI hiring tools can discriminate against individuals with certain names or appearances which can have a negative effect on individuals with underrepresented identities. Removing this sensitive information from applications in the first rounds of consideration can allow for more objective hiring decisions made and more equal hiring patterns.

CHALLENGE	SOLUTION(S)
Individuals with underrepresented or protected identities are often rejected or written off before they have a chance to demonstrate their unique perspective and value.	Implement the practice of seriously considering all candidates with protected minority identities.

Candidates with protected identities are frequently rejected from positions before having been seriously considered, often due to differences in background or experience connected to their identity. Companies should take an approach based loosely upon the Rooney Rule instituted by the NFL in the United States which requires teams to interview at least one minority candidate while hiring for a head coach position. Whether candidates are vetted by human hirers or AI hiring tools, protected minority candidates must receive serious consideration, possibly meaning a guaranteed interview or a "second look" during screening processes.

SPEAKER LIST

Steve Crown, Vice President and Legal Counsel Global Human Rights, Microsoft

Mitchell Baker, CEO of Mozilla

Mark Surman, Executive Director Mozilla

Craig Newmark, Founder of Craigs List

Chenai Chair, Mozilla Fellow, South Africa

Dr. Merhnoosh Sameki, Microsoft's Chief Technology Officer on Responsible AI

Wendy Chisholm, Chief Accessibility Architect, Microsoft

Rob Goldman, Former Vice President Ads Google

Deborah Raji (Nigeria/Canada), MIT Media Lab 30 under 30 Forbes Tech Leader

Deborah Raji is one of the three "Face Queens" along with Dr. Timnit Gebru.

Rati Thanawala, Harvard Kennedy School and former Vice-President Bell Labs

The first woman of color to graduate with a computer sci ence engineering PhD from Yale.

Ya Xu, Head of Data Science Practice at LinkedIn

Safiya Noble, Author of Algorithms of Oppression: Time 100 video interview with the Duke and Duchess of Sussex in October on Times Most Innovative People

Sandra Wachter Berkman, Professor at Oxford, Berkman Klein Center, Harvard and Visiting faculty, Harvard Law School

Jenni Olson, LGBTQ filmmaker, Criterion Collection

Dr. Gitanjali Swamy, Harvard Business School, and Managing Partner at IoT Task

Sanjay Sarma, Dean and Fred Fort Meyers and Daniel Fort Meyers Professor of Mechanical Engineering and Vice President MIT

Natalie Jabangwe, CEO of EcoCash, Zimbabwe

Raj Agarwal, Founder and CEO of Medocity- a leading telemedicine provider in New Jersey

Speaking to algorithmic biases in health care data, especially with immigrant populations.

Manish Raghavan,PhD Cornell Engineering

A.T. Cooper, Cornell Engineering

Anu Chittrapu, Head of The Indus Valley Entrepreneurs- Boston and Vice President Bank of America

Silda Spitzer, Former First Lady of New York and head of private equity group led by women

Eric Rosenblum, Managing Partner at Tsingyuan Ventures, Largest Venture Capital Firm for Chinese Diaspora Innovators