A New Generation of Biases: Reflections from Members of the Penn Law AI & Bias Policy Lab

Penn Law
UNIVERSITY of PENNSYLVANIA LAW SCHOOL

The students in the Penn Law Policy Lab on AI and Bias present these policy briefs to industry leaders and policymakers.

Detecting a New Generation of Biases in AI-Related Recruitment Platforms

April 2021

Table of Contents

Mihir Awati

Dear United States Oversight Subcommittee on Health Care, Benefits, and Administrative Rules,

My name is Mihir Awati and I am a data researcher who is perturbed with the vast healthcare inequalities that we see in the US among different demographic groups. COVID-19 has exposed the inequities within the US healthcare system and how access to treatment and healthcare benefits are not tailored to a representative population. Paula Johnson, the president of Wellesley College, says "black Americans are disproportionately affected by the coronavirus…our health system too often gives short shrift to women and people of color".[1] These healthcare disparities occurs for a variety of reasons, but with respect to disproportionate artificial intelligence (AI) data collection in health status, it boils down to the following:

1. The health status data that we are currently collecting is unrepresentative. As just one example, the heart attack data is significantly more self-reported by white men, which leads to health policy decisions implemented that do not take into account the full population's experience and medical needs with heart attacks.
2. The importance of having a diverse data set is critical to ensure representative data and policy decisions based off of health status that better embody our population and all voices in our democracy.
3. Democracy is ultimately representation, and unrepresentative medical data goes against this principle. There is a history and legacy of discrimination in the medical system and unrepresentative data sets will only exacerbate this further, particularly in

[1] Paula A. Johnson, *A 30-year-old teacher's Covid-19 death tells us volumes*, May 31, 2020, https://www.cnn.com/2020/05/31/opinions/higher-education-coronavirus-opportunity-for-a-more-just-world-johnson/

the 21st century when we move would like to move things forward in a more equitable manner.

Thus, the past and current state of play has been less than ideal, but there is an opportunity to correct past wrongs and move forward effectively. AI plays a critical role here and it can be an effective corrective tool as it relates to health status on hiring and recruiting platforms. It is important to note that while AI has significant potential for efficiency and reducing subjectivity, there are some pitfalls that must be addressed. As Dr. Manish Raghavan indicates, biased AI can create "algorithms [that] behave in discriminatory ways simply due to negligence or insufficient attentiveness to these issues."[2]

For one thing, AI's data collection methods can be deemed a "lazy" way of gathering. It is prone to self-reporting and selection bias, as AI data samples are far more skewed towards white men than black women, as just one example. These biased data samples further exacerbate existing health inequities that exist between different demographic groups; a study in the Journal of Internal Medicine indicates black women are three times more likely to die of COVID-19 than white and Asian men.[3] These disparities are more related to social than biological factors, and if AI data sets produce biased health status data samples, these socially driven disparities will continue.

In addition, men go to the doctor far more than women because many women have other family obligations that make them unable to visit the doctor as frequently. Many women feel going to the doctor involves additional judgments about them, and many of their true physical

[2] Manish Raghavan, *Testimony of Manish Raghavan before New York City Council Committee on Technology regarding int. 1894",* November 9, 2020, https://www.cs.cornell.edu/~manish/int_1894_testimony.pdf
[3] Kate Gibson, *Black women 3 times more likely to die from COVID-19 than white men*, April 8, 2021 https://www.cbsnews.com/news/black-women-three-times-as-likely-to-die-from-covid-19-than-white-men/

symptoms are attributed to mental health instead. Achieving a more granular data set where a woman is not subject to judgment review is a goal of AI. With respect to negative personality, one example of the bias women face can be seen in the tech sector. A report from Bias Interrupters indicates in tech performance evaluations, 66% of women's performance reviews contain negative personality criticism; in contrast only 1% of men's reviews include this information.[4] This creates bias in recruiting / hiring platform's based off of women's health status that must be reduced.

The risk is that normalizing these skewed data sets that are not representative unfortunately become the norm and is then further perpetuated through AI. The impact is unrepresentative policy that has deleterious second order effects as groups left out of the data sets are not incorporated into eventual policies; this must be avoided.

Here are some specific policy recommendations I would like to see to address these issues

1. Increased telemedicine where AI can collect information for free; the benefit is AI prevents an individual from being humiliated for feeling like they are being judged (as the information is based on cold hard data and less subjective).
2. Improved public awareness education on the benefits of AI telemedicine. The impact of AI telemedicine is it can take specific symptoms a patient is reporting (for example coughing or throwing up) and effectively attribute it to an individual medical issue that needs to be resolved, rather than explaining it away due to an individuals circumstances (such as a stressful job / school situation). Historically

[4] *Bias Interrupters Tools for Individuals*, https://biasinterrupters.org/toolkits/individualtools/

outside factors have been used to explain away how they are feeling, rather than acknowledging the patient has bodily symptoms that must be treated.

3. Innovation in AI that improves the representation of data sets. Currently the unrepresentativeness of AI data is leading to skewed health policy decisions that are not representative of the population. Reducing this bias through innovation in AI data collection methods that may eliminate potential sources of bias should be a top policy goal.
4. Increased regulation and data audits by both the US Department of Health and Human Services (HHS) and the Food and Drug Administration (FDA) to ensure AI health status data sets are fully representative. Congress should provide robust funding to ensure both HHS and the FDA can perform this task effectively.

Ultimately, I urge this committee to pursue these recommendations to ultimately reduce the health status data disparities we see and utilize AI properly to achieve more equitable and representative data sets. When health policy decisions are made based off of representative health status data with less bias, ultimately not only will bias on hiring and recruiting platforms reduce, but simultaneously our health system will more accurately represent the healthcare needs in our country (which embodies good healthcare public policy).

Kimerly Biesinger

AI Methods to Decrease Bias in Hiring Practices

Introduction

Equitable hiring is critical from both legal and moral standpoints. When making hiring judgements, however, humans consistently make biased decisions. In recent years, artificial intelligence has been utilized in some elements of hiring with the hope of improving and expediting the process. Although AI assisted decision-making is not bias free, it does have the potential to decrease disparity.[9] There are many AI methods already in practice across industry which have proven useful in decreasing bias in hiring. With additional development, AI can be used to improve the fairness of hiring processes.

Decision-making processes differ between AI and humans in ways that are readily applicable to corporate hiring. Unlike humans, AI is not affected by factors such as weather, mood, or fatigue, but is totally consistent with candidate selection given the same data.[9] AI systems can also process data extremely quickly, whereas time and energy constraints force human hirers to shrink the number of candidates they seriously consider.[18] Hiring algorithms are not able to hide the reasoning behind their final choices, but humans can even fail to realize which factors and biases they use in their decisions.[22] Each of these characteristics demonstrates that the use of AI in hiring has the potential to decrease bias that permeates the hiring process. In the words of Dr. Andrew McAfee of MIT, “If you want the bias out, get the algorithms in.”[22]

Technological Solutions

AI can be utilized to decrease hiring bias by providing transparency. Unlike humans, algorithms can be forced to demonstrate which data components were used to make a decision which is critical in removing bias from hiring processes.[3,16] Such complete transparency allows human hiring managers to understand much more clearly where unfairness may exist in their hiring processes.[14] Often, further technology is necessary to interpret reasoning data from AI. Quantitative input influence tools can show how impactful each input is on the final decision of an algorithm.[5,9] In a hiring context, an algorithm can review a candidate profile and decide whether to interview or reject the candidate. The quantitative input influence tool will show which pieces of information were used by the system to make its judgement, and how much importance was placed upon each data point. Reviewing this information allows human reviewers to ensure that the AI is making judgements based on relevant information. Similarly, graphical causal models can be developed to help understand cause-and-effect pathways created while AI makes a decision on a candidate.[2] Local Interpretable Model-Agnostic Explanations (LIME) are another set of tools that can be used to explain the conclusions of a variety of AI decision-making tools. LIME can present an explanation along with each decision made about a candidate, which then allows for human reviewers to judge the validity quickly and efficiently.[19,20] Implementing these or similar methods into hiring processes allows for insightful auditing and improved equality.

In addition to finding unfairness, AI can be used to remove bias from AI hiring programs.[9] Several organizations have published open-source solutions which can be used to remove bias from algorithms such as IBM's AI Fairness 360 toolkit and Pymetrics open-source de-biasing methodology.[1,23] Companies that want to implement AI into their hiring processes can use these sources to screen and remove bias from programs and data sets. Facebook's Fairness Flow, a similar system, analyzes code to find, eliminate, and re-test for bias in AI within their own systems,[11] a process similar to those used by AI hiring companies such as HireVue.[12] The implementation of these fairness solutions can help remove any initial bias from hiring algorithms so they perform responsibly and equitably.

An increasingly common use of AI in hiring is to remove information from hiring data such as names and photographs. Hirers can use these types of data to infer information about a candidate that is typically protected during the hiring process such as race or sexual orientation.[9] By removing this information from consideration completely, hiring companies such as Unbiasify[26], Seekout[17], TalVista[24], and Slack[8] prevent potential employers from making decisions based on stereotypes. One company, Entelo, has developed a system which goes as far as to hide school names, employer names, employment gaps, and all but the most recent five years of work history to prevent discrimination against older individuals, caregivers, and other vulnerable populations.[7] Removing this data prevents both AI and human decision makers from erroneously selecting discriminatory information and stereotyping potential candidates.

AI is particularly useful for reviewing written communications throughout the hiring process, such as job listings or initial messages with candidates. These types of written communications often take place during the earliest stages of the hiring process, and biased language can prevent diverse candidates from pursuing a position. Using inclusive language prevents bias that can take place before the official hiring process begins. Chat bots powered by AI such as the one offered by the company Mya can prevent bias from pushing away potential employees in the earliest hiring stages by communicating with fairness and clarity.[6] Instead of messaging with a potentially prejudiced hiring manager, candidates can communicate with AI specifically designed to be fair and inclusive. AI can also be used to highlight jargon and language that is gendered, ageist, or ableist in posted listings.[10,24,13,25] Removing bias in these earlier stages of the hiring process helps companies find and recruit more diverse candidates.

Alternative, more inclusive testing methods are made possible with AI as well. Pymetrics is one hiring company which has developed games to measure attributes and skills of candidates. The employees of a company play the games which allows the AI to learn which attributes are found in successful employees. That data is then used to find candidates who are a good fit based on their own game scores. These attributes are measured without having to consider ethnicity, gender, or work history.[15] Unique methods such as these combined with de-biasing technology allows for candidates with potential to demonstrate their skill set in more concrete ways.[23] Utilizing AI can help candidates with differing backgrounds be considered based on their abilities and potential more fully.

Conclusion

AI technology has potential to decrease bias in hiring which will only grow as methods and technologies improve. Although AI can be a valuable tool for fairness, it is essential that extreme caution be exercised in the implementation of these methods. Without careful and continuous review, bias is easily overlooked in AI decision-making as has been demonstrated repeatedly by systems implemented for hiring in the past.[4] Constant review and improvement is absolutely necessary to continue progress. Although AI hiring tools improve on existing methods and may perform well enough for legal purposes, there is no “good enough” when considering subconscious discrimination. These processes must be refined repeatedly as our understanding of fairness changes and develops over time.[16,21]

References

[1] AI Fairness 360. (n.d.). Retrieved April 04, 2021, from https://aif360.mybluemix.net/

[2] Chiappa, S., & Gillam, T. P. (2019). Path-specific counterfactual fairness. *Proceedings of the AAAI Conference on Artificial Intelligence, 33*, 7801-7808. doi:10.1609/aaai.v33i01.33017801

[3] Civin, D. (2018, May 21). Explainable AI Could Reduce the Impact of Biased Algorithms. Retrieved April 04, 2021, from https://venturebeat.com/2018/05/21/explainable-ai-could-reduce-the-impact-of-biased-algorithms/

[4] Dastin, J. (2018, October 10). Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women. Retrieved April 04, 2021, from https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

[5] Datta, A., Sen, S., & Zick, Y. (2016). Algorithmic Transparency via Quantitative Input Influence: Theory and Experiments with Learning Systems. *2016 IEEE Symposium on Security and Privacy (SP)*. doi:10.1109/sp.2016.42

[6] Deshmukh, A. (n.d.). Your AI Recruitment Tools Pocket Guide: Everything You Need to Know. Retrieved April 04, 2021, from https://www.mya.com/ai-recruitment-tools/

[7] Diversity Recruiting Software. (n.d.). Retrieved April 04, 2021, from https://www.entelo.com/products/platform/diversity/

[8] Grace, J. (n.d.). A Walkthrough Guide to Finding an Engineering Job at Slack. Retrieved from https://slack.engineering/a-walkthrough-guide-to-finding-an-engineering-job-at-slack/

[9] Houser, K. (2019). Can AI Solve the Diversity Problem in the Tech Industry? Mitigating Noise and Bias in Employment Decision-Making. *Stanford Technology Law Review*, *290*. Retrieved from https://law.stanford.edu/wp-content/uploads/2019/08/Houser_20190830_test.pdf

[10] Humanly Diversity. (n.d.). Retrieved April 04, 2021, from https://humanly.io/diversity

[11] Kloumann, I., & Tannen, J. (2021, March 31). How We're Using Fairness Flow to Help Build AI that Works Better for Everyone. Retrieved from https://ai.facebook.com/blog/how-were-using-fairness-flow-to-help-build-ai-that-works-better-for-everyone/

[12] Larsen, L. (2018, June 20). HireVue Assessments and Preventing Algorithmic Bias. Retrieved from https://www.hirevue.com/blog/hiring/hirevue-assessments-and-preventing-algorithmic-bias

[13] McCormic, K. (2020, March 26). Bias Interruption Beyond Gender. Retrieved April 04, 2021, from https://textio.com/blog/bias-interruption-beyond-gender/27525611553

[14] Mission. (n.d.). Retrieved April 04, 2021, from https://www.pymetrics.ai/mission

[15] Pymetrics. (n.d.). Retrieved April 04, 2021, from https://www.pymetrics.ai/candidates#client-advisory-council

[16] Raghavan, M., Barocas, S., Kleinberg, J., & Levy, K. (2020). Mitigating Bias in Algorithmic Hiring: Evaluating Claims and Practices. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. doi:10.1145/3351095.3372828

[17] Remove Unconscious Bias with Blind Hiring Mode (And How to Change the Cat Picture). (n.d.). Retrieved April 04, 2021, from https://support.seekout.io/hc/en-us/articles/360054116291-Remove-unconscious-bias-with-Blind-Hiring-Mode-and-how-to-change-the-cat-picture

[18] Removing Bias from Talent Decisions with Artificial Intelligence. (n.d.). Retrieved April 04, 2021, from http://go2.pymetrics.ai/l/863702/2020-09-14/p4yqt/863702/42452/RemovingBiasFromTalentDecisionswithAI_FINAL.pdf

[19] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. doi:10.1145/2939672.2939778

[20] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August 12). Local Interpretable Model-Agnostic Explanations (LIME): An Introduction. Retrieved from https://www.oreilly.com/content/introduction-to-local-interpretable-model-agnostic-explanations-lime/

[21] Ruab, M. (2018). Bots, Bias, and Big Data: Artificial Intelligence, Algorithmic Bias and Disparate Impact Liability in Hiring Practices. *Arkansas Law Review, 71*(2). Retrieved from https://scholarworks.uark.edu/alr/vol71/iss2/7

[22] Silberg, J., & Manyika, J. (2019, June 6). Tackling Bias in Artificial Intelligence (And in Humans). Retrieved April 04, 2021, from https://www.mckinsey.com/featured-insights/artificial-intelligence/tackling-bias-in-artificial-intelligence-and-in-humans

[23] Solutions. (n.d.). Retrieved April 04, 2021, from https://www.pymetrics.ai/solutions

[24] TalVista. (n.d.). Retrieved April 04, 2021, from https://www.talvista.com/

[25] Textio. (n.d.). Textio. Retrieved April 04, 2021, from https://textio.com/products/tone/

[26] Unbiasify. (n.d.). Retrieved April 04, 2021, from https://chrome.google.com/webstore/detail/unbiasify/affijhegklbkdinpepgphhlgphnhbenk

Emily Friedman
Socioeconomic Status Bias

There are several characteristics—such as gender, race, and religion—that we all agree should play no role in hiring decisions. Employers are encouraged to judge based on "neutral" characteristics thought to be relevant to an applicant's ability to do the job—experience, education, and "fit" in an organization.[1] In reality these characteristics are largely determined by a person's socioeconomic status (SES). Further, new research shows that not only is use of these characteristics discriminatory, but they are also poor predictors of job performance.[2] Artificial intelligence (AI) has the potential to reduce the impact of SES-tainted characteristics in hiring and make the entire process fairer. However, without intervention, AI can replicate existing patterns of discrimination and create additional barriers for low-SES applicants.

Socioeconomic Status Bias in Hiring

The traditional hiring process generally begins with the solicitation and sorting of resumes. Recruiters typically spend only a few seconds on each resume (if they look at them at all), scanning for certain key qualifications.[3] They rarely apply standardized scoring methods, instead basing their evaluations on an "often-flawed combination of pedigree and gut instinct."[4]

Socioeconomic status manifests itself in several ways on an applicant's resume, and can be a large determinant of hiring decisions. For "high status" jobs, prestige of university attended is the first factor used to screen applicants.[5] Elite employers limit their recruiting to the most

---

[1] *See, e.g.*, Sigal Alon, *The Evolution of Class Inequality in Higher Education: Competition, Exclusion, and Adaptation*, 74 AM. SOC. REV., 731, (2009) ("[E]xclusion on the basis of educational credentials is accepted as more legitimate than exclusion on the basis of race, ethnicity, religion, or gender because it appears to be based on individual merit.").

[2] *See* Josh Constine, "Pymetrics Attacks Discrimination in Hiring with AI and Recruiting Games," *TechCrunch* (blog), accessed Apr. 6, 2020, http://social.techcrunch. com/2017/09/20/unbiased-hiring/ (explaining research that found very little correlation between resume qualifications and job performance).

[3] Laura A. Rivera, *Ivies, Extracurriculars, and Exclusion: Elite Employers' Use of Educational Credentials*, 29 RES. IN SOC. STRATIFICATION AND MOBILITY, 71, 77 (2011).

[4] FILIPPO A. RASO ET AL., ARTIFICIAL INTELLIGENCE & HUMAN RIGHTS: OPPORTUNITIES & RISKS 46 (Berkman Klein Center for Internet & Soc'y at Harv. U., 2018).

[5] Rivera, *supra* note 3 at 75.

prestigious universities and will rarely even consider applicants from lower-ranked universities.[6] As one investment bank recruiter explained, resumes of applicants from non-elite schools "pretty much go[] into a black hole… You need to know someone, you need to have a connection, you need to get someone to raise their hand and say, 'Let's bring this candidate in.'"[7] Thus, only applicants from elite universities and those from lower-ranked universities but are part of elite networks are considered.

While employers highly value attendance at an elite university, they do not generally believe that these universities better prepare students for employment.[8] Some value prestige for its own sake, while some believe that top universities better "groom" their students for admittance into elite circles.[9] Others see the prestige of the university attended as a proxy for an applicant's intelligence.[10] In reality, parental SES is one of the largest determinants of whether a student will attend an elite university.[11] Many hiring professionals give no consideration to how SES affects admission to elite universities or financial constraints on choice.[12] Even the ones who recognize that some lower-SES students may have chosen to go to lower-ranked university

---

[6] *Id.* at 75-76.
[7] *Id.* at 76.
[8] *See id.* at 72 ("Employers value prestige of the school over "the length (e.g., number of years of schooling) or content (e.g., degrees completed, coursework taken, skills acquired) of education," even though these factors are arguably more important determinants of whether applicants are qualified.") Some applicants even think elite universities do a worse job preparing students for the work they will be doing: "evaluators tended to believe that elite and, in particular, super-elite instruction was 'too abstract,' 'overly theoretical,' or even 'useless' compared to the more 'practical' and 'relevant' training offered at 'lesser' institutions." *Id.* at 78.
[9] *Id.* at 72, 80.
[10] *Id.* at 78.
[11] Alon, *supra* note 1 at 740. Some believe that lower-SES students tend to attend lower-ranked universities because of inferior academic ability. However, "socioeconomic gaps in higher education persist even when accounting for differences in academic ability and other background characteristics." *Id.* Sigal Alon explains that higher SES students have superior access to elite universities because they tend to attend better high schools with a larger focus on college admissions, have higher test scores because of expensive test prep tools, and have parental assistance in the admissions process. *Id.* at 736-37.
[12] Rivera, *supra* note 3 at 79.

because they received scholarships there, they penalize students for not attending an elite university because they "should be smart enough to invest in their future."[13]

Participation in extracurricular activities is the second qualification that elite employers look to when screening resumes.[14] Extracurriculars are so important that applicants are unlikely to be called in for an interview without them.[15] Many employers consider extracurriculars to be more important than prior work experience or grades.[16] These employers see participation in extracurriculars as an indication of an applicant's social skills, time-management skills, and initiative.[17] This focus on activities disadvantages lower-income applicants. Lower-SES grade school students are less likely to participate in organized sports and non-athletic extracurriculars and were less likely to have leadership roles within these activities.[18] As a result, they do not have the experience or training necessary to participate in these extracurriculars at college or graduate school. Further, some extracurriculars, such as tennis and crew, are expensive, and lower-SES students might not have the financial resources to participate.[19] Additionally, lower-SES students who have paid jobs in addition to attending school full-time are less likely to have the time to invest in these activities. Employers tend to value paid work less than extracurricular activities: one study found that while students with jobs "often received 'points' for their 'work ethic' from sympathetic evaluators, they still were often penalized on the dimensions of "interestingness," sociability, and 'well- roundedness' because they had fewer 'activities.'"[20]

---

[13] *Id.* (quoting an attorney).
[14] *Id.* at 72.
[15] *Id.* at 82.
[16] *Id.*
[17] *Id.* at 82-83.
[18] Kaisa Snellman et al., *The Engagement Gap: Social Mobility and Extracurricular Participation among American Youth*, 657 THE ANNALS OF THE AM. ACAD. OF POL. AND SOC. SCI. 194, 199-201 (2015).
[19] Rivera, *supra* note 3 at 83.
[20] *Id.* at 84.

Exacerbating this class-discrepancy, employers tend to value exclusive, traditionally "upper-class" extracurriculars over activities with lower barriers to participation. Research shows that employers prefer "activities that were time-and resource-intensive because the investment such cultivation entailed indicated stronger evidence of 'drive' and an orientation towards 'achievement' and 'success.'"[21] These are the very types of extracurriculars that lower-SES students are least likely to be able to participate in because they require time and financial investment since childhood."[22] For example, one investment banker said that he would prefer a candidate who went to Costa Rica to build houses for Habitat for Humanity over a candidate who delivered meals for meal on wheels because the former was a "true accomplishment" while the latter was something "anyone could do."[23]

Research also shows that class-based stereotypes of extracurricular activities have a significant effect on hiring prospects independent of time-commitment or skill required. In one study, researchers sent resumes to law firms that were identical except some listed traditionally "upper-class" activities (sailing, polo, and classical music) and some listed "lower-class activities (track and field, soccer, and country music).[24] 16.25% of the men with "upper-class" activities were called back for interviews compared to 1.28% of the "lower-class" men.[25] These results suggest that employers not only prefer activities that happen to be more prevalent among higher-SES applicants, but they also prefer these activities *because* they are more prevalent among higher-SES applicants.

---

[21] *Id.* at 83.
[22] *Id.*
[23] *Id.*
[24] Laura Rivera & András Tilcsik, *Research: How Subtle Class Cues Can Backfire on Your Resume*, HARV. BUS. REV. (Dec. 21, 2016), https://hbr.org/2016/12/research-how-subtle-class-cues-can-backfire-on-your-resume.
[25] *Id.* The results also revealed intersectional effects of gender and class: while men with upper-class activities were advantaged, women with upper-class activities were penalized. *Id.*

Some employers highly value an applicant's prior work experience. Many said that they would choose a candidate with a prestigious internship and a lower GPA over a candidate with a 4.0 who did not have impressive internship experience.[26] At the same time, these employers "placed no value having a 'work for money' job such as waitressing or cashiering."[27] These preferences disadvantage lower-SES students, who often do not have the elite connections necessary to secure a prestigious unpaid internship, and who cannot accept them even if they receive an offer because they must earn money over the summer to support themselves.

At the interview stage, class manifests itself in new ways. The first thing interviewers may notice about candidates is their appearance, which can trigger unconscious biases. "Beauty bias" is defined as "the favorable treatment that individuals receive when they are deemed more attractive, regardless of whether this happens consciously or unconsciously."[28] People tend to perceive attractive people as more intelligent, talented, social and honest[29]—qualities that employers look for in candidates. Studies confirm that attractive people are more likely to be hired.[30] Physical attractiveness is not necessarily determined by class. However, there are several characteristics that tend to be viewed as unattractive and are associated with lower-SES—for example, being overweight,[31] having tattoos,[32] and having broken and crooked teeth.[33] One researcher explained that "[m]ore than any other marker in America, teeth indicate class

---

[26] *Uncovering Bias: A New Way to Study Hiring Can Help*, KNOWLEDGE@WHARTON (July 18, 2019), https://knowledge.wharton.upenn.edu/article/uncovering-hiring-bias/.
[27] *Id.*
[28] Tomas Chamorro-Premuzic, *Attractive People Get Unfair Advantages at Work. AI Can Help*, HARV. BUS. REV. (Oct. 31, 2019).
[29] *Id.*
[30] *Id.*
[31] *Id.*; Jeffery RW et al, *Socioeconomic Status Differences in Health Behaviors Related to Obesity: The Healthy Worker Project*, 15 INT. J. OF OBESITY 689, 689 (1991) (finding an inverse correlation between SES and obesity).
[32] Chamorro-Premuzic, *supra* note 28; Eric Silver et al., *Bodily Signs of Academic Success: An Empirical Examination of Tattoos and Grooming*, 58 SOC. PROBLEMS 538, 538 (2011) (finding that adolescents with tattoos are less likely to go to college).
[33] Daneli Evans Peterman, *Socioeconomic Status Discrimination*, 104 VA. L. REV. 1238, 1322 (2018).

status…. [B]eing too poor to have respectable teeth is like wearing an 'L' for loser on your face."[34] Thus, lower SES applicants who are unable to invest as much in their appearance are at a disadvantage in the hiring process.

As soon as the conversation begins, before the interviewer even has time to process the content of the interviewees' answers, the interviewer is making unconscious judgments. After hearing a speaker say just seven random words, the listener can determine the speaker's social class with relatively high accuracy.[35] This is problematic because interviewers tend to rate the applicants who "sound" higher class as more qualified for the job, even if they do not realize that perceived social class is influencing their perception.[36] Researchers concluded that "while most hiring managers would deny that a job candidate's social class matters, in reality, the socioeconomic position of an applicant or their parents is being assessed within the first seconds they speak — a circumstance that limits economic mobility and perpetuates inequality."[37]

Potential for AI to Mitigate Class Bias in Hiring

The use of AI in the hiring process can reduce class bias. As AI researchers explain, "[a]utomated hiring systems promise to evaluate job candidates on the basis of their bona fide qualifications, rather than on qualities…that often lead human decision-makers astray."[38] First, AI can find more socioeconomically diverse candidates than traditional recruiting methods can. Elite employers say that they only recruit at elite universities because it is "'time' and 'cost'

---

[34] *Id.* (quoting Susan Sered, What Pennsatucky's Teeth Tell Us About Class in America, Bitch Media (July 1, 2014), https://www.bitchmedia.org/post/what-pennsatucky%E2%80 %99s-teeth-tell-us-about-class-in-america [https://perma.cc/L9FV-GCG5]).
[35] Mike Cummings, *Yale Study Shows Class Bias in Hiring Based on Few Seconds of Speech*, YALENEWS (Oct. 21, 2019), https://news.yale.edu/2019/10/21/yale-study-shows-class-bias-hiring-based-few-seconds-speech.
[36] *Id.*
[37] *Id.*
[38] RASO ET AL., *supra* note 4 at 7.

saving, while wading through 'lower caliber' candidates to find 'diamonds in the rough,' [is] considered wasteful."[39] AI programs can scour the internet to find "passive candidates" who are not actively targeted by recruiters.[40] They can identify candidates from overlooked sources without requiring the human recruiter to expend any additional time or effort on recruiting. Further, these programs can reduce the importance of connections and referrals in the hiring process.[41]

AI programs that automate the resume screening process can also lead to the consideration of more socioeconomically diverse candidates. Employers say time-constraints are the main reason they do not look at the resumes of candidates from lower-ranking universities.[42] Applicant tracking software can analyze exponentially more resumes than human recruiters can, increasing employers' ability to consider candidates from a broader range of sources.[43]

In addition to evaluating *more* candidates, AI can evaluate candidates *more consistently* than human recruiters. Humans decisions are often made based on the amorphous and biased concept of "gut instinct" and two individuals evaluating the same candidates often reach different conclusions.[44] In contrast, resume and interview-evaluating AI algorithms apply consistent evaluative standards.[45] This means, for example, that a Princeton graduate cannot informally give preference to other Princeton graduates in the hiring process. One program in particular, Mya, is a chatbot that interviews candidates and evaluates their answers based on

---

[39] Rivera, *supra* note 3 at 81.
[40] Falon Fatemi, *How AI is Uprooting Recruitment*, FORBES (Oct. 31, 2019), https://www.forbes.com/sites/falonfatemi/2019/10/31/how-ai-is-uprooting-recruiting/?sh=4cdae7c646ce.
[41] *Id.*
[42] *See supra* note 39 and accompanying text.
[43] Fria Polli, Using AI to Eliminate Bias from Hiring, HARV. BUS. REV. (Oct, 29, 2019), https://hbr.org/2019/10/using-ai-to-eliminate-bias-from-hiring.
[44] Kimberly A. Houser, *Can AI Solve the Diversity Problem in the Tech Industry? Mitigating Noise and Bias in Employment Decision-Making*, 22 STAN. TECH. L. REV. 290, 330 (2019).
[45] *Id.* at 330-31.

predetermined criteria.[46] Kimerly Houser explains how Mya is reducing bias in the hiring process:

> By using chatbots to conduct interviews companies eliminate the variability that a human interviewer or multiple interviewers would bring into the process. Unlike humans, a chatbot asks each interviewee the same set of questions. In addition, by setting criteria for promotions in advance, using an algorithm to assess employees will reduce both the bias of managers and noise in the decision by applying rules uniformly.[47]

AI allows for the consideration of a higher quantity of applicants without sacrificing the quality of the screening process.

Further, AI can be used to evaluate candidates based on less classist criteria. Resume-evaluating algorithms can be "programmed to ignore details such as the names of schools attended and zip codes, that can be proxies for race and socioeconomic status."[48] One of the most promising innovations is Pymetrics, an AI program that evaluates candidates based on characteristics that are more relevant to the job and are less tainted by SES.[49] Pymetrics works by having the top performers at a company play a set of neurological games.[50] Then, it creates a profile of the ideal candidates based on top employees' scores.[51] Applicants then play the same games, and those whose scores resemble the scores of the top employees are ranked the highest.[52] As the CEO of Pymetrics explains, the program "doesn't preference white guys from elite schools who were on the sailing team just like the recruiter."[53]

---

[46] *Id.* at 327.
[47] *Id.* at 332.
[48] Ji-A Min, *5 Intriguing AI Recruiting Innovations*, IDEAL (Apr. 25, 2017), https://ideal.com/ai-recruiting-innovations/.
[49] Josh Constine, *Pymetrics Attacks Discrimination in Hiring with AI and Recruiting Games*, TECHCRUNCH (Sept. 20, 2017), https://techcrunch.com/2017/09/20/unbiased-hiring/.
[50] *Id.*
[51] *Id.*
[52] *Id.*
[53] *Id.* (quoting Frida Polli).

AI can also reduce the beauty and speech bias that permeates the interview process. As discussed above, Mya evaluates candidates' answers without being influenced by a candidate's appearance or voice.[54] Similarly, HireVue conducts video interviews of candidates and evaluates their answers and body language.[55] These and similar programs can infuse more consistency and fairness into job interviews, which are typically highly subjective and influenced by class-bias.

Potential for AI to Perpetuate Class Bias in Hiring

Despite the enormous potential for AI to reduce class-bias in the hiring process, it also has the potential to perpetuate or even exacerbate existing class-bias. AI experts believe that "AI can easily perpetuate existing patterns of bias and discrimination, since the most common way to deploy these systems is to 'train' them to replicate the outcomes achieved by human decision-makers."[56] Although there has been relatively little research on AI and class bias in hiring, conclusions can be extrapolated from research on gender and racial bias.

The first step in creating an algorithm to predict which job applicants will be most successful is to identify successful employees and their characteristics.[57] The definition of "success" can be influenced by class-bias. For example, "if a good employee is defined as never being late to work, then this definition will negatively impact those in a lower socioeconomic class because some employees might choose to live farther away from work where the rent is cheaper;"[58] if top performers are identified as those who have the highest performance reviews, and performance reviews are influenced by beauty or speech bias, lower-SES employees will be

---

[54] *See supra* note 46 and accompanying text.
[55] McKenzie Raub, *Bots, Bias and Big Data: Artificial Intelligence, Algorithmic Bias and Disparate Impact Liability in Hiring Practices*, 71 ARK. L. REV. 529, 538 (2018).
[56] RASO ET AL., *supra* note 4 at 7.
[57] See Raub, supra note 5555 at 534 ("Essentially, what makes a 'good' employee 'must be defined in ways that correspond to measurable outcomes.'").
[58] Yash Kanoogo, *Addressing Bias in HR Algorithms*, MEDIUM (Mar. 18, 2020), https://medium.com/@yashkanoongo/addressing-bias-in-hr-algorithms-2b0f9003ed64.

rated as less successful.[59] When top-performers are identified in a discriminatory fashion, the AI model will identify irrelevant and biased criteria as predictors of success: one AI resume evaluation program was fed biased information about who top-performers were, and thus it identified being named "Jared" and playing high school lacrosse as predictors of job performance.[60]

Relatedly, a phenomenon known as "garbage in, garbage out" can cause AI to replicate societal biases.[61] When the training data is skewed because of past discrimination, the algorithm will replicate this discrimination.[62] Houser explains that "when your data sets contain little or no information about certain groups of people, your algorithm will not accurately evaluate people who belong to that group."[63] For example, if a resume-evaluating algorithm is trained with resumes that are mostly from ivy-league applicants, the algorithm may preference candidates with ivy-league degrees.[64] Similarly, if video interview AI programs are trained with videos of conventionally attractive employees with upper-class speech patterns, the AI algorithm might continue to penalize lower-SES applicants who lack these characteristics.[65]

Further, the use of AI can lend legitimacy to hiring decisions that are actually based on the same biases that influence people's hiring decisions. Researchers call this insidious phenomenon the "veneer of objectivity": [T]he "veneer of objectivity" around high-tech systems

---

[59] *See* MIRANDA BOGEN & AARON RIEKE, HELP WANTED: AN EXAMINATION OF HIRING ALGORITHMS, EQUITY, AND BIAS 8 (Upturn 2018) (finding that when biased performance reviews that favored men were used to determine who the top-performers were, the algorithm predicted that men were more likely to be successful than women).
[60] *Id.*
[61] Houser, *supra* note 4444 at 333.
[62] RASO ET AL., *supra* note 4 at 18.
[63] Houser, *supra* note 44 at 336.
[64] *See id.* at 335 (finding that when algorithm was trained with mostly male resumes, it rated male resumes higher than female resumes).
[65] See *id.* (finding that when facial recognition software is trained with data that disproportionally comes from white males, it will have a higher error rate when presented with female and Black faces).

in general can obscure the fact that they produce results that are no better, and sometimes much worse, than those hewn from the 'crooked timber of humanity.'"[66] Without intervention, AI can perpetuate the vicious cycle of class discrimination in hiring. More research is needed to understand and eliminate the impact of SES on AI in the hiring process.

[66] RASO ET AL., *supra* note 4 at 7.

Germaine Grant

AI Policy, Bias and Ageism

It is estimated that in just five years 25% of all workers will be over 55 - a percentage that has doubled over the last quarter-century.[1] With the average age of retirement constantly rising, more people are looking to either stay in the workforce longer or re-enter it when they're over fifty-five.[2] In fact, nearly half of the people aged 55-64 will exit and re-enter the workforce during this time in their lives.[3] Age discrimination in AI is one of the more blatant flaws in terms of technological advancements. With society continuously moving towards a more digital regime, those who lack the technical skills to navigate the new online world are left behind and often excluded. Given the nature of bias, it can be assumed that a group that implicitly struggles with technology would face exclusion from the coding that is involved with creating AI. Though civil discourse and law has built in mechanics to protect elderly citizens, more work and research needs to be done within the AI space. This can be seen in recruiting, in which AI bias is a problem but also a solution.

AI age bias is a systematic problem influenced by the societal belief that older employees are not as valuable as younger employees.[4] This has been exasperated by the technology boom of the early twenty-first century and onward.[5] This trend has been further heightened because of the weakened economy with Covid-19 and the need for online remote work.[6] In fact, the U.S.

---

[1] "Labor Force Participation: What Has Happened since the Peak? : Monthly Labor Review." U.S. Bureau of Labor Statistics, U.S. Bureau of Labor Statistics, 1 Sept. 2016, www.bls.gov/opub/mlr/2016/article/labor-force-participation-what-has-happened-since-the-peak.htm.

[2] *Id.*

[3] *Id.*

[4] "The Case for Hiring Older Workers." Harvard Business Review, 26 Sept. 2019, hbr.org/2019/09/the-case-for-hiring-older-workers.

[5] *Id.*

[6] "Age Discrimination and COVID-19: What to Do When Employees Are in High-Risk Groups." The National Law Review, www.natlawreview.com/article/age-discrimination-and-covid-19-what-to-do-when-employees-are-high-risk-groups.

Bureau of Labor Statistics noted that long-term unemployment for working Americans 55 and older spiked to 26.4% from 14% last September vs. an increase to 18.2% from 11.3% outside that category.[7] Even though rapidly advancing technology is one of the main barriers for elderly citizens, AI also has the ability to lower the entry barrier of technology. One such innovation is conversational AI. Conversational AI is the technology that makes that possible. It allows artificial intelligence technologies like chatbots to interact with people in a humanlike way. By bridging the gap between human and computer language, it makes communication between the two easy and natural.[8] AI technology experts believe that conversational AI can help ensure the safety of older talent as employees return to workplaces without ageism causing harm, especially post-pandemic.[9] AI technology can be helpful as it " bases talent-sourcing decisions solely on skills, attributes and performance indicators, not physical characteristics."[10] As a result, the playing field can be leveled for millions of older workers whose depth of experience and talent is often overlooked by recruiters.[11] This concept was exemplified in the article "Ageism in Technology Hiring: Can A.I. Stop It for Good?" where the pros and cons of AI on ageism were explored. They found that "A.I. could take the whole hiring pipeline into account, making companies consider candidates who might have gone unseen (or quickly disregarded) in other circumstances."[12] This is premised on using AI to exclude the use of ageist criteria and

---

[7] Shutan, Bruce. "AI Can Identify Age Discrimination in Recruiting." Employee Benefit News, Employee Benefit News, 1 Apr. 2021, www.benefitnews.com/news/ai-can-identify-age-discrimination-in-recruiting.

[8] "What Is Conversational AI and How Does It Work?" Bold360, www.bold360.com/learn/what-is-conversational-ai#:~:text=Conversational%20AI%20is%20the%20technology,the%20two%20easy%20and%20natural.

[9] Shutan, Bruce. "AI Can Identify Age Discrimination in Recruiting." Employee Benefit News, Employee Benefit News, 1 Apr. 2021, www.benefitnews.com/news/ai-can-identify-age-discrimination-in-recruiting.

[10] *Id.*

[11] *Id.*

[12] Nick Kolakowski November 5, 20197 min read. "Ageism in Technology Hiring: Can A.I. Stop It for Good?" Dice Insights, 5 Nov. 2019, insights.dice.com/2019/11/05/ageism-technology-hiring-stop-good/.

constantly exploring which standards companies have that have a disparate impact on elderly applicants.

Another aspect of age discrimination is that the older one gets the more likely one is to be disabled, which is proven to lead to discrimination.[13] This leads to a complex between age discrimination and ableism. The risk of arthritis, strokes, vision problems, and hearing problems increase with age, each of which could qualify as a disability. While age discrimination and disability discrimination used to be viewed independently (and still occur independently), there is a strong link between the two.[14] A solution to this is focusing on the hiring phase where older workers exit and re-enter the workforce. Using AI in this capacity should significantly impact hiring strategies and create more age-diverse workspaces and cultures. AI and automation help companies identify potential biases in their hiring patterns.[15] One concrete example is that it can help companies move beyond age restrictions or a candidate's "years experience" as criteria, which has been shown to hurt both younger and older applicants and cause a less diverse candidate pipeline.[16] Capricious age restrictions eliminate large pools of qualified workers.[17] This not only promotes bias but undercuts efforts to find the best candidate available for a job.[18] AI

---

[13] "Disability Discrimination." Disability Discrimination | Equality and Human Rights Commission, www.equalityhumanrights.com/en/advice-and-guidance/disability-discrimination.

[14] Spencer, Jennifer, and Jennifer Spencer. "Age and Disability Discrimination & Your Rights." Jackson Spencer Law, 5 Mar. 2020, jacksonspencerlaw.com/age-and-disability-discrimination/#:~:text=While%20age%20discrimination%20and%20disability,more%20than%20one%20federal%20law.).

[15] Blank, Andres. "How AI & Mindfulness Can Tackle Age Bias in the Modern Workplace." HR Technologist, www.hrtechnologist.com/articles/ai-in-hr/how-ai-mindfulness-can-tackle-age-bias-in-the-modern-workplace/.

[16] *Id.*

[17] Institute, The McQuaig. "Artificial Intelligence In A New Decade." Medium, Medium, 19 Feb. 2020, medium.com/@McQuaig/artificial-intelligence-in-a-new-decade-4895069602.

[18] Blank, Andres. "How AI & Mindfulness Can Tackle Age Bias in the Modern Workplace." HR Technologist, www.hrtechnologist.com/articles/ai-in-hr/how-ai-mindfulness-can-tackle-age-bias-in-the-modern-workplace/.

can also help businesses work around other proxies for ageism, like salary requirements, which can unintentionally filter out older applicants.[19]

In conclusion, removing ageism will require a willingness to unlearn concepts about older workers and a desire to use new strategies and systems in correcting these issues. Correcting these systemic wrongs will not be done quickly, however, hopefully attitudes should eventually shift to reflect changing demographics. Greater recognition of the issue, advanced technologies and a willingness to fight suggests that we can and will ultimately make progress. Until then, it is our obligation to continuously update AI as to allow proper recruiting methods and employment for elderly citizens.

---

[19] *Id.*

Marleca Higgs

AI and Implicit Bias: How do I Look?

The use of algorithms in hiring and recruiting has become increasingly popular. Employers use algorithms in hopes of reducing the time needed to find the "perfect" candidate for the job. These algorithms can operate in a variety of ways during the hiring and recruiting process. Some function to direct job advertisements toward certain candidates, while others flag passive candidates for recruitment.[1] Other algorithms function as predictive tools parse and score resumes, and help hiring managers assess candidate competencies in new ways, using both traditional and novel data.[2] The months previously spent analyzing and evaluating hundreds or thousands of applicants could now be reduced to a few hours. Employers claim that the hiring process is a "drain on organizational productivity."[3] Others utilized algorithms in hopes that the biases often present in hiring and recruiting would be eliminated. Many thought that the tendency to choose applicants based on factors irrelevant to the job's qualifications would be significantly reduced, or eliminated, because of the consistency of algorithms. By removing human decision makers, employers earnestly believed that the process and the outcomes would be more fair and equitable for all.

But algorithms also present their own challenges—or some of the very same challenges—in hiring and recruiting. These systems and the data they are fed are still very much based on institutional and historical biases.[4] In a Harvard Business Review article, Miranda Bogen

---

[1] Miranda Bogen, *All the Ways Hiring Algorithms Can Introduce Bias*, Harvard Business Review (May 2019), https://hbr.org/2019/05/all-the-ways-hiring-algorithms-can-introduce-bias

[2] Id.

[3] Antonio Simental, *Higher Standards for Hire: Algorithmic Bias in the Job Application Process*, Business Today Online Journal (Jan. 2021), https://journal.businesstoday.org/bt-online/2020/higher-standards-for-hire-algorithmic-bias-in-the-job-application-process

[4] Bogen, *Supra* note 1

mentions that these algorithms drift towards bias by default.[5] She argues that these technologies fail to “tackle deeper disparities [that]…can help promote equity, rather than erode it.”[6] Where these algorithms are not properly vetted, they can ultimately lead to reinforcement of stereotypes, bias, and discrimination. In hiring and recruiting, bias algorithms can lead to disastrous results. For example, a class of people can be deemed unqualified and automatically rejected by the system. Or, in recruiting, the system may only show a job advertisement to certain groups of people. In 2019, Facebook advertisements for a cashier position were shown to an audience that was 85% women, while taxi driver advertisements were shown to an audience that was 75% black.[7] These algorithms not only impacts who will get selected for the job—but it determines who will even have access to hiring process.

Additionally, though many assume that reduced human decision will lead to less bias, this is not truly the case. These algorithms, in part, function based on the human-fed data; algorithms can also learn the preferences of the employer.[8] Job boards like ZipRecruiter learns the recruiters’ preferences—and then uses those preferences to solicit similar applicants.[9] Where a recruiter typically chooses white men or ivy league graduates, the system will replicate that behavior. This undercuts any noble efforts that companies may have to increase diversity and inclusion in the work place. Unfortunately, where a company has historically engaged in biased hiring, the algorithms will simply replicate that outcome. “… A hiring algorithm will be more likely to pick someone who resembles the company’s previous employees, thus perpetuating racism and discrimination.”[10]

---

[5] Id.
[6] Id.
[7] Id.
[8] Id.
[9] Id.
[10] Simental, Supra note 3

In other cases, algorithms may be using arbitrary information or data with little to no relevance to the job description. This is particularly true with appearance. HireVue is a company that uses algorithms to enable a more efficient hiring process. HireVue video interviewing process has been the subject of various disputes and litigation. Their video interviewing algorithms, in part, evaluates candidates based on their facial expressions/features, mannerisms, tone, etc.[11] The candidate is then given an employability score.[12] Though Kevin Parker, the CEO of HireVue claims that the overwhelming majority of the assessment focuses on word choice/language, appearance is still a part of a candidate's evaluation.[13] In 2019, a complaint was filed with the Electronic Privacy Information Center, which claimed that the algorithms were ultimately evaluating emotions and facial expressions differently based on race, gender, or sexual orientation.[14] One article discusses how this facial/appearance analysis "could cost the careers of countless candidates."[15] The article goes on to mention that facial analysis software can perpetuate existing biases for overweight candidates—and also give an advantage to candidates that are "more physically attractive."[16]

HireVue claimed that their video interviewing system tools are "more predictive of job performance than human interviewers conducting the same structured interviews."[17] But the development of programs like this has disturbed many. Meredith Whittaker, co-founder of AI

---

[11] Dave Zielinski, Addressing Artificial Intelligence—Based Hiring Concerns, HR Today (May 2020), https://www.shrm.org/hr-today/news/hr-magazine/summer2020/Pages/artificial-intelligence-based-hiring-concerns.aspx
[12] Drew Harwell, A Face-scanning algorithm increasingly decides whether you deserve the job, The Washington Post (Nov. 2019), https://www.washingtonpost.com/technology/2019/10/22/ai-hiring-face-scanning-algorithm-increasingly-decides-whether-you-deserve-job/
[13] Zielinski, Supra note 11
[14] Id.
[15] The Potential Biases Caused by Machine Learning and AI Algorithms, We Pow, https://www.wepow.com/en/blog/the-potential-biases-caused-by-machine-learning-and-ai-algorithms/
[16] Id.
[17] Rebecca Heilweil, Artificial Intelligence will help determine if you get your next job, Vox Recode (Dec. 2019), https://www.vox.com/recode/2019/12/12/20993665/artificial-intelligence-ai-job-screen

Now Institute says, “It’s a profoundly disturbing development that we have proprietary technology that claims to differentiate between a productive worker and a worker who isn’t fit, based on their facial movements, their tone of voice, their mannerisms.”[18] She goes on to say that systems like this simply gives companies a “license to discriminate.”[19]

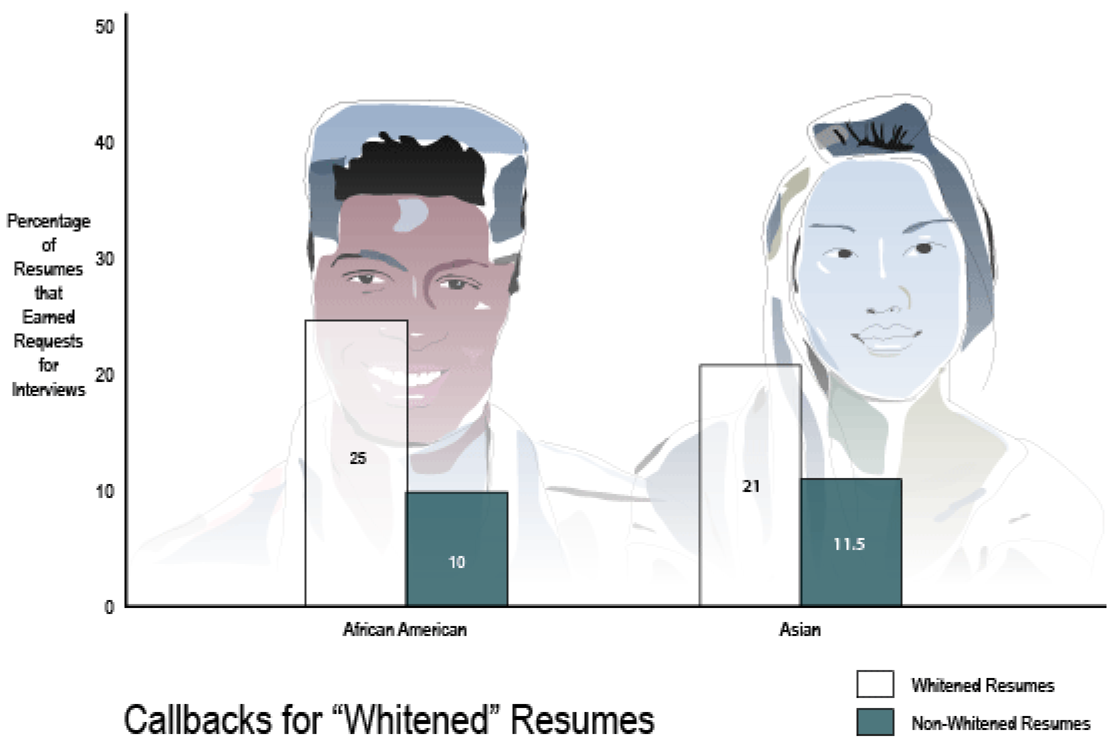
Video interviewing analyzed by algorithms has led to increased anxiety for some seeking employment. One student described feelings of confusion by the recorded interviewing process. When she did not receive a job offer, she wondered what the “AI hiring system believed she had gotten wrong.”[20] But, unlike in-person interviews, she could not inquire about her interview performance. Though HireVue abandoned its video interviewing facial analysis tool in 2020, the issue still remains.[21] There is a huge need for policy that holds corporations responsible for bias algorithms and data sets. It is more important than ever that companies are required to understand what the algorithms are doing—and how that may be impacting who gets a seat at the table. Those that are rejected based on algorithms deserve an explanation on how they were evaluated and what metrics were used. Otherwise, we will continue to have a hiring and recruitment process that plays upon stereotypes and discrimination of certain groups without any accountability or transparency.

---

[18] Harwell, Supra note 12
[19] Id.
[20] Id.
[21] Katie Bishop, Your Next Job Interview May be with a Robot—whether you realize it or not, Observer (March 2021), https://observer.com/2021/03/artificial-intelligence-job-interview-problems-bias-tips/

Otis Hill – How to Combat Racial Bias in AI Hiring

I thought you were white. Those were the words I heard from one of my former managers at a large tech firm. Although I had been working for months at that point we had never met in person because he worked in a different city. My resume and my voice on the phone led him to the conclusion that I must be white. I began to wonder if this played a role in me being hired. A study done at the University of Toronto found that minority candidates that removed all racially identifying information from their resumes received more than twice as many callbacks for interviews.[1] Crucially, this was true even at firms that espoused strong diversity initiatives.



Graphic by Blair Storie-Johnson (Source: “Whitened Resumes: Race and Self-Presentation in the Labor Market”)

[1] Sonia Kang, et al., Whitened Resumes: Race and Self-Presentation in the Labor Market (http://www-2.rotman.utoronto.ca/facbios/file/Whitening%20MS%20R2%20Accepted.pdf)

Employers have increasingly turned to AI to alleviate pressure in the hiring process. The potential benefits appear good on the surface. A neutral party that can screen resumes without bias. As Amazon found out, this is far from reality.[2] When they employed their AI solution, it exhibited a strong preference for white males in its screening process. AI requires data sets to model the specific outcomes that the developers desire. The data that Amazon fed into their AI screening algorithm showed that their historical hiring process was biased towards white males and as a result the AI saw that as the desired outcome. The algorithm was biased because humans are.

The first step in realizing a future where AI can be used in the hiring process to promote racial equity is recognizing that humans are biased. Biased or nonrepresentative data sets lead to biased outcomes according to Harvard Business Review.[3] XXX suggests audits of both the data sets and the algorithms on a regular basis is one of the best ways to deal with this issue. When auditing algorithms and data sets, we should be asking ourselves if the outcomes are producing fair outcomes. Fair in this context means that the outcomes are representative of the demographics of society. This is an important distinction because if the outcomes are instead made to mimic society, the outcomes would continue to be biased.

Researchers from the University of Cambridge demonstrated that even society's portrayal of AI is racialized.[4] From movies to google search results, AI is largely depicted as white. Cambridge's Dr. Kanta Dihal posited that, "Given that society has, for centuries, promoted the

---

[2] Jeffrey Dastin, Amazon scraps secret AI recruiting tool that showed bias against women (https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G)
[3] Frida Polli, Using AI to Eliminate Bias from Hiring (https://hbr.org/2019/10/using-ai-to-eliminate-bias-from-hiring)
[4] Alexander McNamara, Racial inequality exacerbated by 'white AI' (https://www.sciencefocus.com/news/racial-inequality-exacerbated-by-white-ai/)

association of intelligence with white Europeans, it is to be expected that when this culture is asked to imagine an intelligent machine, it imagines a white machine."[5] This represents another barrier in using AI as a tool to promote racial equity rather than exacerbate the problem because it often means that the people directly working on AI are not diverse.

Despite what may seem lack an insurmountable task, some companies have found success in using AI to increase their number of diverse hires. Several websites dedicated to the use of AI to remove personal identifiers such as race, gender, and national origin have had a positive effect. Use of skill-based testing AI resulted in a jump from 20% of diverse candidates getting an initial interview based on resume screening alone to 60%.[6] Outcomes that are more fair are being realized, but there is no standard, and all of these methods are still in the experimentation phase. AI isn't where it needs to be yet, but it can still be very useful in the hiring process with proper auditing and a focus on fair outcomes and improved data sets. Mitigation of the downsides of using AI while reaping the benefits should be the goal, but that takes continued learning and a desire to do better from the industry, lawmakers, and society.

[5] Id.

[6] Kimberly A. Houser, Can AI Solve the Diversity Problem in the Tech Industry? Mitigating Noise and Bias in Employment Decision-Making (https://law.stanford.edu/wp-content/uploads/2019/08/Houser_20190830_test.pdf)

Lindsay Holcomb

AI & Bias: Credentialism in the Hiring Process

Imagine you are a senior at a small, rural college in the town where you grew up. The college is well-known regionally but lacks broader name recognition. You are a straight-A student, majoring in Business. Your resume highlights your internships at a local accounting firm, credit union, and municipal government office where you gained superb, hands-on experience in the financial sector and took on significant responsibility. By all accounts, you have the skills required to be a valuable employee in any financial organization. Now, with graduation approaching, you are interested in a competitive entry-level position at a financial institution that receives thousands of applicants every year. Recently, in order to manage applications, the institution has contracted with an AI-based resume review vendor in order to weed out weaker applicants before stronger ones are passed along to recruiters in human resources. You prepare your application and submit it, confident that you meet the requirements listed on the institution's hiring page. But your resume never makes it to the next level of review. Not only have you been denied an opportunity for professional advancement, but the financial organization has been denied the opportunity to retain a potentially outstanding employee. What happened?

In recent years, AI-based hiring platforms have been touted as an antidote to the well-documented[1] biases of human decision-makers in the resume evaluation and interview processes.[2] "When done well, using algorithms in hiring has the potential to reduce bias that may

[1] *See* Anthony Greenwald & Linda Hamilton Krieger, *Implicit Bias; Scientific Foundations* 94 CAL. L. REV. 945, 961 (2006) ("Evidence that implicit attitudes produce discriminatory behavior is already substantial").
[2] *See, e.g.*, Kimberly Houser, *Can AI Solve the Diversity Problem? Mitigating Noise and Bias in Employment Decision-Making* 22 STAN. TECH. L. REV. 290, 293 (2019) ("The responsible use of artificial intelligence in employment decision-making not only increases the diversity of candidates and employees, but actually result in more successful employment outcomes"); Charles A. Sullivan, *Employing AI* 63 VILL. L. REV. 395, 399 (2018) ("There are reasons to expect AI to do better than humans by reducing irrationality in employee selection, and that

come from subjective human decisions, such as quick resume scans and interviews based on intuition," writes Jenny Yang, former commissioner of the U.S. Equal Opportunity Employment Commission.[3] Kimberly Houser has echoed this sentiment, writing, "The responsible use of artificial intelligence can mitigate unconscious bias by reducing the impact of human decision-makers on the process, and create better employment decisions which are based on skills, traits and behaviors rather than factors (such as sex, race, or pedigree)."[4] Despite the significant excitement surrounding the potential for AI-based hiring algorithms to effectively predict which job applicants will perform best on the job,[5] fit in best with their coworkers,[6] and stay at the organization the longest,[7] these algorithms have not lived up to their promise and instead have replicated many of the biases they were designed to avoid.[8] Of the three factors mentioned by Houser, sex and race-based biases have received ample scholarly attention,[9] but very little

---

includes performing better than we humans have done in avoiding discrimination"); Emily Peiffer, *Algorithms Risk Perpetuating Bias in Hiring. How Can Employers Use Them to Make Hiring More Inclusive?* URBAN INST. (Nov. 8, 2018) https://www.urban.org/urban-wire/algorithms-risk-perpetuating-bias-hiring-how-can-employers-use-them-make-hiring-more-inclusive ("With the rise in online job boards over the last decade, companies that used to get 20 resumes for one position now get 2000. A lot of companies don't have the capacity to individually look at every application that comes in. Employers need a more sophisticated way of screening because there are so many people to choose from").

[3] Peiffer *supra* note 2.

[4] Houser *supra* note 2.

[5] Rebecca Heilweil, *Artificial Intelligence Will Help Determine If You Get Your Next Job* VOX (Dec. 12, 2019) https://www.vox.com/recode/2019/12/12/20993665/artificial-intelligence-ai-job-screen.

[6] Nicolas Rivero, *How to Use AI Hiring Tools to Reduce Bias in Recruiting* QUARTZ (Oct. 9, 2020) https://qz.com/1914585/how-to-use-ai-hiring-tools-to-reduce-bias-in-recruiting/.

[7] Karen Hao, *An AI Hiring Firm Says It Can Predict Job Hopping Based on Your Interviews* MIT TECH. REV. (Jul. 24, 2020) https://www.technologyreview.com/2020/07/24/1005602/ai-hiring-promises-bias-free-job-hopping-prediction/.

[8] *See* Miranda Bogen, *All the Ways Hiring Algorithms Can Introduce Bias* HARV. BUS. REV. (May 6, 2019) https://hbr.org/2019/05/all-the-ways-hiring-algorithms-can-introduce-bias (Analyzing predictive tools across the hiring process and finding that "most hiring algorithms will drift toward bias by default"); *see also* Manish Raghavan & Solon Barocas *Challenges for Mitigating Bias in Algorithmic Hiring* BROOKINGS INST. (Dec. 6, 2019) https://www.brookings.edu/research/challenges-for-mitigating-bias-in-algorithmic-hiring/ ("On their surface, algorithmic screening tools seem to be entirely evidence-based, making it an appealing alternative to biased human evaluations. However, there is mounting evidence that such tools can reproduce and even exacerbate human biases manifested in the datasets on which these tools are built").

[9] *See* Rachel Kraus, *Amazon Used AI to Promote Diversity. Too Bad It's Plagued with Gender Bias* MASHABLE (Oct. 10, 2018) https://mashable.com/article/amazon-sexist-recruiting-algorithm-gender-bias-ai/ (explaining that Amazon's proprietary AI-based hiring program was found to down-rank female candidates); *see also* Betsy Williams et al., *How Algorithms Discriminate Based on Data They Lack: Challenges Solutions, and Policy*

inquiry has been made into one of the most pernicious, yet difficult to detect, means of hiring bias: academic pedigree.

Credential bias has the opportunity to sneak in at many junctures in the algorithmic resume analysis, particularly in the context of hiring entry-level or internship positions. To evaluate a batch of candidates, an algorithm will rank the candidates or assign them scores based on the presence of specific keywords in their resumes.[10] Unlike a senior hire, though, whose resume's work experience might offer professionally-rooted keywords like "investment analyst" or "manager" or "increased revenue," an entry level candidate or an internship applicant likely has little content outside of credentials on her resume. The algorithmic rules dictating which credential-specific keywords deserve which scores in a resume review are not necessarily written by a human. Rather, they are developed by the program's machine learning system based on the past data that are fed to the algorithm by data scientists and engineers.[11] The machine learning system might be given the resumes of current employees from when they were entry-level applicants and that data might be paired with their on the on-the-job performance, or the length of time they stayed at the company, or any other metric hiring managers might find valuable in a candidate.[12] Taken together, the computer can then identify keywords that successful employees have tended to use in their resumes and develop a rule based on those keywords.[13] In that sense, all of the biases of real world, offline office experience are transmuted into the algorithmic hiring process even though the rules of hiring are developed automatically through machine learning.

Implications 8 J. OF INFO. POL. 78, 92 (2018) (confirming "statistical discrimination in hiring based on online applications to entry-level jobs coded with names that are characteristic of particular races").

[10] Raghavan & Barocos *supra* note 8.

[11] Manish Raghavan et al., *Mitigating Bias in Algorithmic Hiring: Evaluating Claims and Practices* 2 (Dec. 6, 2019) https://arxiv.org/pdf/1906.09208.pdf.

[12] *Id*.

[13] Raghavan & Barocos *supra* note 8.

Returning to the process of algorithmic resume review described above in the context of a large financial institution, when fed resumes from past entry-level candidates, the algorithm may notice some particular keywords and draw conclusions about those keywords based on their associated on-the-job performance. The names of particular colleges and universities might be assigned more weight than others; the names of internship programs might be rated more or less high based on prestige; a bachelor's in science might be weighted more heavily than a bachelor's in arts; and the names of particular majors might be assigned higher scores than others. Because hiring managers in the past have been biased towards candidates from Princeton or Harvard, who have interned at J.P. Morgan or Morgan Stanley, and who have bachelor's degrees in finance or economics, the algorithm reviewing candidates like the hypothetical one described above may learn to also have a preference towards those traits.[14] As Cathy O'Neil has written, "When we blithely train algorithms on historical data, to a large extent we are setting ourselves up to merely repeat the past...The data is, after all, simply a reflection of our imperfect culture."[15]

Aside from overt name recognition in keywords, the hiring algorithm may also develop rules based on more subtle cues about a candidate's academic pedigree. It may learn to associate the term "concentration" with elite academic programs given that only a small minority of Ivy League schools refer to majors as such; it may learn to negatively rate degrees in "business" or "accounting" given that very few high-ranking schools have such a degree; and it may downgrade resumes with professional experience that appears to be in a different class or category than financial services.[16] For example, while some hiring managers might be thrilled to

---

[14] *See* KAREN HO, LIQUIDATED: AN ETHNOGRAPHY OF WALL STREET 52 (2009) (explaining that Wall Street has an "intense and persistent focus on Princeton, Harvard, and a few other campuses" and "elite universities are naturally conflated with investment banking").

[15] Eric Rosenbaum, *Silicon Valley is Stumped: AI Cannot Always Remove Bias from Hiring* CNBC (May 30, 2018) https://www.cnbc.com/2018/05/30/silicon-valley-is-stumped-even-a-i-cannot-remove-bias-from-hiring.html.

[16] *See* Ho *supra* note 14 at 76-77 (explaining that banks tend to associate lesser schools with "back-office" positions and therefore recruit them less heavily).

see that a candidate worked as a waitress at a restaurant – thinking that such experience lends customer service skills, an ability to quickly relay and retain information, and a strong work ethic – an algorithm might not view the position as favorably because it will have no way of drawing such inferences unless the phrase has appeared enough times on past resumes. For an applicant like the hypothetical individual described above, who attends a college without widespread name recognition, who majors in Business, whose work experience does not include the names of fancy firms, but rather terms like "credit" and "accounting," a hiring process guided by such biased algorithmic rules does not look promising. Rather than recognizing the person's skillset or potential for success on the job, an algorithmic review of the person's resume would likely yield a low score simply because the data points of her resume do not conform to those of an idealized applicant.

This type of ingrained credential bias is harmful in two ways. First, it denies opportunities to qualified applicants who would be the first in their school's history, or in their family, or in their town, to obtain a position at an elite firm. Second, it deprives organizations of talented applicants who have the requisite skillset and commitment to be a successful employee.[17] These harms are all the more pernicious in that they arise from the efforts of an ostensibly neutral, evidence-based tool. As outlined above, algorithmic hiring tools are intended to remediate human biases, protecting applicants from whatever unfair prejudices or opinions a particular recruiter or hiring manager may have.[18] But as Manish Raghavan writes, "Algorithms, by their nature, do not question the human decisions underlying a dataset. Instead, they faithfully attempt to reproduce past decisions, which can lead them to reflect the very sorts of biases they

[17] McKenzie Raub, *Bots, Bias and Big Data: Artificial Intelligence, Algorithmic Bias and Disparate Impact Liability in Hiring Practices* 71 ARK. L. REV. 529, 532 (2018).
[18] Gregory Bekken, *The Algorithmic Governance of Data Driven Processing Employment* 7 PSYCHOSOCIOLOGICAL ISSUES IN HUMAN RESOURCE MANAGEMENT 25, 26-30 (2019).

are intended to replace."[19] When trying to evaluate a candidate's value to the company, algorithmic hiring technologies can only represent the candidate within the same framework and on the same terms as previous candidates.[20] Thus, though vendors of algorithmic hiring tools claim to represent candidates fairly and objectively, in fact, their tools emphasize many of the human biases that have for so long impeded fair candidate evaluations.

This is a complex issue that demands immediate solutions. As more and more companies use AI-based hiring programs to conduct a first screen of their applicants, it is essential that the vendors of such programs are held accountable for their work. As New York Times technology columnist Meredith Broussard has written, "The fundamental problem [is] algorithms are designed by people, and people embed their unconscious biases in algorithms. It's rarely intentional, but it doesn't mean we should let data scientists off the hook."[21] To that end, vendors must be more transparent with respect to the construction, validation, and use of their algorithmic screening tools as well as the data sets used to build them. This is not to say that vendors should be required to turn over proprietary algorithms and sensitive employee information, but rather that they should disclose the mechanisms by which they aim to achieve unbiased assessment.[22] The "black box" approach to hiring is an unacceptable means of mitigating human bias as it merely adds more mystery to the process and gives applicants no recourse to action if are curious why they were not called back for a second round interview or if they are wondering what they

---

[19] Raghavan & Barocos *supra* note 8.

[20] Tom Sühr et al., *Does Fair Ranking Improve Minority Outcomes? Understanding the Interplay of Human and Algorithmic Biases in Online Hiring* ASSOC. FOR COMPUTING MACHINERY (Jul. 17, 2017) https://arxiv.org/pdf/2012.00423.pdf.

[21] MEREDITH BROUSSARD, ARTIFICIAL UNINTELLIGENCE: HOW COMPUTERS MISUNDERSTAND THE WORLD 150 (2018).

[22] Candice Schumann et al., *We Need Fairness and Explainability in Algorithmic Hiring* BLUE SKY IDEAS TRACK (May 9, 20202) http://www.ifaamas.org/Proceedings/aamas2020/pdfs/p1716.pdf.

can do to improve their hiring prospects in the future.[23] Where so little is known about the internal operations of these applicant screening tools, applicants may find themselves at a loss in terms of how best to represent themselves to hiring managers and recruiters.[24] In revealing the methods employed to achieve objective, neutral applicant review, vendors can increase the public's confidence in their products and hold themselves and each other accountable to their stated aims. Ultimately, transparency is crucial as sunlight may be the best disinfectant of credential-based algorithmic bias.

[23] John Horton, *The Effects of Algorithmic Labor Market Recommendations: Evidence from a Field Experiment* N.Y.U. STERN SCHOOL OF BUSINESS 25-28 (Mar. 4, 2016) http://john-joseph-horton.com/papers/algo_labor_rec.pdf.
[24] Min Kyung Lee, *Understanding Perception of Algorithmic Decisions: Fairness, Trust, and Emotion in Response to Algorithmic Management* 7 BIG DATA & SOC. 1, 3-5 (2018).

<u>Wooseok Ki</u>

<u>Factors to Consider in AI Hiring Bias – Social Media Presence</u>

In the digital age of heightened social media use in which information is inevitably mediated by technology, both recruiters and job candidates utilize social media platforms for effective employment outcomes. The abundance of platforms allows the user to freely express one's own achievements through various content, whether it be a LinkedIn profile boasting job experiences or an Instagram page detailing personal narratives. In the recruitment context, I pose the challenges AI-based hiring algorithms driven by social media presence must face; would they advocate unbiased employment practices or merely aggravate existing biases towards minority applicants? I explore the current social media landscape across different platforms with a focus on LinkedIn, consider policy implications for companies' recruiting efforts particularly the screening of applicants' social media presence, and outline what should be considered for the future in eliminating bias for an AI-based algorithm approach.

Even as early as 2013, companies were increasingly searching and screening applicants' social media profiles to gather information; albeit this was an era when social media was not fully understood by researchers and human resources professionals alike.[1] According to surveys from the Society for Human Resource Management, companies reporting social media use for recruiting practices increased over 40% from 2008-2013, and 65% in 2015 reported using social media.[2] In today's landscape it's not a stretch to assume social media screening in recruiting decisions has escalated exponentially. Companies may inadvertently display overt discriminatory

---

[1] Philip L. Roth et al., *Social Media in Employee-Selection-Related Decisions: A Research Agenda for Uncharted Territory*, JOURNAL OF MANAGEMENT (2013).
[2] Enrica N. Ruggs et al., *Online Exclusion: Biases That May Arise When Using Social Media in Talent Acquisition*, *in* SOCIAL MEDIA IN EMPLOYEE SELECTION AND RECRUITMENT 289 (Richard N. Landers and Gordon B. Schmidt, eds., 2016).

practices in the process.[3] Most applicants already lack the negotiating power to be at an even playing field as the companies; it is even more difficult for applicants to detect hiring bias without sufficient company disclosure to determine if it has actually occurred.[4] In the midst of this, minority applicants based on gender, race, sexual orientation, etc. are disproportionately affected as social media further perpetuates the power dynamic and biases already present in the hiring process.[5]

There is currently a lack of research on the relationship between employers' social media screening of applicants and underlying biases and stereotypes.[6] Yet, it is widely accepted knowledge amongst applicants to any institution to be wary of how they would be presented pre-selection, even in the college admissions context.[7] Personally if I am applying to a job position, I would also be cautious of how I present myself on *personal* social media accounts (ex: Instagram, Facebook) versus *professional* platforms (ex: Linkedin). Yet I question whether social media presence should be a factor in hiring outcomes at all, because my social media decisions imply that there are biases outside of my control based on my public presentation. A 2011 Belgian study indicated that professionals tended to judge extraversion and maturity based on applicants' Facebook profile pictures.[8] Given the growth of social media in a decade and the

---

[3] Carl Samson, *Tech Company Says Job Ad for 'Non-Asians' on LinkedIn Had No 'Discriminatory Intent'*, NEXTSHARK (Feb. 16, 2021), https://nextshark.com/job-advertisement-tech-company-non-asians/.
[4] Marc Bendick, Jr. and Ana P. Nunes, *Developing the Research Basis for Controlling Bias in Hiring*, JOURNAL OF SOCIAL ISSUES VOL.68, NO.2, 238-262 (2012); Manish Rhagavan et al., *Mitigating Bias in Algorithmic Hiring: Evaluating Claims and Practices* (Dec. 6 2019).
[5] *Supra* note 2
[6] Payton McCarthy Stewart, Preventing Implicit Biases from Influencing Employer Decisions in Social Media Screen (Aug. 2020) (unpublished Ph.D dissertation, University of Houston).
[7] Rachel Nuwer, *To Avoid College Admissions Scrutiny, High Schoolers Are Changing Their Names on Facebook*, SMITHSONIAN MAGAZINE (Nov. 16, 2012), https://www.smithsonianmag.com/smart-news/to-avoid-college-admissions-scrutiny-high-schoolers-are-changing-their-names-on-facebook-128508950/.
[8] Ralf Caers and Vanessa Castelyns, *LinkedIn and Facebook in Belgium The Influences and Biases of Social Network Sites in Recruitment and Selection Procedures*, SOCIAL SCIENCE COMPUTER REVIEW 29(4) 437-448 (Nov. 2011).

ease of accessing personal information online, common selection biases pre-selection are likely to be even greater.

In the hiring context of mitigating AI bias, I believe that the algorithm must be structured to assess inherent biases in underrepresented minority groups prior to focusing on social media presence. Despite a push to incorporate AI to supplement human decision-making in the hiring context and the optimism for consistency, I think that ultimately social media presence is largely a visual stimulus and thus will initially be prone to error.[9] For example, AI could definitely be trained to detect certain keywords in LinkedIn profiles that suit the job description; an AI algorithm for a software engineer (SWE) position should detect keywords such as “back-end”, “programming”, etc. Yet SWE is very apparently a male-dominated position – if there is inherently more of a certain demographic in a field, what would be the best way to ensure underrepresented talent are equally screened despite thousands of candidate profiles? In contrast, what about for law firm associate positions which generally lack pre-screening competency tests or personal portfolios unlike those in tech? Should AI algorithms attract candidates and evaluate applicants based sheerly on what is displayed on the LinkedIn profiles, when plenty of lawyers don’t update them?

Hence when designing and applying AI algorithms to detect certain applicants based on their social media presence, companies should take extra caution to train and inform employees to be cognizant of the already inherent biases in social media screening. The demographics differ per occupation, and anonymous pre-interview assessments may have greater weight. Employers must be cognizant that personal and professional social media use have become indispensable

---

[9] Kimberly A. Houser, Can AI Solve the Diversity Problem in the Tech Industry? Mitigating Noise and Bias in Employment Decision-Making, 22 STAN. TECH.L.REV. 290, 333 (2019) (discussing possibility of errors and societal prejudices in social media data from Internet).

today. Even though there should be a socially accepted threshold that determines a candidate's character and fitness based on the social media presence, the question remains how much AI algorithms should probe into non-professional platforms. This paper focused on LinkedIn specifically because candidates' Facebook, Instagram and TikTok presence invite such a different set of concerns surrounding privacy and biases not directly relevant to credentials. Therefore when designing an AI algorithm to reduce hiring bias based on social media presence, companies must take into consideration major factors such as prevalent demographics of the job position, the diversity of LinkedIn profiles and content, and the inherent shortcomings in hiring methods that should carefully focus on underrepresented groups for equal opportunity.

Ashley Marcus

Gender Bias in AI

The underrepresentation of women in STEM is more than just a pipeline issue—it is a product of gender bias that is now reflected and reinforced in AI programs. According to The Elephant in the Valley survey, women working in Silicon Valley/Bay Area technology companies experience inhospitable work environments that negatively impact the industry's ability to successfully recruit and retain talented women.[1] To better their AI programs, technology companies must start by improving their culture, increasing their gender diversity, and creating more inclusive training data.

Technology companies must carefully develop gender inclusivity trainings and promote communication about gender bias and inappropriate conduct to improve the culture of the industry. The longstanding tradition of a homogenous male-dominated workspace has bred an inhospitable and oppressive environment wherein women face pressure to walk a tightrope—they must be masculine enough to be seen as competent but feminine enough to be likeable.[2] Men within the technology industry must recognize women as equal members of a team rather than as maternal or daughter-like figures conforming to gender expectations for their acceptance. Open discussions about existing biases and their damaging effects brings a rather taboo topic to the forefront. Employees must be trained to recognize behaviors that contribute to a harmful environment.[3] Once they identify ways in which they contribute to the problem, they can take steps to actively change their behavior and improve the culture of the technology industry.

---

[1] ELEPHANT IN THE VALLEY, https://www.elephantinthevalley.com/ (last visted Mar. 24, 2021).
[2] *See* Joan C. Williams, *The 5 Biases Pushing Women Out of STEM*, HARVARD BUSINESS REVIEW (Mar. 24, 2015), https://hbr.org/2015/03/the-5-biases-pushing-women-out-of-stem (revealing that "[m]ore than a third (34.1%) of scientists surveyed reported feeling pressure to play a traditionally feminine role . . . .).
[3] *See* ELEPHANT IN THE VALLEY, *supra* note 1 (noting "90% [of women surveyed] witnessed sexist behavior at company offsites and/or industry conferences[,]" 87% received "demeaning comments from male colleagues[,]" and "60% of women in tech reported unwanted sexual advances.").

Individuals charged with designing these programs, however, must be thoughtful in their planning. The program ought to be informative while avoiding alienating men and leaving them feeling defensive rather than motivated to help make a positive change.[4]

Recruiting departments must remove applicant names to eliminate gender bias in assessing competence when first evaluating candidates. In 2012, a randomized, double-blind study found that both male and female faculty members in charge of hiring perceived candidates with identical application materials as more hirable and competent when those applications were attached to a man's name rather than a woman's.[5] Removing names for an *initial* assessment may allow recruiters to evaluate a candidate unencumbered by gender biases; however, names and gender may later be revealed so that the recruiters become aware of the gender-makeup of their incoming class of new hires. In this way, recruiting departments remain accountable for the gender-diversity in their hiring. Relying on a name-blind practice alone would create a lack of accountability and responsibility that leads to discrimination and unequal access to jobs within the technology industry.[6] The two-step process would strike the optimal result of allowing women's applications to be assessed without gender bias while maintaining accountability within recruiting departments.

AI programming teams must increase the number of women represented in training data. The underrepresentation of women at technology companies mirrors the underrepresentation of images of women in training data, which translates into AI programs' substantial limitation: it

[4] According to McKenzie Raub's article, *Bots, Bias and Big Data: Artificial Intelligence, Algorithmic Bias and Disparate Impact Liability in Hiring Practices*, "[P]oorly implanted diversity programs and messaging in work environments can signal to white male candidates and employees that 'they might be undervalued and discriminated against.'"). McKenzie Raub, *Bots, Bias and Big Data: Artificial Intelligence, Algorithmic Bias and Disparate Impact Liability in Hiring Practices*, 71 Arkansas Law Review 529 (2018).

[5] *Supra*, note 2 citing Moss-Racusin et al., *Science Faculty's Subtle Gender Biases Favor Male Students*, 109 Proceedings of the National Academy of Sciences of the United States of America 16474 (2012).

[6] *See supra*, note 4.

works much better for light-skinned men than it does for the rest of the population. For example, when it comes to AI programs identifying faces, IBM's program has the largest gap in accuracy with a 34.4% error rate between light-skinned men and dark-skinned women.[7] Using a greater number of images of women with a variety of skin tones would improve AI programs' ability to identify women correctly.

Improving AI goes beyond simply increasing the number of images in training data. The limitations on AI programs are rooted in the systemic gender biases that exist in the technology industry. Only in increasing gender diversity within the industry will these programs reach their optimal potential. Creating a gender-inclusive culture and workforce increases the number and types of perspectives working on these programs. Greater gender diversity on AI programming teams ensures that women have the opportunity to make sure they see themselves represented in the data and the programs prove effective in identifying not just the light-skinned man but also themselves.

---

[7] Joy Buolamwini, *Project Gender Shades*, MIT MEDIA LAB (last visited Mar. 24, 2021, 11:10 PM), https://www.media.mit.edu/projects/gender-shades/overview/.

Facial Recognition: A Recognition of its Current Biases

Levi I. Marelus

Facial Recognition

Facial recognition allows for the matching of a person's facial image on a photo or a video frame against a database repository of facial images. By measuring facial features, facial recognition technology allows for the identification, verification, and analysis of people in a streamlined manner.

Facial recognition has been around for some time. In the 1990s already, facial recognition technology was used to identify and prevent people from obtaining multiple driving licences[1], which are often used for identification purposes. The technology has come a long way since, and is likely to remain an essential and important tool in various areas of life, ranging from criminal justice, to consumer advertising, to healthcare diagnoses and preventions. The National Human Genome Institute Research Institute, as an example, using facial recognition, has detected and successfully diagnosed DiGorge syndrome, a rare disease and one difficult for healthcare providers to identify, in 96% of cases.[2] The focus of this article is facial recognition's expansion into the hiring process. The article will thus first discuss facial recognition's relevance in hiring, its uses and its benefits. Next, it will discuss the risks of using facial recognition in hiring and its possible detrimental effects on equality and fairness, and its exacerbation of technological and human biases. Finally, the article will identify and discuss possible mitigatory practices and mechanisms to ensure facial recognition is used responsibly in a way beneficial for all.

Facial Recognition in Hiring

Facial recognition is being used more and more extensively in hiring and recruiting processes. Saving time and costs, recruiters have the ability to use facial recognition to analyse photo

---

[1] Gates, Kelly (2011). Our Biometric Future: Facial Recognition Technology and the Culture of Surveillance. NYU Press. p. 53.

[2] Kruszka, P. *22q11.2 deletion syndrome in diverse populations*, American Journal of Medical Genetics *Part A*, 2017

images or video frames of an applicant's face and thereby evaluate the applicant's facial expressions for the purposes of gaining a better understanding of the applicant's emotions, personality and character traits, such as the applicant's integrity, passion, and agreeableness.

Facial recognition technology can also assist recruiters in finding applicants that would best fit their specific company's cultural ecosystem, by matching the applicant's facial image to the company's database of 'ideal' employees and past "successful" hires.

In the right circumstance and with the right input, facial recognition in hiring has the strong and important ability to "remove […] the 'damaging' human bias" that may cloud one's judgment.[3] As Yi Xu, CEO of Human claims: "We are likely to make recruitment decisions based on chemistry, mood or context rather than on skills, suitability for the role or level of emotional intelligence." Using facial recognition processes in recruitment "provide[s] a level of intelligence that was previously unattainable" removing potential human recruiting biases.

Nonetheless, study after study have demonstrated that facial recognition in hiring have – at least so far - exacerbated and perpetuated biases and discrimination in the recruitment process.

Bias and Discrimination

Joy Buolamwini has demonstrated in her Gender Shades project that the facial recognition software developed by Microsoft, IBM and Face ++ misidentified 0.8% of lighter skinned males, compared to 34.7% of darker skinned women.[4] Confirming these findings, the National Institute of Standards and Technology (NIST), has found that facial recognition technologies across 189 algorithms demonstrated the lowest accuracy rate with women of colour[5]. Women were thus more likely to be falsely identified than men, and the study found that it was middle-aged white men who largely profited from the highest accuracy rates. As a shocking example of racial bias in facial recognition technology, in 2015, Google's photo-tagging tool appallingly misidentified African-Americans as gorillas.[6]

[3] Yi Xu, CEO Human, as reported by Gentle, S. *Onrec*, 2018

[4] Buolamwini, J., *Gender shades: Intersectional accuracy disparities in commercial gender classification*, Conference on Fairness, Accountability and Transparency; See also, Simonite, T. *Photo Algorithms ID White Men Fine—Black Women, Not So Much,* WIRED, 2018

[5] Grother, P. *et al*, *Face Recognition Vendor Test (FRVT) Part 3: Demographic Effects,* 2019

[6] Zhang, M., *Google Photos Tags Two African-Americans As Gorillas Through Facial Recognition Software*, Forbes, 2015

In addition to the racial and sexual biases that have been found in facial recognition hiring software, disability discrimination is often a prevalent, and an often forgotten, element. 26% of Americans suffer from some form of disability,[7] and tens of thousands have been diagnosed with a craniofacial condition.[8]

A disabled person's facial, or even vocal expression, may thus be misinterpreted as negative, and when matched against the database repository of "ideal" employees, will be considered as different to the ideal model, thereby being denied the opportunity to serve in a certain capacity, even if they were perfectly suited for that particular position.[9]

Wendy Chisholm movingly stressed her concern for facial recognition's susceptibility for disability bias: "How do we represent ourselves to technology? How does someone with a disability establish herself to job matching and career matching algorithms. 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?"

Mitigatory Approaches:

AI and facial recognition have numerous benefits, and have the ability to change our world for the better. Whilst facial recognition has been found to be wanting when it comes to eliminating biases, even goes so far as perpetuating them, facial recognition can play a strong and effective part in their elimination with the right approaches.

Training Data: Sarah Drinkwater, Director of Responsible Technology at Omidyar Network, shared her thoughts on the use of the right training data as follows: "Yes, humans often have prejudices that lead to discriminatory decisions. 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."

---

[7] Centre for Disease Control and Prevention, *Disability Impacts All of Us*, 2018
[8] National Institute of Dental and Craniofacial Research, *Prevalence of Cleft lip and Cleft Palate*, 2018
[9] Engler, A., *For Some Employment Algorithms, Disability by Default*, Techtank, 2019

A core element in the perpetuation of biases in facial recognition, and indeed in other AI processes, is the data bank repository that has been inputted with biased data, or simply with an incomplete array of diverse data. With the right "training", Google's facial recognition would, for example, not have misidentified African-Americans, nor would there be any discrepancies in the identification accuracy rates between darker skin females and lighter skin males (see the Gender Shades project). Similarly, when it comes to disabilities for example, AI training data should include many more people with diverse disabilities, allowing facial recognition technologies to better "match" an applicant's facial features and expressions with a more accurate and diverse database.

AI Developers: The biases found in AI technologies are affected by the diversity, or lack thereof, of the people building them.[10] As an example, recently, Google shared that 54.4 percent of its workforce identified as white and 68.4 percent identified as male.[11] With a more diverse workplace and diverse talent, companies would be able to much better combat the various forms of bias found in AI.

Transparency and Constant Auditing: The machine learning and facial recognition technology systems are often clouded in mystery to the casual observer. This lack of transparency has dual implications: first, it means employers may not necessarily have a clear-cut understanding of the data inputs used, and the AI's underlying reasoning for its determinations to either hire or not to hire. Second, the flipside: the applicants do not always necessarily have a clear-cut understanding of the decisions either. This lack of transparency can allow for bias to creep up and continue unabated. More transparency to both employers and employees, and constant auditing would allow companies to get a better understanding of the systems, the decisions they affect, and the possible ways to mitigate any and all biases that may be found.

[10] Michael L, *To Build Less-Biased AI, Hire a More-Diverse Team,* 2020; See also, *AI is in Danger of Becoming too Male - New Research*, The Conversation, 2019
[11] Dickey, M. R., *The Future of Diversity and Inclusion in Tech*, Techcrunch, 2019

Alex Mueller

Political Beliefs

Although Title VII of the Civil Rights Act of 1964 prohibits discrimination based on attributes classes like race, religion, sex, or national origin, this protection typically does not extend to an individual's political beliefs or affiliations.[1] In fact, private employers in most states face very little restrictions on their ability to discriminate against job candidates based on political views. It may be tempting to dismiss concerns about political bias in hiring, especially since claims about purported widespread political discrimination and the importance of "ideological diversity" are commonly brought up to deflect or trivialize calls for greater gender or race-based diversity.[2] Yet, political discrimination still presents a number of ethical concerns that have implications for designing fair and rights-preserving AI-based hiring systems.

Issues with Political Discrimination in Hiring

Despite the absence of a federal law prohibiting employers from soliciting information about a job applicant's political views, this is something that almost never occurs. Such information is still perceived as highly sensitive and something that most people are hesitant to disclose, often out of fear that it could be used for discriminatory purposes. While regulated almost exclusively by social convention in the U.S., the sensitivity of this type of information and the need for restrictions on its use in automated decision making can be found in the laws of some foreign jurisdictions. The EU's GDPR, for example, explicitly identifies personal data

---

[1] 42 U.S.C. §§ 2000e-2(a)(1).

[2] *See, e.g.*, Kate Conger, *Exclusive: Here's The Full 10-Page Anti-Diversity Screed Circulating Internally at Google*, GIZMODO (Aug. 6, 2017), https://gizmodo.com/exclusive-heres-the-full-10-page-anti-diversity-screed-1797564320; *see also* Camila Domonoske, *James Damore Sues Google, Alleging Discrimination Against Conservative White Men*, NPR (Jan. 9, 2018), https://www.npr.org/sections/thetwo-way/2018/01/09/576682765/james-damore-sues-google-alleging-discrimination-against-conservative-white-men.

revealing one's political opinions as part of a category receiving special protection and can only be used in automated decision-making under a few specific conditions.[3] When a hiring firm obtains data capable of revealing a data subject's political opinions from a third-party source or by depriving the subject of a meaningful choice over whether to disclose, it negatively impacts individual privacy rights.[4] Further, incorporating this information into hiring decisions can lead to a chilling effect that harms important societal values like free expression and association due to concerns that certain public political affiliations or views will impact one's employment chances.[5] Finally, political discrimination in hiring can lead to legal issues where political ideology or affiliation effectively serves as a proxy for protected characteristics like race or religion.[6]

<u>Is all political-based discrimination bad?</u>

Much of the fairness concerns surrounding bias against protected attributes (e.g., race or gender) in hiring come from its capacity to disadvantage or deny equal opportunities to candidates on the basis of factors that have little effect on future job performance. There is nothing inherent about these factors, most of which are immutable and/or out of one's control, that make an individual of one variety better suited for a given job than an individual of any other variety. Yet, this is not necessarily true with respect to all political beliefs, something demonstrated by the existence of certain ideologies motivated by hate or prejudice. An individual espousing political views that are openly hostile or promote violence towards certain racial

---

[3] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016, 2016 O.J. (L 119) art. 9, 22.
[4] *See* Filippo Raso, et al., *Artificial Intelligence and Human Rights: Opportunities & Risks*, Berkman Klein Center for Internet & Society Research Publication 42-46 (September, 25 2018).
[5] *See id*. at 46.
[6] *See* Katharine Pickle, *Why Google May Be in the Right: An Analyses of Political Discrimination in the Workplace*, 5 EMORY CORP. GOVERNANCE & ACCOUNTABILITY REV. PERSPECTIVES 179 (2018).

groups would be a poor fit in just about any modern company. While fostering diversity of viewpoints in a workplace can yield positive effects, the presence of harmful ideologies can lead to increased workplace conflict and create a toxic culture which makes it hard to attract and retain employees that are diverse with respect to other attributes.

Similarly, there may also be instances where the business activities of a private employer are fundamentally incompatible with an individual's political beliefs to the extent that would likely render them incapable of performing the job. Such instances are rare due in part to self-selection by candidates, as individuals with conflicting political views have little reason to voluntarily join the applicant pool.[7] Yet, where such irreconcilable political differences do exist, treating a candidate differently on the basis of their views, so long as it serves the legitimate business need of employing individuals that are willing and able to perform expected job duties, is not unfairly discriminatory.

Addressing Political Bias in AI-Based Hiring Systems

Since there is a legitimate argument that some degree of differential treatment towards political beliefs can be justified when making hiring decisions, the important question becomes whether AI-based systems can be trusted to limit the differential treatment to instances where it does not offend fairness. In other words, is an algorithm capable of reliably drawing the distinction between political beliefs that interfere with an individual's job performance or the workplace, and those political beliefs which do not? The likely answer is no, which is why in the

---

[7] *See* Andrew F. Johnson and Katherine J. Roberto, *Elections and selection: The role of political ideology in selection decisions*, 29 HUMAN RESOURCES MANAGEMENT REV. 21-22 (2019).

interest of ensuring fair and ethical automated hiring systems, data revealing political beliefs should not be used as a model input in order to prevent disparate treatment.

Although it would be ideal if an AI-based hiring system could filter out candidates with dangerous or incompatible political beliefs while treating all other political beliefs equally, there are simply too many challenges involved for this to be realistic. In order to train a model to do this, information about individual political beliefs would need to be included in the training dataset. However, self-reported information about one's political beliefs or affiliations is highly susceptible to social-desirability bias, as individuals who possess controversial politics are likely to censor their responses to appear more socially acceptable. This means there is a significant possibility of a model being trained on unreliable data. Further, unlike attributes such as gender or race which can be fully captured by simple nominal data types, the most common and obtainable measures of political orientation, things like party affiliation or position on a political spectrum, lack the nuance to adequately capture an individual's full political beliefs. [8] A person affiliated with a party may disagree with many of its official positions or take them to a much further extreme. This creates a risk of individuals with non-problematic beliefs being disadvantaged by a model solely because of their party affiliation.

Omitting data that represents one's political views from a model's inputs can only do so much to control bias. Where outcomes in the training data were affected by a hiring managers' conscious or unconscious political biases, which limited evidence suggests can occur, this bias

[8] There are also advanced methods that purport to accurately construct detailed profiles of a person's political beliefs using a wide variety of data inputs obtained from third parties (e.g., internet tracking, facial imaging). *See generally* Michael Kosinski, et al., *Private Traits and Attributes Are Predictable from Digital Records of Human Behavior*, 110 Proc. NAT'L ACAD. SCI. U.S. 5802 (2013). However, even if these methods were demonstrated to be reliable, there would be serious privacy and ethical concerns involved with obtaining this data without consent and using it generate sensitive inferences that serve as a basis for making important decisions about an individual.

can still manifest itself in a model through the inclusion of correlated factors.[9] Instead of discriminating directly on political affiliation, a screening algorithm might filter out an applicant because they attend a liberal university or grew up in conservative zip code. A further complication here is that methods of testing for disparate impact among applicants of different political beliefs requires knowledge of these beliefs, sensitive information that hiring employers rarely obtain from applicants and which is subject to certain types of response bias.[10] If the challenges involved with testing a model for disparate impact with respect to political beliefs make it practically infeasible for vendors of AI-based hiring systems, or alternatively, then it should consider some other means of identifying and removing factors that are reasonably likely to serve as proxies for political beliefs. Even though federal law does not compel vendors to prevent this type of disparate impact, as political views or affiliations are not protected attributes, this is still something worth aiming for if the ultimate goal is to make these AI-based systems even fairer than the human processes they are replacing.

---

[9] *See* Philip L. Roth, Jill E. Ellingson and Jason B. Thatcher, *Is political affiliation the new discrimination? Our research suggests 'yes'*, THE HILL (November 27, 2019); *see also* Pauline T. Kim, *Big data and artificial intelligence: New challenges for workplace equality*, 57 U. Louisville L. Rev. 320 (2018).

[10] *See* Manish Raghavan, et al., *Mitigating bias in algorithmic hiring: evaluating claims and practices*, arXiv preprint arXiv:1906.09208 (2019), https://arxiv.org/pdf/1906.09208.pdf.

Dignity in the Age of Data: How Interdisciplinary Leadership and Human Rights Break Open the Black Box of AI Bias

Maryam Nasir

The increasing sophistication of Artificial Intelligence (AI) technology and its rapid proliferation into modern society makes it clear that the technology is here to stay. These powerful prediction engines silently recognize patterns, predict choices, and shape the way we use online platforms. Harnessing the benefits of increasing personalization by AI is not without risk. The growing use of AI in the context of employment and automated hiring has placed the technology under a regulatory spotlight and growing evidence of its irresponsible deployment in several high-profile cases are raising tangible concerns for vulnerable populations.[1] As AI continues to augment human-decision making, this report argues that, in the absence of comprehensive, federal oversight, industry stakeholders must take it upon themselves to interrogate algorithmic decisions and formulate interdisciplinary teams to mitigate the complex risks AI poses to human rights principles.

The use of AI in hiring decisions is touted for its ability to neutralize decision making and the interpersonal human prejudices it can precipitate.[2] However, algorithmic parameters are imbued with the values of their developers and thus can introduce data blind spots that, when deployed to make hiring decisions for thousands of individuals, embed biases and produce discriminatory outcomes at scale. According to researchers, even when characteristics protected under Title VII of the Civil Rights Act of 1964 are purposefully omitted from training sets, biases can be reintroduced through proxy variables.[3] For instance, apparently benign attributes like one's zip code can serve as a proxy for race due to historical patterns of housing discrimination.[4] Even when developers are extremely careful in their implementation of training data, they can still codify discriminatory results as machine learning algorithms can discover

---

[1] Dastin, Jeffrey. "Amazon Scraps Secret AI Recruiting Tool That Showed Bias against Women." *Reuters*, Thomson Reuters, 10 Oct. 2018, www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G.

[2] Simon Chandler, *The AI Chatbot Will Hire You Now*, WIRED (Sept. 13, 2017, 6:45 AM), https:// www. wired. com/ story/ the-ai- chatbot- will- hire- you- now/ [https://perma.cc/XK5U-5PUP]

[3] Raghavan, M., Barocas, S., Kleinberg, J. and Levy, K. (2019). Mitigating Bias in Algorithmic Hiring: Evaluating Claims and Practices. ACM Conference on Fairness, Accountability, and Transparency (FAT*), 2020. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3408010.

[4] Solon Barocas and Andrew D. Selbst, "Big Data's Disparate Impact." *California Law Review* vol 104, no. 3 (2016: 671–732. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2477899.

subtle correlations for protected characteristics revealing socioeconomic and racial status.[5] In one example, even when the sex of applicants in a hiring algorithm was removed, the algorithm was still able to teach itself to look for verbs more commonly associated with male engineers, such as the word "executed." Because the algorithm had been fed data from primarily males it awarded higher scores to the resumes submitted by male applicants.[6]

At their core, AI systems replicate human behavior[7] and as Kleinberg et al. suggest, it would be naive —to conflate "algorithmic" with "objective," or to think that the use of algorithms will necessarily eliminate discrimination against protected groups.[8] Even if flawed algorithms were expunged from the public space and manicured to address disparate impact, we would still have to account for the Silicon Valley engineers and entrepreneurs who are being increasingly pressured to not only diversify their white, male-centric workforce but also account for the culture that has forced women, such as Dr. Timnit Gebru, out of the space entirely. [9]

Biased decision making, whether by humans or machines, has led many tech companies to roll out high level ethics principles such as fairness and accountability and while such developments are encouraging, the difficulty lies in operationalizing these metrics. No matter how laudable these principles are, there must be a pathway for stakeholders both affected by AI-hiring platforms and civil society at large to access the black box decision making that occurs within the AI space. As researchers from the AI Now Institute at New York University (NYU) point out, "without a framework that accounts for social and political contexts and histories, these mathematical formulas for fairness will almost inevitably miss key factors and can serve to paper over deeper problems in ways that ultimately increase harm or ignore justice."[10] Ethical commitments to fairness that are not tied down to mechanisms of accountability raise concerns for a false sense of security. Similarly, reducing transparency to the revealing of an algorithm's source code is diluted by the consideration that nonexperts are illiterate in the language of a machine's algorithms.

---

[5] Miranda Bogen and Aaron Rieke. "Help Wanted: An Examination of Hiring Algorithms, Equity, and Bias" (Washington, DC: Upturn, 2018), https://www.upturn.org/reports/2018/hiring-algorithms/.
[6] Kimberly Houser. Can AI solve the diversity problem in the tech industry? mitigating noise and bias in employment decision-making. Stanford Technology Law Review, 22, 2019.
[7] Solon Barocas and Andrew Selbst, Big Data's Disparate Impact, 104 Calif. L. Rev. 671 (2016), http://www.californialawreview.org/wp-content/uploads/2016/06/2Barocas-Selbst.pdf.
[8] Jon Kleinberg, Jens Ludwig, Sendhil Mullainathan, Cass R Sunstein, Discrimination in the Age of Algorithms, *Journal of Legal Analysis*, Volume 10, 2018, Pages 113–174,
[9] West, S.M., Whittaker, M. and Crawford, K. (2019). Discriminating Systems: Gender, Race and Power in AI. AI Now Institute. Retrieved from https://ainowinstitute.org/ discriminatingsystems.html.
[10] Whittaker, Meredith, et al. *AI now report 2018*. New York: AI Now Institute at New York University, 2018.

As evidenced by the literature review for the 2021 Penn Law Spring Policy Lab, the current debate around AI governance is rich and takes various forms including external auditing, *FAT* principles, and public statements. This report recognizes the importance of this discourse, but further contends that there is more reason to be critical than accepting of ethics initiatives. As Selbset et al. argue, most of the current regulatory approaches to fairness and accountability fail to "account for the full meaning of social concepts such as fairness, which can be procedural, contextual, and contestable, and cannot be resolved through mathematical formalisms"[11] Addressing the ecosystem of challenges that this report has laid out—at the data and designer level—necessitates a meaningful transition; one that strays from adherence to broad benchmarks of parity and, instead, upholds an interdisciplinary commitment to framing principles of non-discrimination as a responsibility to human rights shared among business leaders, designers, and developers alike.

In agreement with Christiaan Van Veen of NYU and Corinne Cath of Oxford Internet Institute, adopting a corporate responsibility initiative grounded in a human rights framework "is itself a source of power because human rights carry significant moral legitimacy" and for industry leaders, "the reputational cost of being perceived as a human rights violator can be very high." [12] This call to action is echoed by research from Deborah Raji and Joy Buolamwini that illustrates how multistakeholder pressure and risks of company integrity in the face of performance results, separate from legal developments and regulatory compliance, were able to motivate companies to prioritize addressing classification bias in their systems within less than a year. [13] Mutale Nkonde, a fellow at Harvard's Berkman Klein Center for Internet and Society at Harvard Law School, argues that corporations could shoulder social responsibility by using CSR budgets to combat the underlying social issues that lead to the introduction of bias in the first place. In anticipation of corporate cooptation in the face of what Safiya Noble calls "techlash,"[14] supplanting a human rights *by design* framework, as outlined by the UN Guiding Principles on

---

[11] Selbst, Andrew D., et al. "Fairness and abstraction in sociotechnical systems." *Proceedings of the conference on fairness, accountability, and transparency*. 2019.

[12] Christiaan van Veen and Corinne Cath, "Artificial Intelligence: What's Human Rights Got To Do With It?," Data & Society Points, May 14, 2018, https://points.datasociety.net/artificial-intelligence- whats-human-rights-got-to-do-with-it-4622ec1566d5.

[13] Raji, Inioluwa Deborah, and Joy Buolamwini. "Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial ai products." *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*. 2019.

[14] Le Bui, Matthew, and Safiya Umoja Noble. "We're Missing a Moral Framework of Justice in Artificial Intelligence." *The Oxford Handbook of Ethics of AI* (2020): 163.

Business and Human Rights[15], could provide teeth to growing social responsibility initiatives by companies.

The importance The Guiding Principles place on contemplating the relationship between rights holders and duty bearers across industry and civil society through the principle of due diligence transcends the focus on technical outcomes and instead offers a normative standard to identify, anticipate, and mitigate risks for harm at both the deployment and product design level. Recognizing the responsibility of enterprises to respect human rights, as illustrated by Guiding Principle 19, would offer redressal mechanisms and also ensure that new and emerging AI technologies are designed, deployed and regulated in a way that is enabling of, rather than threatening to, the human rights enshrined in established bodies of law.[16] In the context of the workplace, the burdens of error-prone data sets are borne by groups that have already been historically marginalized by discrimination. While a rights-based approach necessitates technical considerations, it importantly invites input from these groups as well as designers, industry leaders, and civil society to crystallize a commitment to upholding human dignity in the age of AI.

[15] United Nations Human Rights Office of the Human Rights Commissioner. 'Guiding Principles on Business and Human Rights', 2011. https://www. ohchr.org/Documents/Publications/ GuidingPrinciplesBusinessHR_EN.pdf.

[16] Fjeld, Jessica, et al. "Principled artificial intelligence." *Berkman Klein Center, February* 14 (2020).

TOOLKIT FOR MITIGATION: AI BIAS AGAINST DISABILITY

By Yinran Pan

I. BACKGROUND

Historically, society has marginalized and excluded individuals and groups such as minorities, females, and those with disabilities from resources and opportunities. As AI systems learn how to operate by studying history and society, societal biases become embedded into their DNA. Lack of data points also skews the machine learning process to recognize and prefer groups that display traits shown in the dominant data—often, able-bodied, white males.[1] As such, AI systems are inherently biased against groups that display outlier tendencies and even have the potential to amplify the biases that these groups face on a daily basis.[2] Because data on disability is tied to individuals' medical records, this data is difficult to obtain and largely unavailable for machine learning, thereby creating a "data desert" that hinders the development of solutions.[3] For individuals with disabilities, this means they are often targeted as outliers by AI systems and thus, deemed as lesser qualified and excluded from job opportunities.

For example, HireVue, an AI interviewing software used by many companies, uses characteristics such as facial expressions and speech enunciation to determine a potential candidate's suitability for the role.[4] Yet, an individual with motor skills impairment or a speech impediment who could otherwise perform all other essential responsibilities of the job would not fit into the archetype this AI system was taught to recognize and thus be parsed out, regardless of their qualifications.[5] Furthermore, disability is a blanket term for many physical and mental conditions, so biases that individuals with disabilities face can come in many shapes and forms of varying degrees.[6] If an AI system is not trained to recognize the different ways in which disabilities can manifest, it will indiscriminately screen out candidates with behaviors it has never encountered.

---

[1] Alexandra Reeve Givens, *How Algorithmic Bias Hurt People With Disabilities*, Slate (Feb. 6, 2020, 5:15PM), https://slate.com/technology/2020/02/algorithmic-bias-people-with-disabilities.html.

[2] AI Now Institute, *Disability, Bias, and AI*, MEDIUM (Nov. 21, 2019), https://medium.com/@AINowInstitute/disability-bias-and-ai-4434dc7065f0.

[3] Jennifer Langston, *Shrinking the 'data desert': Inside efforts to make AI systems more inclusive of people with disabilities*, MICROSOFT - THE AI BLOG (Oct. 12, 2020), https://blogs.microsoft.com/ai/shrinking-the-data-desert/.

[4] Alex Engler, *For some employment algorithms, disability discrimination by default*, BROOKINGS (Oct. 31, 2021), https://www.brookings.edu/blog/techtank/2019/10/31/for-some-employment-algorithms-disability-discrimination-by-default/.

[5] *Id.*

[6] Givens, *supra* note 1.

II. Current Policy Landscape

The Americans with Disabilities Act (ADA), passed in 1990, aims to prevent discrimination against individuals with disabilities in all public sectors, including employment, government, transportation, schools, and other public spaces.[7] Despite many companies purporting to be EEO employers in accordance with this law, they rely on outdated guidelines issued by the EEOC that allow bias to persist in the recruiting and hiring process.[8] Thus, the ADA is insufficient in preventing discriminatory behavior perpetuated by AI systems.

New legislation has been introduced at the federal and state levels seeking to regulate AI. In 2019, California passed Senate Joint Resolution 6, a measure urging the President and Congress "to adopt a comprehensive artificial intelligence policy" in light of increased use of AI by both government and private entities.[9] The Algorithmic Accountability Act, which aims to regulate and increase the transparency of systems that use AI or machine learning, was first introduced in the House of Representatives in 2019, and its proponents plan to reintroduce a updated version to Congress this year.[10] Similarly, many states have introduced various bills regarding the regulation of AI technologies.[11] Although many such bills did not pass, acts that have passed include Illinois' H.B. 2557, requiring employers to notify and acquire the consent of potential job applicants prior to AI-assessed video interviews, and New York's S.B. 3971, which created a temporary state commission to research ways to regulate AI systems.[12] Alabama, California, New Jersey, Utah, Vermont, and Washington have also created similar task forces or commissions regarding the research and regulation of AI in the past few years.[13] Recently, New York City introduced Int. 1894, an amendment that attempts to regulate systems that use algorithms to evaluate candidates for jobs by requiring such systems be audited.[14] With regards to Int. 1894, Manish Raghavan, a researcher studying the impacts of AI in hiring, points out that while this legislation is a good first step at mitigating bias, it is limited in scope because of the

---

[7] U.S. Department of Justice, *A Guide to Disability Rights Laws*, (Feb. 2020), https://www.ada.gov/cguide.htm.
[8] Givens, *supra* note 1 (noting how the EEOC's 4/5 rule is outdated and too simplistic for our diverse society).
[9] S.J. Res. 6 (Cal. 2019).
[10] *Id*.
[11] *Legislation Related to Artificial Intelligence*, NCLS (Jan. 17, 2021), https://www.ncsl.org/research/telecommunications-and-information-technology/2020-legislation-related-to-artificial-intelligence.aspx.
[12] *Id*.
[13] *Id.*
[14] Manish Raghavan, *Testimony of Manish Raghavan before New York City Council Committee on Technology regarding Int. 1894* (Nov. 9, 2020), https://www.cs.cornell.edu/~manish/int_1894_testimony.pdf.

limitations of current laws and because audits cannot detect information on disabilities which candidates choose not to disclose.[15] In the private sector, companies such as Microsoft are researching ways to make AI systems more inclusive of individuals with disabilities by creating more inclusive and representative training datasets for AI systems' machine learning but are still encountering "data deserts".[16] All of these efforts are paving the way to make AI systems less biased and more inclusive but more needs to be done, especially to protect those with disabilities.

III. Policy Options to Address AI Bias Against Disability in the Hiring Process

A. Policy 1: Requiring Data Sets Used to Train AI Systems to be More Inclusive

1. Establish a **quota** of data points representing certain groups that must be used to train AI.
2. Form a **regulatory body** to enforce standards for AI technologies and oversee the collection and use of training data.

Data sets currently used in machine learning training are largely skewed towards able-bodied individuals. By requiring AI systems or machine learning technologies to meet a certain standard regarding the diversity and inclusivity of their training data, the resulting assessment of these systems may be less biased. For example, there could be a required quota of data points representing certain groups for the set of data used to train AI systems. In conjunction with this policy, the government should set up a regulatory body that will not only enforce this quota and oversee the collection and collaborative-sharing of such "rare" data but will also assess AI technologies in general. With AI and machine learning on the rise, it would be prudent for the government to form a regulatory agency that can monitor the various aspects of life and diverse communities that AI affects and that will combat biases that taint the AI processes and results.

B. Policy 2: Providing Channels to Access Evaluations and Report Perceived Bias

1. Allow job candidates **access to their AI-generated evaluation reports** and provide **clear channels to report perceived** bias in such reports.
2. Give candidates the **choice to opt-out** of AI evaluations, no questions asked.

Though many companies use AI systems to assess potential candidates, they often do not provide information on *how* the AI system evaluated the candidates. Without knowing what their shortcomings were, these applicants may be left wondering if they have been discriminated against. Requiring employers and AI hiring systems to retain the evaluation reports and make them available to job applicants will allow candidates—and potentially other stakeholders, such

---

[15] *Id*.
[16] Langston, *supra* note 2.

as AI researchers and developers—to assess whether biases have occurred. Furthermore, employers should provide clear channels for candidates to report biases they believe have occurred during the AI recruiting and hiring process. Finally, individuals should be given the option to request an evaluation not performed by AI, with no questions asked, since not all individuals with disabilities may want to disclose their condition.

C. Policy 3: Increasing Transparency of AI Hiring Process for the Public

1. Require *all* AI hiring technologies to be subject to **bias audits**, using **standardized metrics**, and make these reports **accessible to the public.**
2. Require the resulting audit reports to **acknowledge the shortcomings** of the audits.

AI technologies, especially those used in hiring, are still fairly new and thus largely unregulated. Their evaluation processes require no formal auditing and there are few, if any, mechanisms in place to hold these systems accountable for their harmful and discriminatory practices. This policy, like NYC's Int. 1894, aims to require bias audits of all AI programs used in hiring, and the resulting audit reports should be made available to the public. Moreover, such audits must have standardized metrics for what auditors should be accounting for—including how the AI system differs in its capability to assess different groups of people—and must explain which forms of discrimination, such as those against certain types of disabilities or certain groups of people, cannot be detected or accounted for in the result of these audits. This acknowledgement ensures that employers and the public are not misled into thinking that the AI systems used in hiring are bias-free and fair. As such, programmers of these systems will be accountable to the public for creating AI algorithms that are more equitable and inclusive. These audit reports will also help highlight shortcomings that can be improved upon.

IV. Conclusion

While each policy option would act to prevent or mitigate biases against individuals with disabilities during the hiring process, legislative action can only go so far. A cultural shift and the way in which society perceives and frames disability is paramount to the success of these policies. Conversations on reducing biases against individuals with disabilities must include members of the disabled community and center around their lived experiences. Leaders, in both the private and public sector, should take advantage of their large platforms to help advocate for the disabled community. Moreover, anyone can be an ally and help bring about greater awareness of the difficulties individuals in the disabled community face. As society becomes more diverse, inclusive, and equitable, so too will AI systems be modeled after these behaviors.

Religion and Bias in AI
By: Sara Shayanian

Candidates applying for jobs on hiring and recruiting platforms may encounter several forms of bias in their quest to find a career. In many instances, employers are "using algorithms to decide who gets interviewed, hired, or promoted".[1] One of these prejudices can come in the form of religious bias. However, there are ways for AI to help identify and mitigate bias on these platforms through a technological changes.

Unfortunately, employers may harbor personal prejudices against certain faiths. In one extensive study, researchers found that candidates who listed membership in faith-based student organizations received fewer responses from U.S. employers than those with no mention of religion.[2] This prejudice was particularly stronger in southern states in the U.S. than in New England states. According to Michael Wallace, a co-author of the study, employers may both be biased towards certain religions while others may fear that people who decide to reveal their religious beliefs on their resumes may one day clash with co-workers.[3] Muslims, pagans, and atheists applicants bear the brunt of the bias in the United States.[4]

Outside of the United States, religious prejudice exists globally in the hiring process. Studies in France and Greece have found hiring bias for certain religious groups.[5] For example, a Muslim applicant with African heritage was two-and-a-half times less likely to get a foot in the door for an interview in France than an equally qualified Christian applicant with the exact same ethnic background.

---

[1] Pauline T. Kim, *Data-Driven Discrimination at Work*, 58 William & Mary L. Rev. 857 (2017).
[2] Ronald Alsop, *Does religious bias begin with your CV?*, BBC (July 30, 2014), https://www.bbc.com/worklife/article/20140730-reveal-religion-on-your-cv.
[3] *Id.*
[4] *Id.*
[5] *Id.*

Another example of this discrimination came from a study of Greece hiring practices.[6] Compared to Greek Orthodox applicants, job seekers who identified as Pentecostal, evangelical, or Jehovah's Witnesses were given less job interviews and received lower wages.[7]

Closer to home, a peer who does not wish to be named, applied on LinkedIn for a writing position she was overqualified for. Her first name and last name are traditional to her Middle Eastern culture and Islam. Because of her background her name on LinkedIn was not what one would consider a typically 'white name' or 'Caucasian name'. She was also a member of a Muslim student organization on her undergraduate campus. Ultimately, she was not hired or given an interview. Later, she noted how a white candidate and friend of hers with far less credentials was hired for the position. To this day, this candidate feels that it was her name and participation in the organization that cost her the job. This anecdote is just one example of job candidates beginning to "rationally fear that their participation in potentially observable activities", such as religious groups, will "trigger discrimination".[9] Members of a "religious group will often try and "avoid posting their group affiliation on social media or forego… partaking in another action associated with the group out of fear of repercussion".[10]

Fortunately, there are some ways for recruiting platforms to use innovative technologies to hire fairly and with as little unconscious or conscious bias as possible. First, recruiting and hiring platforms have already taken strides in changing their algorithms to help mitigate bias in the initial evaluation process.

---

[6] *Id.*
[7] *Id.*
[8] Anya E.R. Prince & Daniel Schwarz, *Proxy Discrimination in the Age of Artificial Intelligence and Big Data,* 105 Iowa L. Rev. 1293 (2020).
[9] *Id.*
[10] *Id.*

Many platforms have already made moves to counter unconscious and conscious bias by using algorithms to remove national origin, which can often lead employers to deduce religion.[11] One example of Unbias.io, "removes faces and names from LinkedIn profiles".[12] Interviewing.io eliminates unconscious bias in recruiting by providing an anonymous interviewing platform.[13]

However, despite technology being changed to mitigate bias in recruiting, recruiters should beware of the problems AI can have. In many cases, AI tools may still have problems that lead to discrimination in the hiring process. To help combat against this, recruiting and hiring platforms may take extra steps to remove names of religious organizations automatically from resumes to provide a neutral slate for whatever employer is reviewing resumes. Using technology to create black box recruiting on hiring platforms in this way could result in the hiring of people of diverse religious backgrounds. Furthermore, employers could try to control for issues with AI technology by controlling "for such issues by running algorithms and comparing such results to human decision makers' results, to help account for any differing outcomes between the two processes".[14]

By utilizing technology and continuing to ensure AI is tested, employers and employees of all religions will have a level-playing field to apply for and get hired for the jobs they deserve.

---

[11] Kimberly A. Houser, *Can AI Solve the Diversity Problem in the Tech Industry? Mitigating Noise and Bias in Employment Decision-Making*, 22 Stan. Tech. L. Rev. 290, 325 (2019).

[12] *Id.*

[13] *Id.*

[14] Nakimuli Davis-Primer & Aldo Leiva, *Artificial Intelligence and Bias: Considerations to Prevent Bias and Mitigate Legal Risk of Employers*, JD Supra (Nov. 12, 2020), https://www.jdsupra.com/legalnews/artificial-intelligence-and-bias-97199/.

Anna Sheu

Potential in Limitations: How AI Can Incorporate and Eliminate Unconscious Gender Bias

As a human invention, artificial intelligence (AI) reflects human unconscious bias. However, AI can also help eliminate bias. AI is merely a tool that can be used positively and negatively.

The limitations of AI

Our class has studied the downsides of AI. Job posting algorithms may target candidates based on incorrect gender stereotypes (such as showing nursing job postings to women and doctor postings to men).[1] In addition, companies may use candidate-filtering algorithms to winnow down their applicant pool by eliminating resumes that do not contain key phrases such as Ivy League brands.[2] Even more insidiously, AI can be susceptible to confirmation bias, as Amazon's job automation system showed.[3] AI relies on 'training' datasets to 'learn,' and a dataset unintentionally consisting of mostly male resumes, especially in the male-dominated tech industry, will be more likely to produce a male-favoring algorithm. In Amazon's case, graduates from women's colleges were downgraded, as were candidates participating in "women's" clubs. Harvard Business Review recently pointed out that under current U.S. law, hiring assessments can be biased if they are "job-related" (meaning that successful candidates had certain characteristics) – but if all previous successes are white men, the assessments are intrinsically be biased against women.[4]

---

[1] Miranda Bogen, *All the Ways Hiring Algorithms Can Introduce Bias*, HARVARD BUSINESS REVIEW (May 6, 2019), https://hbr.org/2019/05/all-the-ways-hiring-algorithms-can-introduce-bias
[2] *Id.*
[3] Jeffrey Dastin, *Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women*, REUTERS (Oct. 10, 2018), https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G
[4] Frida Polli, *Using AI to Eliminate Bias from Hiring*, HARVARD BUSINESS REVIEW (Oct. 29, 2019), https://hbr.org/2019/10/using-ai-to-eliminate-bias-from-hiring

The potential of AI

But AI's susceptibility can also be its strength. The World Economic Forum has pointed out that AI can reveal where unconscious biases lie in human processes.[5] Take the previously mentioned use of AI to indiscriminately and quickly filter out candidates that do not have certain keywords on their resume. On the positive side, an algorithm can process many more candidate resumes than a human could, which means that companies can expand their outreach and gather more applicants, ideally creating a larger pool of diverse candidates.

AI algorithms could be trained to combat similarity bias, which often uses the Trojan horse phrase "lack of culture fit" to sneak past equal-opportunity regulations. Again, AI algorithms can be used to categorically exclude non-Ivy League candidates. However, humans are also susceptible to this bias – a human reviewer might dismiss a resume coming from a women's college, regardless of the quality of past experiences. AI algorithms, correctly trained and audited, could remove this bias. Similarly, interviewers often make biased small talk: a male interviewer might ask a female candidate about her children or marital status; the same interviewer might bond with a male candidate over a shared college fraternity. I emphasize again the importance of training and auditing. Algorithms are a tool and can be used to both remove and enforce the glass ceiling for women.

Programmatic suggestions

We must foremost remember that AI is a human creation and, as such, will never be perfectly unbiased. I was glad to hear that class guest A. Cooper Feder agreed with me. Therefore, we should implement transparent disclosures about the training datasets and criteria

---

[5] Ann Cairns, *Why AI Is Failing the Next Generation of Women*, WORLD ECONOMIC FORUM (Jan. 18, 2019), https://www.weforum.org/agenda/2019/01/ai-artificial-intelligence-failing-next-generation-women-bias/

used in algorithms and an audit process to monitor and adjust the AI as needed (Cooper similarly stressed the importance of audit and accountability). In the Amazon example, a human auditor might have suggested that the algorithm look for leadership qualities through gender-neutral school activities ("sports" instead of "football").

The World Economic Forum has proposed a novel hybrid AI system: rather than using an AI-driven recruiting process that is (occasionally, if at all) audited by humans, why not use AI to audit the human-driven process?[6] In this way, AI can be a tool for inclusivity, as suggested by Project Include, which we covered in class.[7]

Our class has extensively discussed Isabel Wilkerson, whose seminal work on caste observes that it is endemic to our infrastructure.[8] Why should the data produced by our society be any different? It is this data that guides our AI; there is an adage about building predictive models: "garbage in, garbage out." AI cannot close the gender gap if it has been trained on data already biased against women. The World Economic Forum calls this lack of good training data a chicken-and-egg problem. Women are underrepresented in industry, leading to less data that is also worse quality. This data then feeds into AI recruiting systems, deepening the discrepancy.[9]

Because of this innate weakness, I suggest that instead of using AI to replace the human process, we use AI to audit our human "algorithms."

---

[6] Nicolás Rivero, *How to Use AI Hiring Tools to Reduce Bias in Recruiting*, WORLD ECONOMIC FORUM (Oct. 13, 2020), https://www.weforum.org/agenda/2020/10/ai-hiring-tools-bias-recruiting-hiring-diversity-fairness-equality/. *See also* Frida Polli, *supra*, who suggests using AI to audit the entire pipeline for fairness, as well as agreeing with my earlier suggestion that AI can be used to process larger pools of candidates, enabling companies to reach out and solicit more applicants from underrepresented backgrounds.

[7] Kate Dwyer, *How Some Silicon Valley Women Are Trying to Solve Sexism*, TIME (Mar. 16, 2017), https://time.com/4704219/silicon-valley-sexism-project-include/

[8] Dwight Garner, *Isabel Wilkerson's 'Caste' Is an 'Instant American Classic' About Our Abiding Sin*, NEW YORK TIMES (Jul. 31, 2020), https://www.nytimes.com/2020/07/31/books/review-caste-isabel-wilkerson-origins-of-our-discontents.html

[9] Cairns, *supra*

A case study

In a positive example of AI used in recruitment, a robotic mannequin in Sweden interviewed first-round candidates, asked the same set of questions to all interviewees, and converted their answers into a typed transcript reviewed by (human) hiring managers. The AI ignored gender as well as age, looks, ethnicity, tone of voice, and dialect.[10]

This Swedish model is not perfect. Certain characteristics may have slipped in and introduced unconscious bias. For example, language characteristics beyond dialect and tone of voice are a source of unconscious bias. Black people are known to 'code-switch.' And the failed Amazon model referenced earlier showed that men used powerful verbs such as 'executed' more often than women. However, this Swedish mannequin is a beacon for future recruiting programs.

Our class read The Elephant in the Valley, detailing the questions women received during side conversations at interviews or performance reviews, including intrusive and irrelevant questions about marriage, family planning, and religion.[11] Interviewers may think small talk lightens the atmosphere. But Claude Steele, whom we studied in class, posits that being aware of negative stereotypes towards our social group diminishes our ability to perform.[12] Steele says we can eliminate this handicap by changing the context, and that is exactly the potential of AI in breaking the glass ceiling: to change the context from a male interviewer to a genderless mannequin. This Swedish AI shifted the focus from a candidate's bias-invoking characteristics to the substance of the candidate's interview responses.

---

[10] Henry Bewicke, *This Robot Interviewer Is Helping Sweden Recruit Without Bias*, WORLD ECONOMIC FORUM (Jul. 18, 2019), https://www.weforum.org/agenda/2019/07/sweden-robot-remove-bias-from-recruitment/
[11] *The Impact of Family*, ELEPHANT IN THE VALLEY, https://www.elephantinthevalley.com/stories/impact-of-family
[12] Claude Steele, *Whistling Vivaldi, How Stereotypes Affect Us and What We Can Do*, https://www.youtube.com/watch?v=-YbE3ljT3t4&ab_channel=FamilyActionNetwork

Anna Sheu

Potential in Limitations: How AI Can Incorporate and Eliminate Unconscious Gender Bias

As a human invention, artificial intelligence (AI) reflects human unconscious bias. However, AI can also help eliminate bias. AI is merely a tool that can be used positively and negatively.

The limitations of AI

Our class has studied the downsides of AI. Job posting algorithms may target candidates based on incorrect gender stereotypes (such as showing nursing job postings to women and doctor postings to men).[1] In addition, companies may use candidate-filtering algorithms to winnow down their applicant pool by eliminating resumes that do not contain key phrases such as Ivy League brands.[2] Even more insidiously, AI can be susceptible to confirmation bias, as Amazon's job automation system showed.[3] AI relies on 'training' datasets to 'learn,' and a dataset unintentionally consisting of mostly male resumes, especially in the male-dominated tech industry, will be more likely to produce a male-favoring algorithm. In Amazon's case, graduates from women's colleges were downgraded, as were candidates participating in "women's" clubs. Harvard Business Review recently pointed out that under current U.S. law, hiring assessments can be biased if they are "job-related" (meaning that successful candidates had certain characteristics) – but if all previous successes are white men, the assessments are intrinsically be biased against women.[4]

---

[1] Miranda Bogen, *All the Ways Hiring Algorithms Can Introduce Bias*, HARVARD BUSINESS REVIEW (May 6, 2019), https://hbr.org/2019/05/all-the-ways-hiring-algorithms-can-introduce-bias
[2] *Id.*
[3] Jeffrey Dastin, *Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women*, REUTERS (Oct. 10, 2018), https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G
[4] Frida Polli, *Using AI to Eliminate Bias from Hiring*, HARVARD BUSINESS REVIEW (Oct. 29, 2019), https://hbr.org/2019/10/using-ai-to-eliminate-bias-from-hiring

The potential of AI

But AI's susceptibility can also be its strength. The World Economic Forum has pointed out that AI can reveal where unconscious biases lie in human processes.[5] Take the previously mentioned use of AI to indiscriminately and quickly filter out candidates that do not have certain keywords on their resume. On the positive side, an algorithm can process many more candidate resumes than a human could, which means that companies can expand their outreach and gather more applicants, ideally creating a larger pool of diverse candidates.

AI algorithms could be trained to combat similarity bias, which often uses the Trojan horse phrase "lack of culture fit" to sneak past equal-opportunity regulations. Again, AI algorithms can be used to categorically exclude non-Ivy League candidates. However, humans are also susceptible to this bias – a human reviewer might dismiss a resume coming from a women's college, regardless of the quality of past experiences. AI algorithms, correctly trained and audited, could remove this bias. Similarly, interviewers often make biased small talk: a male interviewer might ask a female candidate about her children or marital status; the same interviewer might bond with a male candidate over a shared college fraternity. I emphasize again the importance of training and auditing. Algorithms are a tool and can be used to both remove and enforce the glass ceiling for women.

Programmatic suggestions

We must foremost remember that AI is a human creation and, as such, will never be perfectly unbiased. I was glad to hear that class guest A. Cooper Feder agreed with me. Therefore, we should implement transparent disclosures about the training datasets and criteria

---

[5] Ann Cairns, *Why AI Is Failing the Next Generation of Women*, WORLD ECONOMIC FORUM (Jan. 18, 2019), https://www.weforum.org/agenda/2019/01/ai-artificial-intelligence-failing-next-generation-women-bias/

used in algorithms and an audit process to monitor and adjust the AI as needed (Cooper similarly stressed the importance of audit and accountability). In the Amazon example, a human auditor might have suggested that the algorithm look for leadership qualities through gender-neutral school activities ("sports" instead of "football").

The World Economic Forum has proposed a novel hybrid AI system: rather than using an AI-driven recruiting process that is (occasionally, if at all) audited by humans, why not use AI to audit the human-driven process?[6] In this way, AI can be a tool for inclusivity, as suggested by Project Include, which we covered in class.[7]

Our class has extensively discussed Isabel Wilkerson, whose seminal work on caste observes that it is endemic to our infrastructure.[8] Why should the data produced by our society be any different? It is this data that guides our AI; there is an adage about building predictive models: "garbage in, garbage out." AI cannot close the gender gap if it has been trained on data already biased against women. The World Economic Forum calls this lack of good training data a chicken-and-egg problem. Women are underrepresented in industry, leading to less data that is also worse quality. This data then feeds into AI recruiting systems, deepening the discrepancy.[9]

Because of this innate weakness, I suggest that instead of using AI to replace the human process, we use AI to audit our human "algorithms."

---

[6] Nicolás Rivero, *How to Use AI Hiring Tools to Reduce Bias in Recruiting*, WORLD ECONOMIC FORUM (Oct. 13, 2020), https://www.weforum.org/agenda/2020/10/ai-hiring-tools-bias-recruiting-hiring-diversity-fairness-equality/. *See also* Frida Polli, *supra*, who suggests using AI to audit the entire pipeline for fairness, as well as agreeing with my earlier suggestion that AI can be used to process larger pools of candidates, enabling companies to reach out and solicit more applicants from underrepresented backgrounds.

[7] Kate Dwyer, *How Some Silicon Valley Women Are Trying to Solve Sexism*, TIME (Mar. 16, 2017), https://time.com/4704219/silicon-valley-sexism-project-include/

[8] Dwight Garner, *Isabel Wilkerson's 'Caste' Is an 'Instant American Classic' About Our Abiding Sin*, NEW YORK TIMES (Jul. 31, 2020), https://www.nytimes.com/2020/07/31/books/review-caste-isabel-wilkerson-origins-of-our-discontents.html

[9] Cairns, *supra*

A case study

In a positive example of AI used in recruitment, a robotic mannequin in Sweden interviewed first-round candidates, asked the same set of questions to all interviewees, and converted their answers into a typed transcript reviewed by (human) hiring managers. The AI ignored gender as well as age, looks, ethnicity, tone of voice, and dialect.[10]

This Swedish model is not perfect. Certain characteristics may have slipped in and introduced unconscious bias. For example, language characteristics beyond dialect and tone of voice are a source of unconscious bias. Black people are known to 'code-switch.' And the failed Amazon model referenced earlier showed that men used powerful verbs such as 'executed' more often than women. However, this Swedish mannequin is a beacon for future recruiting programs.

Our class read The Elephant in the Valley, detailing the questions women received during side conversations at interviews or performance reviews, including intrusive and irrelevant questions about marriage, family planning, and religion.[11] Interviewers may think small talk lightens the atmosphere. But Claude Steele, whom we studied in class, posits that being aware of negative stereotypes towards our social group diminishes our ability to perform.[12] Steele says we can eliminate this handicap by changing the context, and that is exactly the potential of AI in breaking the glass ceiling: to change the context from a male interviewer to a genderless mannequin. This Swedish AI shifted the focus from a candidate's bias-invoking characteristics to the substance of the candidate's interview responses.

---

[10] Henry Bewicke, *This Robot Interviewer Is Helping Sweden Recruit Without Bias*, WORLD ECONOMIC FORUM (Jul. 18, 2019), https://www.weforum.org/agenda/2019/07/sweden-robot-remove-bias-from-recruitment/

[11] *The Impact of Family*, ELEPHANT IN THE VALLEY, https://www.elephantinthevalley.com/stories/impact-of-family

[12] Claude Steele, *Whistling Vivaldi, How Stereotypes Affect Us and What We Can Do*, https://www.youtube.com/watch?v=-YbE3ljT3t4&ab_channel=FamilyActionNetwork

Jessica Shieh

Policy Lab: AI and Implicit Bias

Toolkit for Mitigation: Location/Zip Code

The ever-increasing use of artificial intelligence (AI) across a panorama of industries and for myriad purposes presents as many pitfalls as it does benefits. Indeed, bias in AI has been a particularly challenging issue that has recently risen to the forefront of the technology world's consciousness. Research is beginning to show the drawbacks of algorithmic decision-making – some algorithms "run the risk of replicating and even amplifying human biases, particularly those affecting protected groups."[1]

A variety of contributing causes results in biased AI solutions. Unrepresentative underlying data samples are a major root cause, as is bias in an algorithm itself.[2] Reliance on training data that is incomplete or on information reflecting historical inequities can also result in outcomes that have a collective and disparate impact on specific groups, if left unchecked – this can happen even when the programmer has no intent to discriminate.[3] With that said, human bias is certainly a major cause as well, as AI researchers are a relatively non-diverse group in terms of gender, racial demographics, socioeconomic background, and disability status.[4]

Bias in AI has not spared the labor market, which has widely deployed AI technology via online recruiting platforms – algorithms are commonly used to screen candidates' cover letters

---

[1] Nicol Turner Lee, Paul Resnick, and Genie Barton, *Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms*, The Brookings Institution (May 22, 2019), https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/.
[2] Craig S. Smith, *Dealing With Bias in Artificial Intelligence*, New York Times (Nov. 19, 2019), https://www.nytimes.com/2019/11/19/technology/artificial-intelligence-bias.html.
[3] Jake Silberg and James Manyika, *Tackling bias in artificial intelligence (and in humans)*, McKinsey (June 6, 2019), https://www.mckinsey.com/featured-insights/artificial-intelligence/tackling-bias-in-artificial-intelligence-and-in-humans#.
[4] Smith, *supra* note 2.

and CVs for information such as years of experience, languages spoken, degrees obtained, and countries of employment, and this information is used to automatically reject applicants who do not meet a preferred set of criteria.[5] The likelihood of a candidate being hired for a job they apply for may be systematically biased, not only due to human bias, but because ostensibly neutral algorithms utilized by employers can deploy human errors at scale.[6] For instance, Amazon discontinued use of a recruiting algorithm after it was discovered that the data which engineers utilized in developing the algorithm was derived from resumés submitted to Amazon over a period of 10 years, largely by white males.[7] The algorithm "was taught to recognize word patterns in the resumes, rather than relevant skill sets, and these data were benchmarked against the company's predominantly male engineering department to determine an applicant's fit."[8] This led to the software penalizing words like "women" and the names of women's colleges on candidates' resumés.[9] While human managers certainly hold biases that lead to unfair outcomes for protected classes, the potential consequences of a biased algorithm that could disadvantage thousands of applicants dwarfs the influence of an individual biased recruiting manager.

One particular type of bias in AI-driven recruiting is geographic data and zip code-based bias. For example, questions on an application about commute time can be construed as discriminating against applicants from minority neighborhoods that may not have stable transportation infrastructure.[10] Studies have also offered distinct evidence that employment

---

[5] Julius Schulte, *AI-assisted recruitment is biased. Here's how to make it more fair*, World Economic Forum (May 9, 2019), https://www.weforum.org/agenda/2019/05/ai-assisted-recruitment-is-biased-heres-how-to-beat-it/.
[6] *Id.*
[7] Lee, Resnick, and Barton, *supra* note 1.
[8] *Id.*
[9] Schulte, *supra* note 5.
[10] Debra Cassens Weiss, *Do Job Personality Tests Discriminate? EEOC Probes Lawyer's Complaint, Filed on Behalf of His Son*, ABA Journal (Sept. 30, 2014), http://www.abajournal.com/news/article/do_job_personality_tests_discriminate_eeoc_probes_lawyers_complaint_filed_o.

discrimination generally favors white applicants, with this discrimination commonly predicted by the racial composition of the employer's neighborhood."[11] Moreover, zip code is often correlated strongly with race due to historical housing segregation and inequality in the U.S., so AI hiring processes that take into account zip code may be perniciously making decisions based on race, with zip code as a proxy.[12] If an algorithm determines that the possibility that particular races live in a certain zip code are either high or low, it could produce biased and exclusionary outcomes. Software "blind" to "demographic data like gender and race can still encode this information through other features that are statistically correlated with protected attributes."[13]

As a more granular dimension of this issue, it is important to note that location-based bias also contributes to the challenges that unhoused individuals often face in finding employment. Unhoused job applicants who are able to use the addresses of relatives or acquaintances have a material advantage over those who can only use a shelter residence, or have no address at all.[14] However, many web-based job application platforms do not allow candidates to move forward without providing an address.[15] Recruiters have no practical reason to require applicants to disclose their address during the early stages of the online application process and should remove this prerequisite – such a requirement needlessly and unfairly shuts out unhoused applicants.[16]

---

[11] Amanda Agan and Sonja Starr, *Do Employers' Neighborhoods Predict Racial Discrimination?*, Columbia University, https://law-economic-studies.law.columbia.edu/sites/default/files/content/docs/starr_do_employers_neighborhoods_predict_racial_discrimination_nov_27.pdf.
[12] Alexandra George, *Thwarting bias in AI systems*, Carnegie Mellon University College of Engineering, https://engineering.cmu.edu/news-events/news/2018/12/11-datta-proxies.html.
[13] Anupan Datta, *3 kinds of bias in AI models – and how we can address them*, New Tech Forum (Feb. 24, 2021), https://www.infoworld.com/article/3607748/3-kinds-of-bias-in-ai-models-and-how-we-can-address-them.html.
[14] Sarah Golabek-Goldman, *Ban the Address: Combating Employment Discrimination Against the Homeless*, the Yale Law Journal (April 2017), https://www.yalelawjournal.org/note/ban-the-address-combating-employment-discrimination-against-the-homeless.
[15] Sarah Golabek-Goldman, *Op-Ed: Homeless shouldn't face job discrimination just because they lack an address*, Los Angeles Times (Oct. 10, 2016), https://www.latimes.com/opinion/op-ed/la-oe-golabek-goldman-homeless-address-job-application-20161010-snap-story.html.
[16] *Id.*

Solutions to location and zip code-oriented bias in AI track closely with solutions to bias in AI generally – aside from companies' self-regulation measures, government regulation with the involvement of industry-specific officials who are deeply familiar with the subject matter is necessary, as well as more rigorous privacy legislation and potentially a formal regulatory framework focused on the ethical use of AI.[17] On the programming level, recruiting algorithms can be set to "ignore details such as the names of schools attended and zip codes that can correlate with demographic-related information such as race and socioeconomic status,"[18] and furthermore, bias may be mitigated in part by "paying attention to the data sets by using known sources and making sure the sets are balanced and representative of all groups."[19] Education is also critical – students should be taught to think deeply about the ethical, democratic, and civic implications of technology as they grow into their roles as leaders of the future.[20] Finally, companies creating critical AI solutions need to take a more intersectional and cross-disciplinary approach in hiring teams that are appropriately representative of the population, allowing for diverse and "heterogenic perspectives overseeing the development of AI."[21] The rise of AI will inevitably continue at a rapid pace, but this technology needs to be developed and utilized in an ethical and responsible way – after all, "science has to be situated in trying to understand the social dynamics of the world, because most of the radical change happens at the social level."[22]

---

[17] Christina Pazzanese, *Great promise but potential for peril*, The Harvard Gazette (Oct. 26, 2020), https://news.harvard.edu/gazette/story/2020/10/ethical-concerns-mount-as-ai-takes-bigger-decision-making-role/.
[18] Nakimuli Davis-Primer and Aldo Leiva, *Artifical Intelligence and Bias: Considerations to Prevent Bias and Mitigate Legal Risk of Employers*, JD Supra (Nov. 12, 2020), https://www.jdsupra.com/legalnews/artificial-intelligence-and-bias-97199/.
[19] Kimberly A. Houser, *Can AI Solve the Diversity Problem in the Tech Industry? Mitigating noise and Bias in Employment Decision-Making*, Stanford Technology Law Review (2019), https://law.stanford.edu/wp-content/uploads/2019/08/Houser_20190830_test.pdf.
[20] Pazzanese, *supra* note 17.
[21] Houser, *supra* note 19.
[22] Smith, *supra* note 2.

Ashley Suarez

<u>Elephant in AI: A Toolbox for Those in the AI Ecosystem to Identify and Mitigate Bias in Recruiting and Hiring Platforms—Housing</u>

When analyzed from a human rights framework, the relationship between housing and social and professional opportunity is one that evidences inequalities not only relating to one's societal wealth but also deep-rooted, systemic inequalities stemming from years of discrimination on the basis of race. Historically, where a person lives largely determines the opportunities and resources that are made available to them. Such opportunities include the availability of quality education—largely funded through property taxes—and access to well-funded community initiatives such as daycare, mental health counselling, neighborhood cleanup taskforces, and local youth mentorship programs. As a result of the relationship between community resources and property taxes, communities that are under-funded, most commonly Black and brown communities, do not gain access to such social and professional opportunities due to low property value and other impacting factors. Notably, property value has historically been determined on the basis of location—consequently skewing the most privileged communities as those that are wealthier, and whiter. Black communities, in particular, have been affected and targeted by the racism underlying the relationship between housing and opportunity. This is evidenced blatantly through the practice of redlining, beginning in the 1930s and continuing for almost 40 years before being outlawed.[1] Redlining, a practice termed by author Richard Rothstein as a "state-sponsored system of segregation," was effectuated predominantly by the Federal Housing Administration, created in 1934.[2] The FHA propagated segregation

[1] Terry Gross, *A 'Forgotten History' Of How the U.S. Government Segregated America*, NPR (May 3, 2017), https://www.npr.org/2017/05/03/526655831/a-forgotten-history-of-how-the-u-s-government-segregated-america.
[2] *Id*. (citing Rothstein's book, THE COLOR OF LAW (2017)).

efforts in the United States at the time "by refusing to insure mortgages in and near African-American neighborhoods" while "at the same time…subsidizing builders who were mass-producing entire subdivisions for whites— with the requirement that none of the homes be sold to African-Americans."[3] The redlining practice severely impacted African Americans, as they were unable to realize the equity appreciation for their home investments as compared to white Americans.[4] Further, neighborhoods that were considered "too risky" for mortgage issuance, that is, any areas where African Americans lived, continued to attract starkly lower property values even after the redlining practice was prohibited by the Fair Housing Act in 1968.[5] As a result, and as Rothstein himself notes, the segregation of communities on the basis of race has become a systemic issue that continues to affect Black and brown communities—

> Today African-American incomes on average are about 60 percent of average white incomes. But African-American wealth is about 5 percent of white wealth. Most middle-class families in this country gain their wealth from the equity they have in their homes. So, this enormous difference between a 60 percent income ratio and a 5 percent wealth ratio is almost entirely attributable to federal housing policy implemented through the 20th century.[6]

Further, Rothstein addresses the relationship between housing and opportunity, as he states that "the segregation of our metropolitan areas today leads…to stagnant inequality, because families are much less able to be upwardly mobile when they're living in segregated neighborhoods

---

[3] *Id*.
[4] *Id*.
[5] Tracy Jan, *Redlining was Banned 50 years ago. It's Still Hurting Minorities Today*, WASH POST (Mar. 28, 2018 6:00 AM), https://www.washingtonpost.com/news/wonk/wp/2018/03/28/redlining-was-banned-50-years-ago-its-still-hurting-minorities-today/.
[6] *See supra* note 1.

where opportunity is absent."[7] It is, thus, apparent how the practice of redlining and the blatant racism underlying the relationship between housing and opportunity remains relevant today.

Because of the historic, systemically flawed, relationship between housing and opportunity, it is perhaps unsurprising that the issue is prevalent amongst modern platforms used for professional hiring and recruiting. As the novelty of advanced technology, specifically artificial intelligence ("AI"), both fascinates and concerns contemporary consumers and businesses alike, such concerns are merited, as AI proves to mimic the social and economic constructs present in the non-AI world. Unfortunately, as noted by various scholars and evidenced by recent litigation involving Facebook's Ad Services algorithm,[8] AI technology mimics and implements systemic inequalities and discriminatory patterns of behavior that plague the non-AI world. For example, the complaint filed in 2018 by the National Fair Housing Alliance against Facebook alleges that, "Facebook has created a pre-populated list of demographics, behaviors, and interests that makes it possible for housing advertisers to exclude certain home seekers [that is, racial, ethnic, and gender minorities] from ever seeing their ads."[9] This practice is especially egregious given that Facebook had previously pledged in a 2016 "Newsroom" publication that they planned to "disable the use of ethnic affinity marketing for ads that [they] identify as offering housing, employment or credit."[10] Facebook, however, failed to enact such changes, and, as a result of litigation, was forced to correct its ad tech to prevent discriminatory ad delivery on its platform.[11] Further, discriminatory practices occurring on

[7] *Id*.
[8] *Nat'l Fair Hous. All. v. Facebook, Inc.*, No. 1:18-cv-02689 (S.D.N.Y. July 30, 2018).
[9] Complaint, *Nat'l Fair Hous. All. v. Facebook, Inc.* at 1.
[10] *Id*. at 15; *see also* Erin Egan, *Improving Enforcement and Promoting Diversity: Updates to Ethnic Affinity Marketing*, FACEBOOK NEWSROOM (Nov. 11, 2016), https://about.fb.com/news/2016/11/updates-to-ethnic-affinity-marketing/.
[11] Emily Dreyfuss, *Facebook Changes Its Ad Tech to Stop Discrimination*, WIRED (Mar. 19, 2019, 9:22 PM), https://www.wired.com/story/facebook-advertising-discrimination-settlement/.

platforms such as Facebook prove significant, as Facebook and Linked In are social networking platforms that control a considerable market share of the digital hiring and recruiting market, and provide such services for employers to reach potential employees across the globe. It is, thus, apparent how gatekeeper platforms' use of flawed AI technology in this manner perpetuates systemic, economic inequalities between white Americans and minorities of color as advertisements for employment and housing would be subject to the flaws in Facebook's AI algorithm—ensuring that certain job postings reach "target" groups calculated by the AI's data pool, namely where they live, rather than what the advertiser themselves intended. A study conducted by Northeastern University discussed the harmful effects of such flaws in Facebook's AI, concluding that "real-world employment and housing ads can experience significantly skewed delivery" and finding that:

> Ads with the same targeting options can deliver to vastly different racial and gender audiences depending on the ad creative alone. In the most extreme cases, our ads for jobs in the lumber industry reach[ed] an audience that is 72% white and 90% male, our ads for cashier positions in supermarkets reach[ed] an 85% female audience, and our ads for positions in taxi companies reach[ed] a 75% Black audience, even though the targeted audience specified by us as an advertiser [was] identical for all three.[12]

Such findings clearly violate the Fair Housing Act and the Civil Rights Act of 1964, as Facebook's advertisements display a "prohibited preference, limitation, specification or discrimination" in employment recruitment.[13] Put simply, if certain job postings are only delivered to racial, ethnic, and gender minorities on the basis of AI-collected data—such as

[12] Muhammad Ali et al., *Discrimination through optimization: How Facebook's Ad Delivery Can Lead to Skewed Outcomes* (Sept. 12, 2019), https://arxiv.org/pdf/1904.02095.pdf/.
[13] Julia Angwin, *Facebook Lets Advertisers Exclude Users by Race*, PROPUBLICA (Oct. 28, 2016), https://www.propublica.org/article/facebook-lets-advertisers-exclude-users-by-race.

where they live—the inequalities present in the non-AI world in terms of community wealth concentration and employment opportunities are effectively propagated by discriminatory AI ad delivery. Perhaps more difficult than identifying this issue and its broader, historic implications is the task of creating effective solutions to prevent discriminatory practices in AI from occurring. Rob Goldman, the former Vice President of Google's Digital Advertisement department, noted in his discussion during a Policy Lab session on housing that, "…it is extremely difficult to address discrimination in AI because the algorithm's own developers often do not understand why the algorithm did what it did…there needs to be a study done that it sufficiently blind with broad control groups in order to identify why this is happening and how we can fix it." Thus, in order to achieve effective policy solutions to this issue, there needs to be substantive changes to how AI data is collected, how the algorithm is developed by creators, and what metrics are ultimately used by the algorithm to ensure its functionality. Additionally, it is my view that there needs to be an industry-wide priority to rigorously test a company's AI, using research parameters specifically aimed at addressing systemic practices leading to discrimination on the basis of race, gender, and other minority demographic information.