Unmasking Coded Bias

Why we need inclusion and equity in AI

Report on AI and Stereotype Threat in a New Generation of Black Students & Professionals

Foreword

A month before the first meeting of the Policy Lab on AI and Implicit Bias, Dr. Timnit Gebru's public departure from Google made national headlines. Dr. Gebru is a computer scientist who works on algorithmic bias and data mining, an advocate for diversity in technology, and co-founder of Black in AI, a community of black researchers working in artificial intelligence.

The public conversation spurred by Dr. Gebru and her colleagues on algorithmic bias allowed our lab to engage in this national dialogue and expand a more nuanced understanding of algorithmic bias in hiring platforms.

The mission of the lab was to engage in multilateral and multidisciplinary conversations and interrogate top industry and academic leaders working in the area. In this effort, I was privileged be partnered by Steve Crown, Microsoft's Vice President and Deputy General Counsel of Global Human Rights. Steve Crown's global experience as a Rhodes Scholar, his scholarly expertise on Russia, and work in China supported me in pushing the frontiers of the lab. He encouraged me in engaging in a conversation with the global South and to understand the power and potential of storytelling to make AI more inclusive.

Our speakers, among others included, Dean Sanjay Sarma of MIT, Professor Sandra Wachter of Oxford (visiting professor at Harvard Law School and Fellow at the Berkman Klein Center), and Time 100's Safiya Noble, the author of Algorithms of Oppression and conversationalist with the Duke and Duchess of Sussex. We were also joined by philanthropist Craig Newmark, Founder of Craigslist, Mitchell Baker, the CEO of Mozilla, leading women AI engineers from LinkedIn, Microsoft, and Ebay, and a new generation of minority computer engineers working on cutting-edge areas of algorithmic bias in employment platforms. The Lab also hosted venture capital principals funding women- and minority-owned technology startups. The Lab also partnered with MIT Media Lab's Deborah Raji. At age 24, Deborah Raji of Nigerian origin and acclaimed speaker in our Lab was chosen as one of MIT Review's youngest innovators of 2020.

Deborah Raji has spent years in partnership with Dr. Timnit Gebru and Joy Buolmwini. The three Black women computer scientists known as fondly by the media as "Face Queens" helped create the groundbreaking "Gender Shades" project at MIT Media Lab. The project pilots an intersectional and inclusive approach to testing AI. Their project, like ours, is concerned with unmasking the assumptions of AI neutrality and questioning whether the remnants of racism and sexism were hardwired into in AI.

Dr. 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 shows that 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 a 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.**

For our final project, we narrowed our focus to address algorithmic bias in employment platforms and contribute to the discussion started by the first diversity and inclusion report in Silicon Valley: *Elephant in Silicon Valley* in 2015. This conversation sparked by the infamous Google memo in 2017 created a new landscape which examined the bias in tech ecosystem. Despite all of dollars spent on diversity training, diversity offices, and reports, little has really changed in 2021.

Our mission was to identify a new generation of biases and "stereotype threat" in AI and help provide context and nuance to the conversation to mitigate those biases.

As part of the first step of identifying biases in AI-related recruitment platforms, students in the class led several informal pilot surveys, including:

1) **The Elephant in AI: Stereotype Threat in AI**

   Perceptions of emerging professionals using recruiting platforms.

2) **Algorithmic Bias in China**

   Comparison data from a new generation of lawyers and engineers using hiring platforms.

3) **Prove it again Bias in Silicon Valley**

   Stories from women and minorities in Silicon Valley

4) **Unmasking Coded Bias**

   Perceptions from a new generation of professionals from the Black community.

This last report and survey led by Amani Carter (with Ziguo Yang's technical assistance) is a critical contribution to understanding algorithmic bas at a time of a public reckoning on racial and intersectional injustice. This is a first-in-kind work to peer beneath the surface and understand the most important human rights issue of our time: how AI can reify and reconstruct bias based on our gender, age, race, and class. What we find here too is that even when the respondents may not yet have experienced bias in AI, stereotype threat can profoundly affect our use of AI, threatening to undermine performance, causing both emotional and intellectual reactions affecting our career choices.

Amani's work helps us recognize these algorithmic threats shared among a new generation of professionals and points us in the direction of new mitigation tools needed to address these threats.

For long, I had been inspired by Microsoft's CEO Satya Nadella's charge to empower every person on the planet and his deeply personal commitment to accessible AI and sustainable technologies. His dedication to the Convention on the Rights of Persons with Disabilities -- the first human rights treaty of the 21s century -- makes our Lab's human rights-based approach to addressing bias doubly important. A humanist and a scientist, Satya Nadella often speaks of the connection between the humanities and technology and quotes from the Pulitzer Prize winning Vijay Seshadri's *Imaginary Number*: "The soul, like the square root of minus 1, is an impossibility that has its uses."

This report is dedicated to Satya Nadella with appreciation.

Rangita de Silva de Alwis
Policy Lab on AI and Implicit Bias

Introduction

The summer of 2020 was characterized by two crises: the COVID-19 pandemic and the struggle for social justice.[i] As the virus tore through the country, millions of protestors and activist groups engaged in demonstrations across the US and around the world that reignited discussions about the role systemic racism plays in every aspect of American life.[ii] Amidst this renewed national conversation about the negative impacts of embedded anti-black bias, industry leaders across the country made statements firmly supporting and advocating for diversity in the workplace.[iii] In an effort to bolster these statements with action, several companies committed to making concrete internal changes including with respect to hiring practices.[iv] For example, Adidas committed to increasing the number of Black employees by filling thirty percent of all new positions with Black applicants.[v] This renewed commitment to a diverse workforce raises the question: how can companies reform their standard hiring practices such that they are able to recruit and maintain a strong pipeline of diverse talent? Rather than making reactive pushes for diverse hiring, how can companies proactively transform their hiring processes to consistently yield diverse hires. Some experts have suggested that artificial intelligence (AI) is the answer.[vi] This report focuses on evaluating how effective an answer AI might be.

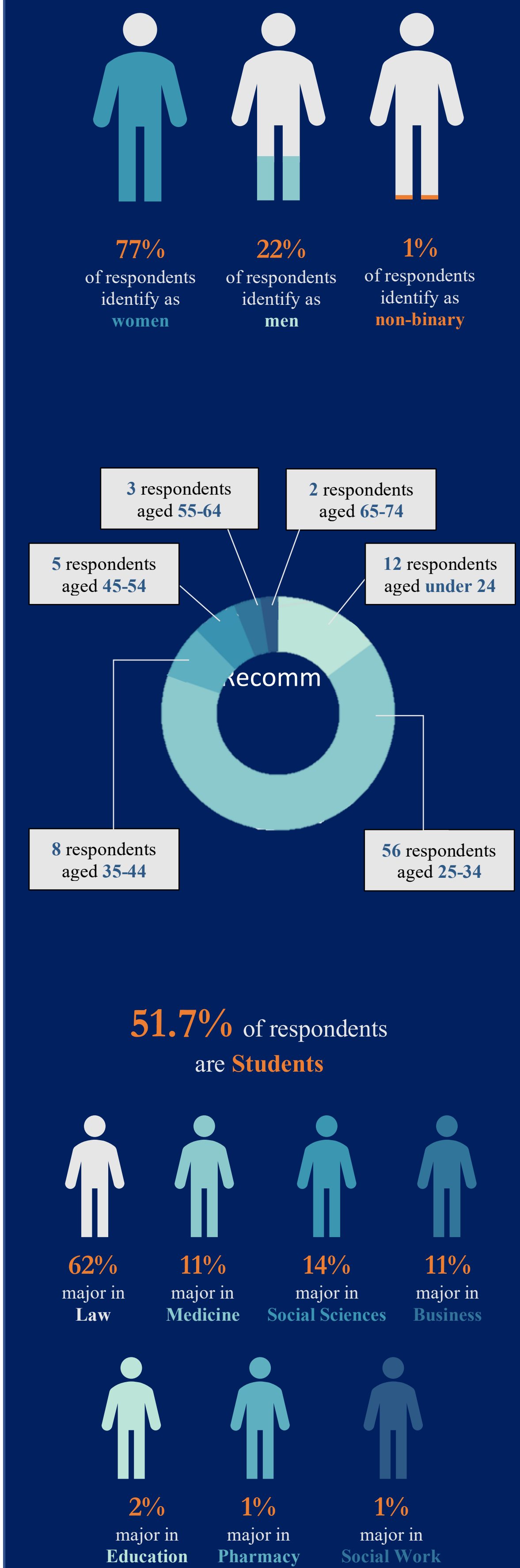
This report principally finds that AI can be a useful tool for increasing diversity but only insofar as the AI itself is designed and used in a broader equity and inclusion context. AI is not a panacea solution to corporate diversity woes. AI, designed without explicit attention to equity and inclusion issues, can be counterproductive to the goal of de-biasing the hiring process. Based on the survey responses collected, respondents were already experiencing the negative consequences of self-censoring bias, bias in design, and stereotype threat in their interactions with AI used by hiring platforms. Respondents also identified instances where the design of AI, both in terms of the type of evaluation and its pattern recognition, may be incorporating anti-Black bias. Furthermore, due to implicit biases, merely incorporating human decision-making is unlikely to cure these issues. If AI is to be the solution that experts suggest it can be, then those designing and using AI will need to be intentional about building equity and inclusion into the AI itself.

Methodology

This report is based on an empirical study of Black professionals and students from a variety of industries and fields of study regarding their perceptions of and experiences with anti-Black bias in online hiring platforms. To assess Black professionals' and students' perceptions, the author surveyed eighty-seven Black professionals and students[vii]. Seventy-seven percent of respondents identified as women, twenty-two percent identified as men, and one percent identified as non-binary. The vast majority of respondents are aged twenty-five to thirty-four. Thirty-five percent of respondents are students in the fields of Law, Medicine, Social Sciences, and Business. Black professional respondents reported working in a host of industries and fields including Consulting, Education, Fashion and Design, Human Resources, Public Health, Law, Real Estate, Sports, Finance, Economics, Technology and Engineering,

The author supplemented these survey results with an evaluation of three-hundred and sixty LinkedIn profiles to assess the impact of appearance bias on Black women specifically. The survey was distributed through the author's social network and the professional communities and associations to which the author belongs.

The survey consisted of three main sections. The first section collected demographic data about the respondents. The second section explored respondents' experiences with hiring platforms such as LinkedIn, Indeed, Monster.com, and ZipRecruiter. And the third section asked respondents about



their concerns regarding bias in hiring platforms. Questions in the second and third sections included a mixture of descriptive questions, Likert scale questions and a qualitative question inviting respondents to leave comments. Just under one-fourth of respondents left comments describing their experiences with and perceptions of anti-Black bias on the hiring platforms. Profiles evaluated as part of the supplementary LinkedIn analysis were selected through three searches, one for "black women," one for "black women business," and one for "black women law."

This report is not designed to make definitive claims about the efficacy of any particular site's algorithm or model. Evaluating the efficacy of an algorithm or model would require access to information that is typically kept private.[viii] Both the models and the sensitive data used to generate the models generally are not publicly available, and as such auditing a site's algorithm to determine how anti-Black bias does or does not manifest in its construction would be extremely challenging. This report's intervention principally attempts to contextualize the *outcomes* of hiring platforms' AI-powered practices in terms of the Black experience. This report's goal is to understand how interactions with these hiring platforms reflect, recreate and reinforce anti-Black bias as experienced by Black professionals and students.

A Definitional Note

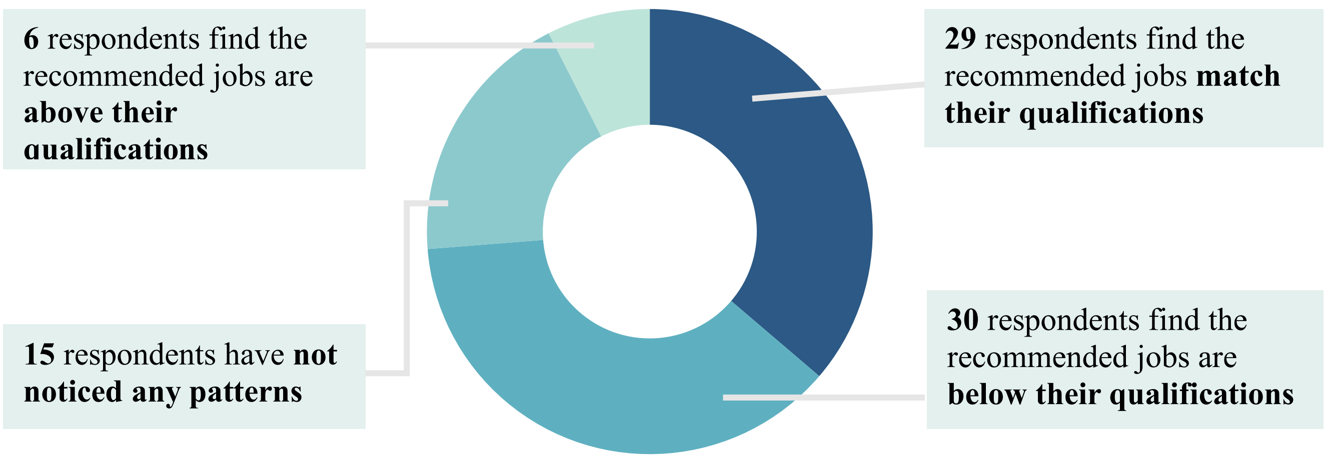
AI can be a nebulous term and is used to refer to a host of technological advancements ranging in sophistication. While industry leaders will be familiar with how AI works in this context, community members may not be. As the goal of this report is to contextualize hiring platforms' AI performance in terms of the Black experience, it is important that the report remains accessible to the Black community. Providing a brief definitional framework can help orient those less familiar with AI and how AI is used by hiring platforms. As explicated in Turner's work, AI can be broken into two broad buckets: narrow and general.[ix] Narrow AI is typified by a system's ability to achieve a certain stipulated goal or set of goals in a manner or using techniques which qualify as intelligent.[x] Examples of these limited goals include natural language processing functions like translation, or navigating through an unfamiliar physical environment.[xi] Narrow AI is suited only to the goal for which it was designed.[xii] General AI, in contrast, is capable of an unlimited range of goals.[xiii] General AI can set new goals independently, including in situations of uncertainty or vagueness.[xiv] General AI is what we typically see portrayed in popular culture, like the humanoid robots portrayed in the 2004 film *I, Robot* or the 1984 film *The Terminator*.[xv]

The AI utilized by hiring platforms falls into the category of narrow AI. The site collects data based on how users interact with the platform, including the information users include in their profiles. Hiring platforms feed that data into an algorithm that identifies interpretable patterns, which are then used to create a model that can make predictions about users – how likely a user is to engage with certain content, whether another user would make a good connection, whether job opportunities would be right for the user, etc.[xvi] When a hiring platform makes recommendations or optimizes search results, the platform takes into account user information like job searches, alerts, profile information, and site activity to push opportunities that its model predicts will be a good match for the user.[xvii] The survey sought to assess whether and how this process reflects, recreates and reinforces anti-Black bias.

Results

The data suggests that hiring platforms may be optimizing for Black users' identities as much or more than their actual credentials. The majority of respondents, seventy-seven percent, reported using hiring websites such as LinkedIn, Indeed, Monster.com, ZipRecruiter, etc. in the past year. Nearly forty-two percent of respondents reported feeling that the employment opportunities recommended to them on hiring platforms were mismatched with their credentials, with nearly thirty-five percent identifying such recommended jobs as below their qualifications. Just about thirty-three percent of respondents reported feeling that the job opportunities recommended to them match their qualifications. This suggests that the hiring AI used by respondents are almost as likely to underestimate Black respondents' abilities by recommending opportunities that are below respondents' qualifications as the AI is to

correctly assess Black respondents' abilities and match them with opportunities matching their qualifications. Furthermore, only six percent of respondents felt that the AI recommended jobs are above their qualifications suggesting that the AI is very unlikely to overestimate Black respondents' abilities. This is even more so the case with Black professionals. Only one percent of Black professional respondents reported feeling that recommended jobs were above their qualifications.

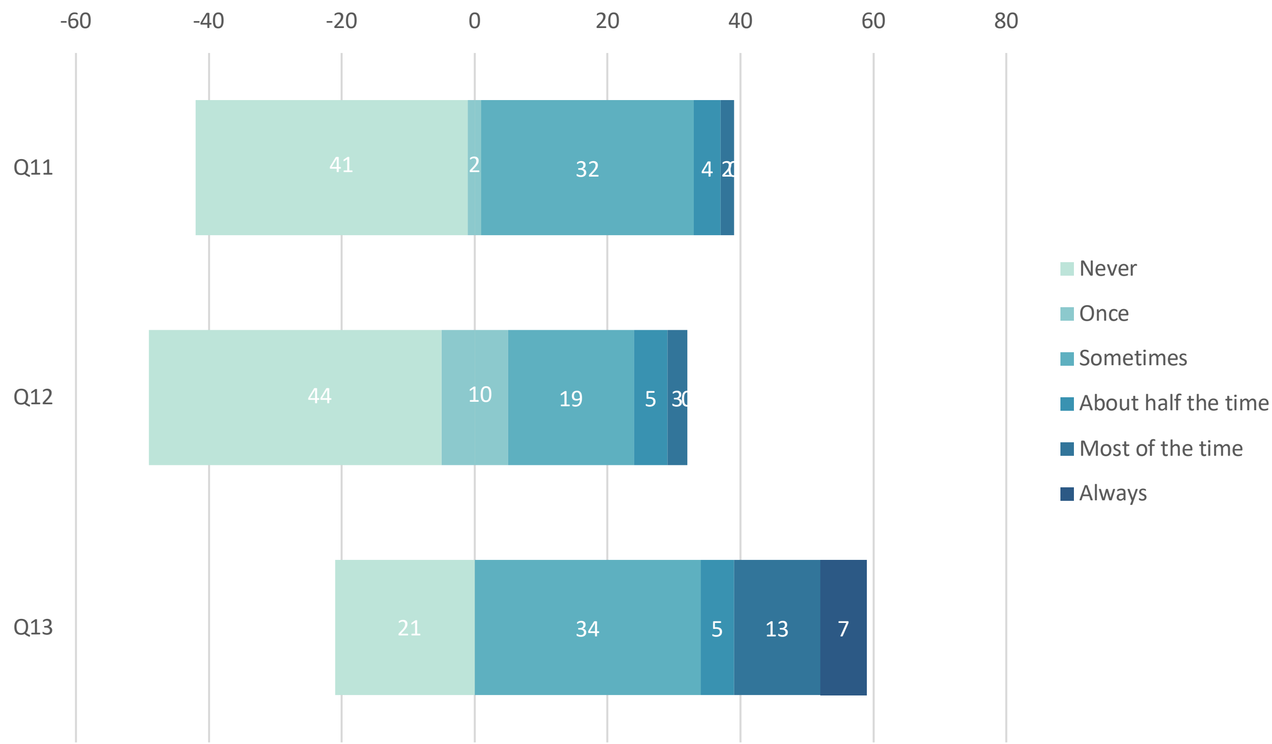


Graph 6: Respondents' answers to "Do you feel that the hiring platforms that you use recommend jobs that match your skills and expertise?"

When asked whether the hiring platforms that respondents used ever recommended a job for the respondent that they felt was target towards a particular aspect of their identity rather than their credentials, forty percent of respondents answered in the affirmative. Similarly, thirty-nine percent of respondents reported finding it difficult to locate job postings on the hiring platform(s) that respondents used because the position respondents were seeking was not one stereotypically held by people with respondents' identity. This suggests that hiring platforms' AI-powered recommendations may be nearly as likely to optimize for respondents' identities as their actual credentials.

Interestingly, when asked whether the hiring platforms that respondents used ever recommended academic programs that respondents felt were not on par with their credentials, sixty-three respondents answered in the affirmative. Nearly twenty percent of respondents indicated that they were recommended these lower tier programs always or most of the time. This is particularly jarring considering Black women have been shown to be the most educated group in the United States and the majority of our respondents identify as Black women.[xviii] Given the clear interest in education, one would think that these platforms would have a vested interest in accurately matching educational programs with Black women's credentials.

Graph 7: Respondents' answers to questions eleven through thirteen

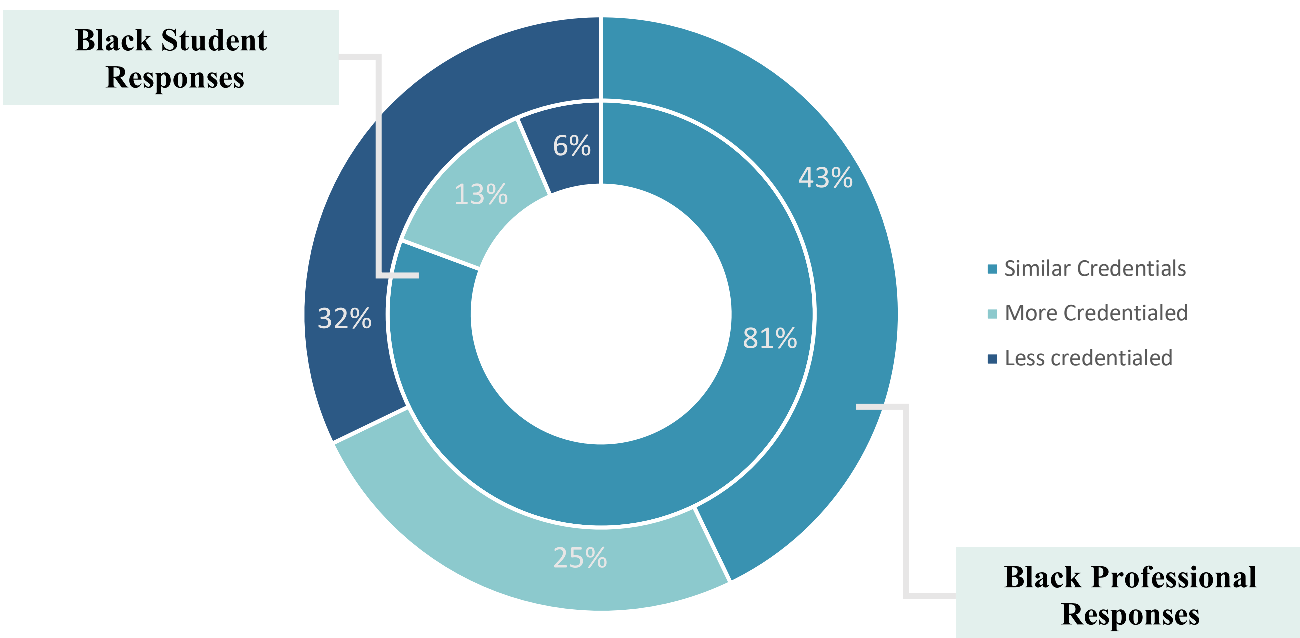


Q11: "Have the hiring platform(s) that you use ever recommended a job for you that you felt was targeted towards a particular aspect of your racial and/or gender identity as opposed to your credentials? (e.g. black man with master's in education recommended physical education/athletic coaching positions)

Q12: "Have you ever found it difficult to locate job postings 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. black male obgyn looking for hospital opportunities but search results primarily returns sports medicine results)

Q13: "Have the hiring platforms that you use ever recommended academic programs that you felt were not on par with your credentials?
(e.g. black woman with Master's in public health from John's Hopkins University recommended medical school program at University of Arizona)

Interestingly, when respondents were asked if they felt that the suggested professional connections recommended to them were similarly credentialed sixty-three percent of respondents answered in the affirmative. Of the Black student respondents that noticed a pattern, eighty-one percent felt that the connections recommended to them had similar backgrounds as them in terms of credentials, with only thirteen percent indicating that suggested connections were less credentialed than them. The picture is markedly different for Black professional respondents. Of the Black professional respondents that noticed a pattern, only forty-three percent felt suggested connections were similarly credentialed, and thirty-two percent felt that suggested connections were less credentialed. This suggests that while hiring platforms may be properly recognizing the credentials of Black students and connecting them to similarly credentialed users, hiring platforms may perform more poorly in this regard for Black professionals.



Graph 8: Respondents' answers to "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?"
*Note this graph denotes percentage of respondents that noticed a pattern

This differential may be due to the propensity for students to be recommended connections with fellow classmates. This may suggest that while hiring platforms are properly recognizing Black student respondents' status as students and recommending other students in their programs, the hiring platform performs less well as an emerging Black professional's career progresses.

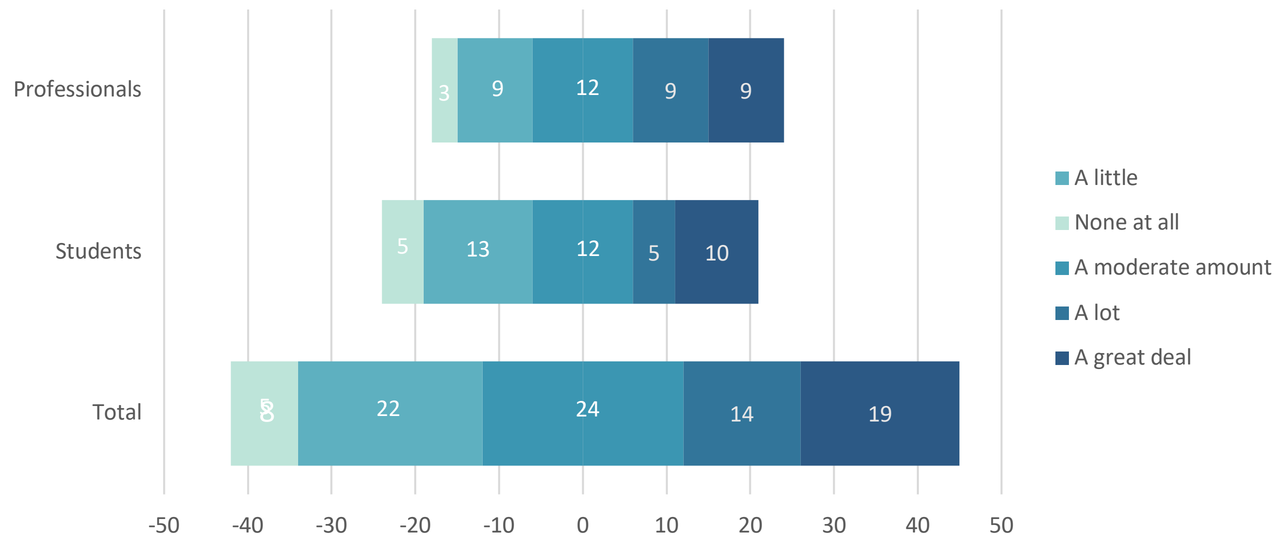
Furthermore, this report found some evidence of self-censoring bias and bias in design, as well as evidence implicating stereotype threat.

Self-Censoring Bias

Racial discrimination in the applicant evaluation process remains a pervasive problem in North American labor markets.[xix] A recent meta-analysis of available field experiments of hiring discrimination – studies in which fictionalized matched candidates from different racial or ethnic groups apply for jobs – found that hiring discrimination against African-Americans[xx] has remained unchanged for nearly five decades.[xxi] Evidence suggests resumes containing minority racial cues, such as a distinctively Black name lead to thirty to fifty percent fewer callbacks from employers than do otherwise equivalent resumes without such cues.[xxii] One of the ways applicants respond to this manifestation of anti-Black bias is self-censoring application materials through practices like resume whitening. Black applicants "whiten" their resumes by deleting references to their race with the hope of boosting their shot at jobs – a strategy which has proven successful.[xxiii] Whitening techniques such as omitting Black professional associations from a resume or emphasizing experiences that signal whiteness have been shown to increase likelihood of callbacks.[xxiv] Black applicants self-censor, exclude certain aspects of their professional or personal experiences that may be associated with Blackness, to avoid being penalized by recruiters. And what's more concerning is that the evidence shows that such self-censoring is working.

When respondents were asked how much, if at all, respondents worry that employers or managers using AI-based recruiting tools might not consider respondents for a position because of respondents' racial identity, only nine percent of respondents answered not at all. Only seven percent of Black professionals indicated the same. Nearly twenty-two percent of respondents indicated that they worried a great deal about not being considered for a position because of their racial identity.

Graph 9: Respondents' answers to "How much, if at all, do you worry that employers or managers using AI-based recruiting tools might not consider you for a position because of your racial identity?"



This report found a clear concern that indicating one's racial identity could limit professional opportunities. This report also found a corresponding impulse to engage in self-censoring techniques to increase the likelihood of favorable outcomes on these hiring platforms. Several respondents expressed the desire to remove their racial identity in such a way that the AI would not be able to categorize the respondent as Black. Two Black professional respondents working full-time in Education, indicated that they would prefer race be excluded as a factor. One said she didn't see the value in adding a race category and would prefer that it be removed altogether. The other preferred that race not even be asked saying she felt it was "super biased." One respondent, a Black professional working full-time in Commercial Real Estate, described attending a "lunch and learn" program designed to teach attendees techniques to "get around" online applicant tracking systems. Another Black professional respondent working full-time in Public Health felt the need to circumscribe her political expression to resonate with colleagues. Rather than "meaningfully communicate her radicalism and Black feminist politics," she engages these topics more shallowly by "repurposing the buzzwords of the day" much like colleagues who she feels "co-opt the language of resistance, progress, [and] struggle." Further, the supplemental LinkedIn study revealed that sixty-two percent of Black women present with flat-ironed or straightened hairstyles as opposed to natural hairstyles. Each of these reflect an impulse to self-censor – to remove indicators of Blackness from employment applications for fear of anti-Black bias or outright discrimination. A fear that has been substantiated by historical experience.

This is worrisome for at least two reasons: i) Black applicants that engage in this particular kind of self-censoring may ultimately be overlooked when companies launch programs designed to increase diversity hires, and ii) widespread self-censoring in the Black community can result in a dearth of data used to train the AI that hiring platforms use to identify potential candidates which ultimately could result in less accurate recommendations for Black candidates regardless of campaign. The first concern is salient from a reinforcement perspective. Consider Adidas's pledge to fill thirty percent of open positions with Black and Latinx applicants. If Adidas decides to work with a hiring platform to optimize for Black and Latinx users to ensure Adidas gets the widest applicant pool possible, Black users who have engaged in self-censoring may be at a disadvantage. The AI will have fewer data points to confidently determine that a particular user is appropriate for this campaign, and thus the hiring platform may be less likely to advertise the Adidas positions to Black users who self-censored. This could lead to underrepresentation of Black candidates in the applicant pool and ultimately in the new hire class, which could then reinforce amongst the Black community that anti-Black bias or outright discrimination runs rampant in the hiring process. It could also reinforce amongst other communities that Black candidates are less qualified or deserving. This could lead to an increase in self-censoring – and the cycle continues.

The second concern is salient from recreation perspective. The kind of AI used by hiring platforms require a sizeable amount of data to develop reliable patterns.[xxv] The more data available, the more accurate the predictive model can be. Self-censoring by Black users reduces the number of data points that can be used in a predictive model, which can result in a model that performs less effectively for Black users than other groups that do not engage in self-censoring. In this way, the AI used by hiring platforms could quite literally be recreating this dynamic whereby Black applicants face anti-Black bias or discrimination in the hiring process.

Bias in Design

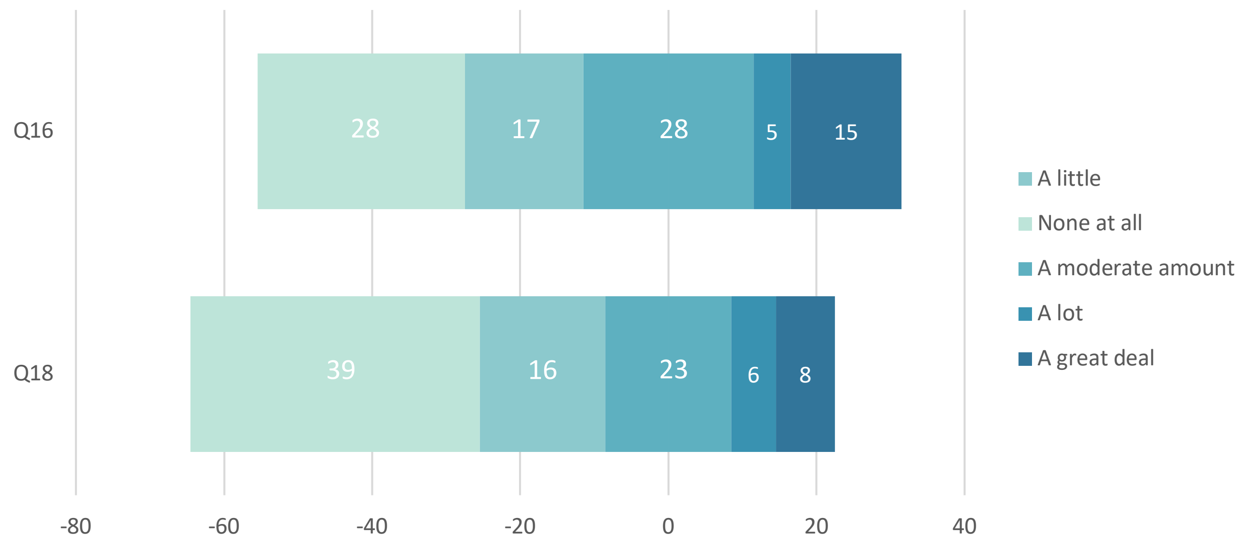
Much has been written about whether and how our biases can be baked into AI.[xxvi] This is especially troublesome when AI is used in the hiring context, given how employment impacts quality of life.[xxvii] Anti-Black bias can be incorporated into AI in several ways, but this report focuses on: i) inequitable determination of target variables, and ii) data inequity. The programmers that design AI for use in the hiring context translate the desired outcome, identify an ideal employment candidate, into a question about the value of some target variable, like communication skills.[xxviii] Defining an ideal employment candidate is challenging because it requires prioritization of numerous observable characteristics that make an employee "good." Programmers could define target variables in ways that correspond to measurable outcomes such as relatively higher sales, shorter production times, or longer tenure.[xxix] Or, in an effort to ostensibly create a more holistic evaluation, programmers could define target variables in terms of previous annual reviews or overall assessments of performance.[xxx] Doing the latter can permit anti-Black bias inherent in those legacy review processes or assessments of performances to leak into AI performance. In this way inequitable determination of target variables can result in anti-Black bias within the AI's design.

The two kinds of data inequity that this report is principally interested in are subjective data labeling and biased sampling. Data labeling is the process by which training data is manually assigned labels by programmers.[xxxi] For example, if a programmer is designing a model that can predict consumer creditworthiness, then data sets containing information about how often consumers pay bills on time, will need to be labeled. The programmer will need to determine which kinds of data should be labeled as defaulting and which should not. Because assigning these labels can be arbitrary, this process is ripe for anti-Black bias.[xxxii] Biased data sampling is briefly discussed above. Decisions that depend on conclusions drawn from incorrect, partial, or nonrepresentative data may discriminate against Black applicants.[xxxiii] Not all data is created or collected equally – dark zones or shadows where Black citizens and communities are overlooked or underrepresented can yield models infused with anti-Black bias.[xxxiv]

This report found that the majority of respondents were concerned about address or interest-based biases negatively impacting their prospects. When asked whether respondents were concerned that AI-based recruiting tools might overlook their profile due to listed interests, just over fifty-five percent of respondents answered in the affirmative with nearly sixty percent of Black professionals reporting the same. When asked whether respondents were concerned that

AI-based recruiting tools might overlook their profile due to address sixty-eight percent of respondents answered in the affirmative.

Graph 10: Respondents' answers to questions sixteen and eighteen



Q16: "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?"

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 of your listed interests or extracurricular involvements? (e.g. concerned you may not be considered because you played basketball instead of golf)

This suggests respondents may be concerned that the areas where they live and the activities they partake in will be underrepresented in data sets. Given the segregated nature of housing in the United States,[xxxv] respondents' concern is reflective of the worry that Black neighborhoods and interests may fall into a "dark zone" – the data sample used to train AI-models may have little to no information about how applicants from Black neighborhoods or interested in activities more common amongst the Black community will fare in the role.

This report found further qualitative evidence showing respondents' concern about and experience with anti-Black bias in design. One respondent, a Black woman professional working full-time in Human Resources, noted that she has observed implicit bias in her office's uses AI in their recruitment screening process. She noted that the Black applicants "may not have access to the resources that groom them to be able to provide the expected responses" that the AI screens for and as a result Black applicants are "immediately discounted for answering truthfully." This respondent observed a clear disparity in outcomes and attributed this disparity to the AI screening tool used by her employer. The respondent's comment that Black applicants are less likely to be "groomed" to provide "expected responses" may be indicative of a data inequity problem. The training data used to create this particular AI's model may have a biased data sample, one over-inclusive of candidates from other racial backgrounds who do have access to resources that train them to provide expected responses. Unconscious biases of a programmer labeling the training data may also have leaked into the model during the labeling process. The programmer may have arbitrarily and unconsciously labeled responses commonly given by non-Black candidates as favorable while labeling responses commonly given by Black candidates as less favorable.

> Some employers utilize platforms to take personality and skills tests, which ironically are intended to decrease the biases in the hiring processes. However, I fear that these tests still promote bias, such as the skills tests which contained LSAT based questions. It's been discussed how the LSAT and other standardized skills tests might disproportionately impact test takers based on race, socioeconomic status, etc.

Another respondent expressed concern about AI hiring platforms incorporating legacy skills-based tests that have been shown to disadvantage Black test takers. The respondent, a Black woman law student, recalled an experience taking a skills-based assessment as part of the hiring process for a potential opportunity. The assessment incorporated questions based on the Law School Admission Test (LSAT), a standardized exam integral to the law school admissions process in the United States.[xxxvi] Studies have shown that the LSAT disadvantages Black test takers.[xxxvii] Data has shown that Black test takers fare worse on average than their white counterparts.[xxxviii] Performance assessments that incorporate questions based on the LSAT inherit the anti-Black biases present in the LSAT and may function to disadvantage Black law students applying for jobs that use this screening method. This is a perfect example of inequitable determination of target variables. In this case the target variable identified, high score on the skills-based assessment, was defined in terms of a legacy assessment, the LSAT, that has been shown to disproportionately disadvantage Black applicants.

Another Black woman law student respondent remarked that the hiring platform she engages with primarily recommends employment opportunities in diversity, equity and inclusion (DEI) focused roles despite her not having mentioned anything about this area in her profile. This could be indicative of either inequitable determination of target variables or a data inequity problem. A programmer's implicit bias could have snuck into the target variable setting process if the programmer defined ideal employment candidate in terms of racial identity rather than in terms of express interest in DEI career opportunities. Presuming that a Black employment candidate would be a good fit for DEI roles, regardless of indicated interest, discriminates against the candidate by reducing them to their racial identity rather than engaging their full professional potential and can contribute to tokenism in the workplace.[xxxix] Additionally, implicit bias could have snuck into the data sampling process if the data set used to train the AI's model overrepresented Black candidates and underrepresented candidates that expressed interest in DEI opportunities.

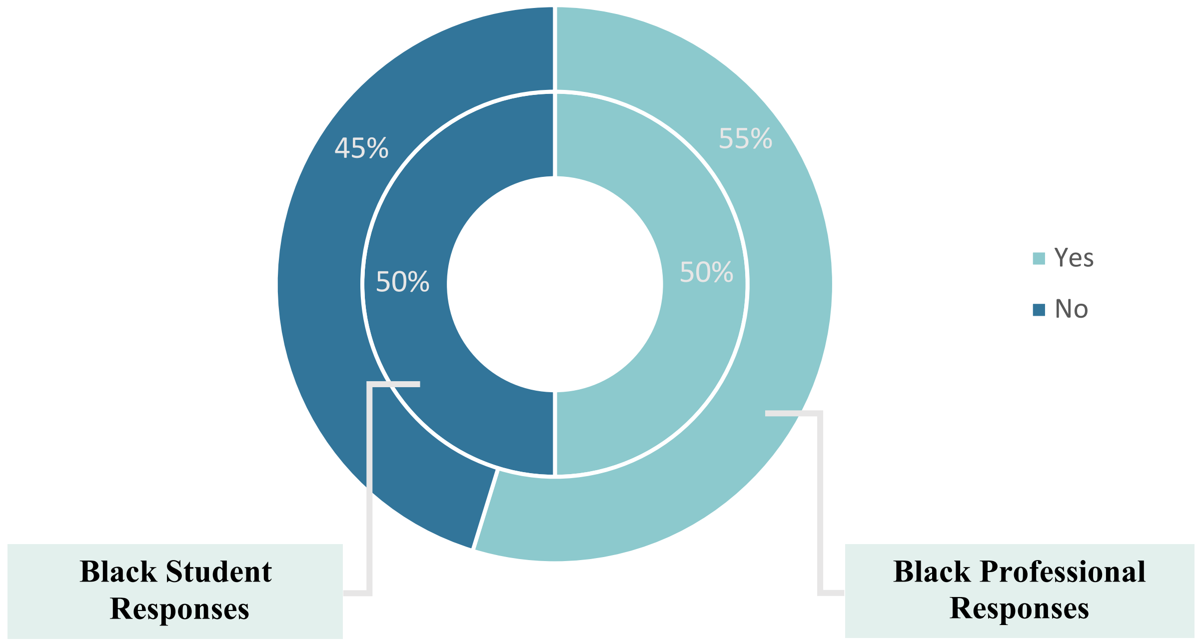
The data pointing to anti-Black bias in design is worrisome particularly in light of the inaccessibility of AI design. AI is a complex technology that can be difficult to understand. Employers and HR departments that utilize AI as a screening tool may not be aware that anti-Black bias can manifest in data labeling, data sampling, and setting target variables. This can cast a veneer of fairness over the AI-powered screening process whereby the employer or HR department presumes that the process is fair because they are unaware of how bias can be baked into AI's design. This is concerning from a reflection perspective and from a recreation perspective. When existing implicit or explicit anti-Black bias leaks into the design of an AI's model, that model reflects the same prejudices that it was supposed to be mitigating. Moreover, the model recreates discriminatory outcomes in a manner that is scalable and difficult to challenge given that the AI appears facially neutral as opposed to facially discriminatory.

Stereotype Threat

Stereotype threat is one of the most widely studied social psychological concepts of the past twenty years.[xl] Stereotype threat is defined as a situational predicament in which individuals are at risk, by virtue of their actions or behaviors, of confirming negative stereotypes about their group.[xli] Take a standardized testing scenario for example. When Black test takers sit for the exam, a situation requiring the test taker to display intellectual ability, the test taker may fear that they may confirm negative stereotypes about Black people's intellectual ability.[xlii] This fear of stereotype confirmation can hijack the cognitive systems required for optimal performance and result in poorer test performance.[xliii] Research over the last two decades has shown repeatedly that stereotype threat contributes to low performance among marginalized groups including Black people.[xliv] The hiring process is precisely the kind of situational predicament wherein stereotype threat can arise.

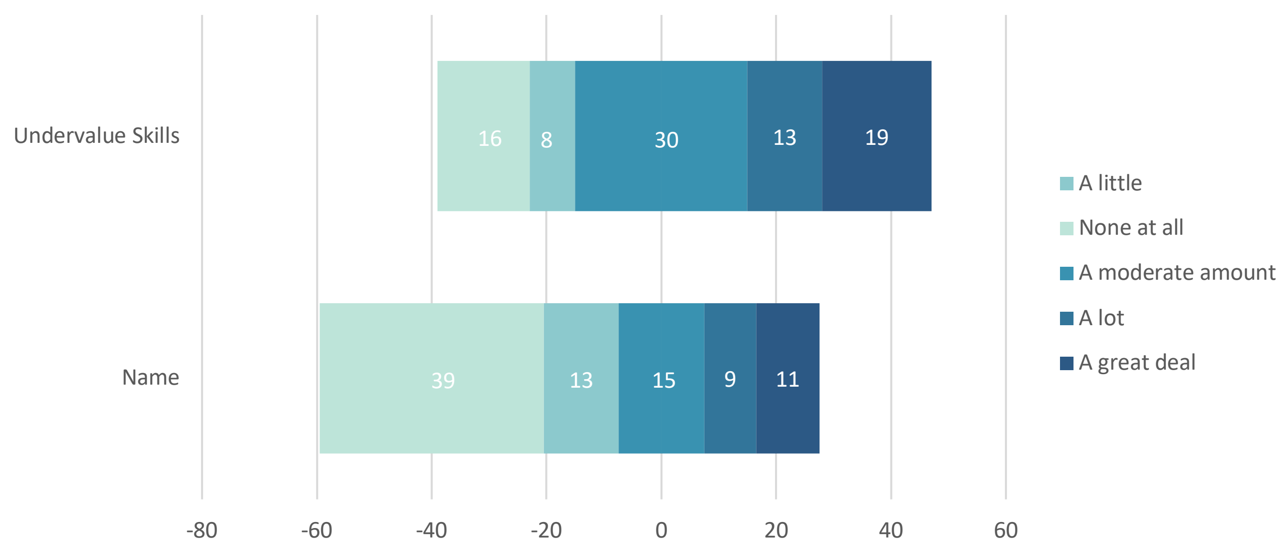
This report found ample evidence suggesting that Black students and professionals are concerned about facing anti-Black bias during the hiring process. Just over half of all respondents report having observed bias in the hiring or recruiting process on hiring or recruiting websites. Black professionals are slightly more likely to have observed such bias with fifty-five percent of respondents indicating having observed bias in the hiring process.

Graph 10: Respondents' answers to "Have you observed any sort of bias in the hiring/recruiting process on hiring/recruiting sites?"



Recall, when respondents were asked how much, if at all, respondents worry that employers or managers using AI-based recruiting tools might not consider respondents for a position because of respondents' racial identity, only nine percent of respondents answered not at all.[xlv] Only seven percent of Black professionals indicated the same. Recall also, when asked whether respondents were concerned that AI-based recruiting tools might overlook their profile due to listed address, just over fifty-five percent of respondents answered in the affirmative with nearly sixty percent of Black professionals reporting the same.[xlvi] When asked whether respondents were concerned that AI-based recruiting tools might overlook their profile due to address sixty-eight percent of respondents answered in the affirmative.[xlvii]

Graph 11: Respondents' answers to questions seventeen and twenty-four



Q17: "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 your name?"

Q24: "How much, if at all, do you worry that employers or managers using AI-based hiring platforms will undervalue some of your skills or experiences? (e.g. concerned that your experience working a paid service position such as waitressing will not be valued as much as an unpaid internship position in your area of interest)

Additionally, respondents indicated concerns that their profiles may be overlooked due to their names and worried that employers would undervalue their skills. The concerns reflected in question seventeen may be rooted in the negative stereotyping of names common amongst the Black community as ghetto, undesirable, and unprofessional.[xlviii] The concerns reflected in question twenty-four may be rooted in concerns that blue-collar or service work is negatively stereotyped as menial and intellectually unchallenging,[xlix] and that the people that do this work are negatively stereotyped as less capable than others.

One respondent noted observing that the recommendations suggested to her on hiring platforms screamed “come, be the underpaid Black woman who will do both paid and unpaid labor for our company!” She also remarked feeling that these “platforms reflect and reinforce the narrow conceptions of Black women’s professional possibilities” which are rooted in negative stereotypes about Black women. Another respondent remarked that they felt employers would either lazily include or exclude them based on their racial identity. Another respondent noticed a bias toward younger white women of a certain socio-economic class, and noted that these women were granted managerial positions over people of color with years of experience.

It is clear that respondents are aware of, are concerned about, and in a substantial number of cases have affirmatively observed anti-Black bias in hiring creating the perfect conditions for stereotype threat to thrive. This suggests that a significant portion of Black students and professionals may be encountering stereotype threat as they assemble their applicant materials, engage with hiring platforms, and take AI-based assessments. This could, as has been observed in the standardized testing context, lead to poorer performances overall for Black applicants. This is particularly concerning from a reinforcement perspective. Stereotype threat, while discussed at length in academic settings, is not a popularized concept. Data showing that Black applicants perform poorer overall than applicants that do not face stereotype threat may be used as evidence indicating that stereotypes Black applicants fear are, indeed, true. In this way stereotype threat can trigger a cyclical pattern wherein outcomes serve to reinforce the negative stereotypes that produce those stereotypes.

Conclusion

The overall value of this report's intervention is that it models precisely the kind of inquiry that hiring platforms, employers, programmers and designers of AI should be undertaking. Much has been written about the need for representation of marginalized peoples on the teams that build AI.[1] This research challenges us to think beyond representation in terms of the engineers and programmers that build AI and begin to broaden the conversation to include the marginalized communities impacted by this technology. This report has shown evidence suggesting that the AI-powered hiring platforms used by Black students and professionals reflect, recreate, and reinforce anti-Black bias. Findings such as this should serve as an indicator that these AI models are, in fact, incomplete. The impulse to deploy products to the market without assessing their impact, in terms of outcome, on marginalized communities is a dereliction of duty but more concerningly is a missed opportunity. While, legacy systems that discriminate against marginalized communities can be especially challenging to supplant and transform, innovative technologies such as AI present us with a unique opportunity to build that discrimination out of our society. Accomplishing this is a daunting challenge, but so was the creation of this technology. We need only decide that AI is unfinished until it works for all, that only AI built in the context of equity and inclusion are market ready, to find that we can meaningfully transform our world.

Endnotes

[i] See Elly Belle, Movements & Memes: How The Struggle For Social Justice Shaped 2020, Refinery29, Last Updated December 10, 2020 https://www.refinery29.com/en-us/2020/12/10207922/social-justice-movements-2020-covid-black-lives-matter (discussing how activism and the social justice movement took center stage during 2020); see also, Kalhan Rosenblatt, A summer of digital protest: How 2020 became the summer of activism both online and offline, NBCNews, Sept. 26, 2020, https://www.nbcnews.com/news/us-news/summer-digital-protest-how-2020-became-summer-activism-both-online-n1241001 ("The summer was marked by a surging movement of activism calling for social change but with the coronavirus pandemic affecting how people interact with one another, many of these calls to action took place online.");

[ii] See Justin Worland, America's Long Overdue Awakening to Systemic Racism, TIME, June 11, 2020, https://time.com/magazine/us/5851840/june-22nd-2020-vol-195-no-23-u-s/ (discussing the debate regarding systemic racism in the US sparked by the protests during the summer of 2020); see also Derrick Bryson Taylor, George Floyd Protests: A Timeline, The New York Times, March 28, 2021, https://www.nytimes.com/article/george-floyd-protests-timeline.html (discussing the protests that erupted in the wake of George Floyd's death and providing a timeline of such protests); Helier Cheung, George Floyd death: Why US protests are so powerful this time, BBC News, June 8, 2020, https://www.bbc.com/news/world-us-canada-52969905 (discussing how this wave of protests in response to the killing of a Black person in police custody was more sustained and widespread with demonstrations in all fifty states including in cities and rural communities).

[iii] See Jay Peters, Big Tech Companies are Responding to George Floyd in a Way They Never Did for Michael Brown, The Verge, June 12, 2020, https://www.theverge.com/2020/6/5/21281017/amazon-apple-facebook-response-george-floyd-michael-brown-tech-companies-google (discussing Amazon, Apple, Facebook and other peers in the tech industry making statements of solidarity against racial injustice for George Floyd); see also Brian Heater, Tech companies respond to George Floyd's death, ensuing protests and systemic racism, Tech Crunch, June 1, 2020, https://tcrn.ch/2Av8vim (discussing the responses of several tech companies reaffirming their commitment to diversity and opposition to racism); Natalie Sherman, George Floyd: Why are companies speaking up this time?, BBC News, June 7, 2020, https://www.bbc.com/news/business-52896265 (discussing corporations' varied responses during the summer of activism).

[iv] See Gillian Friedman, Here's What Companies Are Promising to Do to Fight Racism, The New York Times, Aug. 23, 2020, https://www.nytimes.com/article/companies-racism-george-floyd-protests.html (listing companies' concrete commitmenst in response to the protests following George Floyd's death); see also Message from the Adidas Board: Creating Lasting Change Now, Press Releases, Adidas, June 09, 2020 https://www.adidas-group.com/en/media/news-archive/press-releases/2020/message-adidas-board-creating-lasting-change-now/ (announcing Addidas's commitment to filling 30% of all new positions in the US at adidas and Reebok with Black and Latinx people); Sheryl Sandberg, Supporting Black and Diverse Communities, Facebook, June 18, 2020, https://about.fb.com/news/2020/06/supporting-black-and-diverse-communities/ (highlighting commitment to double the number of Black and Latinx employees working at Facebook and committing to have 30% more people of color, including 30% more Black people in leadership positions).

[v] See Id.

[vi] See Anne Fisher, AI for Hire: 4 Ways Algorithms Can Boost Diversity in Hiring, Fortune, June 1, 2019 https://fortune.com/2019/06/01/ai-artificial-intelligence-diversity-hiring/ (suggesting that AI can be used to address bias in the language used for job postings, widen the pool of eligible workers, identify talent beyond the typical resume inputs, and can make it easier to spot biased assumptions); see also Kathy Caprino, How AI Can Remove Bias From The Hiring Process And Promote Diversity and Inclusion, Forbes, Jan 7, 2021 https://www.forbes.com/sites/kathycaprino/2021/01/07/how-ai-can-remove-bias-from-the-hiring-process-and-promote-diversity-and-inclusion/?sh=762d79214ec5 ("Using proven, ethical AI in the hiring process can serve to provide objective information to hiring teams so that they focus on what really matters during the hiring process: characteristics of a candidate that are relevant to success in the job").

[vii] For the purposes of this report, Black corresponds to respondents that indicated that they identify as Black, Afro-Carribean, or African American. Some respondents indicated that they identify with more than one racial identity, those repondents that selected Black, Afro-Carribean, or African American *in addition to* other racial identities such

as Hispanic or Latinx, Asian or Asian American, White or Caucasian, or Native or Indigenous American were included in the data set as well. The survey contained results for respondents that did not indicate Black, Afro-Carribean, or African-American as one of their racial identities. These responses were removed from the data set so that the data can reflect the Black experience. These responses reflected similar attitudes and concerns perhaps due to these respondents' engagement with social justice causes that affect the Black community.

[viii] See Manish Raghavan et al., Mitigating Bias in Algorithmic Hiring: Evaluating Claims and Practices, Cornell University, accessed March 26, 2019, available at https://arxiv.org/pdf/1906.09208.pdf (discussing how "certain models may be completely inaccessible to the public, whether for practical or legal reasons, and attempts to audit these models by examining their training data or outputs might place users' privacy at risk").

[ix] Jacob Turner, Robot Rules: Regulating Artificial Intelligence, Palgrave McMillan 6 (2019).

[x] Id. There are other ways of thinking about what constitutes AI. There is a vibrant discussion amongst thought leaders regarding how we should be thinking about AI. We use Turner's theoretical framework for understanding AI because of its conceptual accessibility.

[xi] Id.

[xii] Id.

[xiii] Id.

[xiv] Id.

[xv] Id.; James Cameron, The Terminator, (1984) https://www.imdb.com/title/tt0088247/; Alex Proyas, I, Robot, (2004) https://www.imdb.com/title/tt0343818/.

[xvi] See McKenzie Raub, Bots, Bias, and Big Data: Artificial Intelligence, Algorithmic Bias and Disparate Impact Liability in Hiring Practices, 71 Arkansas Law Review 529, 532-33 (2018) (explaining that deep learning, a subset of AI, uses algorithms to collect and analyze data to identify "interpretable patterns" otherwise too subtle or complex for unaided human discernment and that once the data is collected and relationships are identified it is called a model).

[xvii] See Job Recommendations Overview, LinkedIn Help, LinkedIn, (Last updated October 2020) https://www.linkedin.com/help/linkedin/answer/11783 (discussing how LinkedIn's job recommendations function operates).

[xviii] See Degrees conferred by race and sex, Fast Facts, US Department of Education, National Center for Education Statistics, (2016) https://nces.ed.gov/fastfacts/display.asp?id=72; See also, Samuel Osborne, Black women become most educated group in US, Independent, (03 June 2016 8:37) https://www.independent.co.uk/news/world/americas/black-women-become-most-educated-group-us-a7063361.html; Kayla Stewart, Black women are now America's most educated group, Upworthy, (5.27.16) https://www.upworthy.com/black-women-are-now-americas-most-educated-group.

[xix] Sonia K. Kang et al. Whitened Resumes: Race and Self-Presentation in the Labor Market, vol. 61 Administrative Science Quarterly 3, 469 (2016).

[xx] This report intentionally uses the term Black rather than the term African-American. We used African-American in this particular context because the authors of the cited study do so. We also elect to capitalize the term Black. For an in-depth discussion of perspectives on the capitalization of "Black" see John Eligon, A Debate Over Identity and Race asks, Are African-Americans 'Black' or 'black,' The New York Times, June 26, 2020 https://www.nytimes.com/2020/06/26/us/black-african-american-style-debate.html.

[xxi] Lincoln Quillian et al., Meta-analysis of field experiments shows no change in racial discrimination in hiring over time, 114 Proceedings of the National Academy of Sciences of the United States of America 41, October 10, 2017 https://doi-org.proxy.library.upenn.edu/10.1073/pnas.1706255114

[xxii] Supra note xvii.

[xxiii] Id. at 470; see also Diana Gerdeman, Minorities Who 'Whiten' Job Resumes Get More Interviews, Working Knowledge: Business Research for Business Leaders, Harvard Business School, May 17, 2017, https://hbswk.hbs.edu/item/minorities-who-whiten-job-resumes-get-more-interviews.

[xxiv] Supra note xvii at 491.

[xxv] See supra note xvi (explaining that deep learning requires "big data," immensely large collected data sets, to analyze and reveal patterns and trends).

[xxvi] See Claire Cain Miller, When Algorithms Discriminate, The New York Times, July 9, 2015 https://www.nytimes.com/2015/07/10/upshot/when-algorithms-discriminate.html (explaining that even if AI is not designed with the intent of discriminating, they may reproduce social preferences and forms of

discrimination); see also Algorithmic bias and the Weaponization of Increasingly Autonomous Technologies: A Primer, United Nations Institute for Disarmament Research, 2 (2018) (discussing the ways in which biases can be introduced in algorithms); Rebecca Heilweil, Why algorithms can be racist and sexist, Vox, Feb. 18, 2020 https://www.vox.com/recode/2020/2/18/21121286/algorithms-bias-discrimination-facial-recognition-transparency (discussing the ways in which AI can be biased based on who builds them, how they're developed, and how they're ultimately used); see also Annie Brown, Biased Algorithms Learn from Biased Data: 3 Kinds of Biases Found in AI Datasets, Forbes, Feb. 7, 2020 https://www.forbes.com/sites/cognitiveworld/2020/02/07/biased-algorithms/?sh=52000b0276fc (describing how biased data sets can result in biased algorithms).

[xxvii] See Universal Declaration of Human Rights, Article 23, United Nations (recognizing the importance of employment to quality of life by declaring "[e]veryone has the right to work, to free choice of employment, to just and favourable conditions of work and to protection against unemployment").

[xxviii] Supra note xvi at 533.

[xxix] See Solon Barocas & Andrew D. Selbst, Big Data's Disparate Impact, 104 California Law Review 671, 679 (2016) (discussing why target variable setting should focus on observable characteristics).

[xxx] See Id. at 679-80.

[xxxi] See Id. at 681.

[xxxii] The arbitrariness of this process is described in David J. Hand, Classifier Technology and the Illusion of Progress, 21 Statistical Science 1, 11 (2006). The example he explored focuses on the consumer credit space. Take, for example, a data set containing information about consumers' payment history is being utilized to train an AI to determine creditworthiness. Labeling a customer as defaulting if they fall three months behind in payments is more or less arbitrary – alternative definitions (e.g. four months late) might be just as reasonable if not more useful. Once we account for unconscious bias, it is easy to see how

[xxxiii] See Supra note xxviii at 684.

[xxxiv] See Id. at 684-85.

[xxxv] See Joseph P. Willians, Segregation's Legacy: The Report, US News & World Report, (April 20, 2018) https://www.usnews.com/news/the-report/articles/2018-04-20/us-is-still-segregated-even-after-fair-housing-act (discussing how fifty years after the Fair Housing Act was signed, America is nearly as segregated as when President Lyndon Johnson signed the law).

[xxxvi] See What is the LSAT?, Law School Admissions Council (2021) https://www.lsac.org/lsat.

[xxxvii] See Latasha Hill, Less Talk, More Action: How Law Schools Can Counteract Racial Bias of LSAT Scores in the Admissions Process, 19 University of Maryland Law Journal of Race, Religion, Gender and Class 313 (2019) (discussing statistical evidence showing the severity of racial bias in LSAT scores); see also Aaron N. Taylor, The Marginalization of Black Aspiring Lawyers, 13 FIU Law Review 489 (2019) (discussing how the LSAT is used to systematically exclude Black law school applicants from admissions classes).

[xxxviii] See Susan P. Dalessandro, Lisa C. Anthony, and Lynda M. Reese, LSAT Performance With Regional, Gender, Racial/Ethnic Breakdowns: 2007-2008 Through 2013-2014 Testing Years (TR 14-02), LSAT Technical Reports, Law School Admission Council, https://www.lsac.org/data-research/research/lsat-performance-regional-gender-and-racialethnic-breakdowns-2007-2008.

[xxxix] Tokenism can be defined as the policy or practice of making merely a symbolic effort. See Merriam-Webster, Definition of Tokenism, (1960) https://www.merriam-webster.com/dictionary/tokenism; See also, Kara Sherrer, What is Tokenism, and Why Does It Matter in the Workplace, Vanderbilt University Owen School of Management, (Feb. 26, 2018) https://business.vanderbilt.edu/news/2018/02/26/tokenism-in-the-workplace/. In this context, hiring Black candidates into DEI roles without meaningfully making an effort to hire Black candidates into other kinds of roles contributes to the sense that Black employees may be "on the team" but not really "in the game." See Dana Brownlee, The Dangers of Mistaking Diversity for Inclusion in the Workplace, Forbes (Sep 15, 2019,05:00pm EDT) (discussing the perception of tokenism and the negative effect this perception has on tokenized employees).

[xl] Michael Inzlicht & Toni Schmader, Stereotype Threat: Theory, Process, and Application, Oxford University Press (2012).

[xli] See Id.; See also Claude Steele, Whistling Vivaldi: and other clues to how stereotypes affect us, W.W. Norton & Company, (2010) (discussing stereotype threat and the ways in which stereotypes can impair performance).

[xlii] See Id.

[xliii] See Id.
[xliv] See Id.
[xlv] See Graph 9, *infra*.
[xlvi] See Graph 10, *infra*.
[xlvii] See Id.
[xlviii] See Janice Gassam Asare, Are Job Candidates Still Being Penalized For Having 'Ghetto' Names?, Forbes, (Feb. 20, 2020, 11:59am EST) https://www.forbes.com/sites/janicegassam/2020/02/20/are-job-candidates-still-being-penalized-for-having-ghetto-names/?sh=6b8b87ac50ed (discussing the fears held by Black candidates that their names will be negatively judged as ghetto);
[xlix] See Derek Thompson, Busting the Myth of 'Welfare Makes People Lazy,' The Atlantic, March 8, 2018 https://www.theatlantic.com/business/archive/2018/03/welfare-childhood/555119/ (discussing the presumption that lower-wage workers in need of state assistance are lazy).
[l] See, Michael Li, To Build Less-Biased AI, Hire a More-Diverse Team, Harvard Business Review, (October 26, 2020) https://hbr.org/2020/10/to-build-less-biased-ai-hire-a-more-diverse-team.