Stanford University

Bias in Algorithmic Hiring

Elena Sierra

08 December 2021

INTRODUCTION

In recent years, many large companies began using algorithms to determine which candidates get an interview at the screening stage of the recruitment process. They are implementing this for three main reasons: cost, efficiency, and fairness (Köchling). A human screener who analyzes resumes and applications takes a significant amount of time to advance a candidate pool; this slows down the entire hiring process and costs the company a significant amount of money. On the contrary, artificial intelligence programs "streamline" the process into a matter of seconds (Raub). Also, the human screener has individual bias, which influences how they score candidates with certain identities (Köchling). Since a computer has no ingrained bias, data-driven hiring eliminates individual bias from the screening process. While the cost reduction, efficiency increase, and elimination of individual bias appear promising, one crucial concern in algorithmic hiring remains regarding fairness. While an algorithm has no individual bias, data-driven hiring processes often perpetuate systemic discrimination.

Hiring algorithms use training and testing data to optimize a model to select candidates who have similar applications to the successful employees at a company (Raghavan). These models make predictions based on resume text, video interviews, or performance on a game or quiz (Goodman). If the company trains the model on past employment data, it will discriminate against people with identities that do not align with the majority of these successful employees. When companies implement these algorithms for positions in the workforce, these models often rank female and male applicants differently. Although all genders can succeed in any role in the workforce, many fields have significant gender discrepancies that can influence how a computer selects the "optimal" candidate. This paper will examine how common gender[1] discrepancies in the workforce affects hiring algorithms, how companies currently implement algorithmic hiring

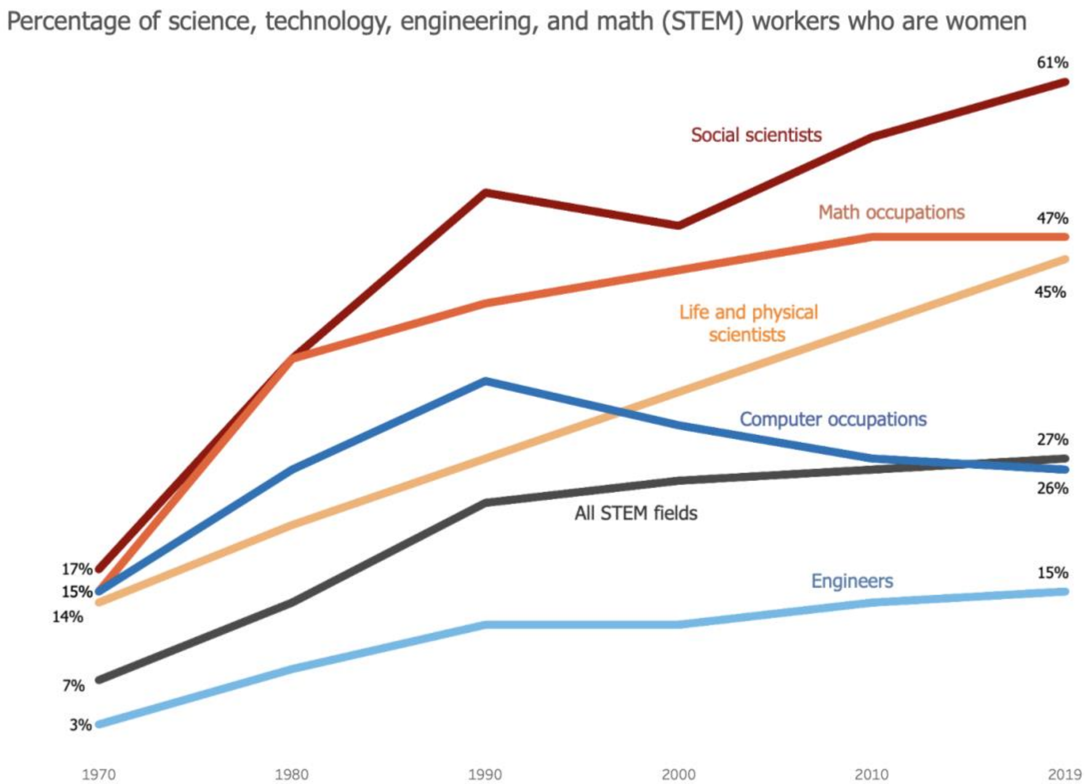
[1]While there are many gender identities beyond male and female, most data on this topic focus on binary gender identities, so this paper focuses on binary gender until the action items later in the text.

models, and how to minimize bias in these algorithms. Algorithmic hiring has the potential to eliminate bias from the recruitment process, but companies build these models in a society with systemic discrimination that becomes embedded in the code, which may be impossible to remove.

GENDER DIFFERENCES IN THE WORKFORCE

Women currently make up forty-seven percent of the United States workforce (U.S. Department of Labor). Although women and men have nearly equal representation in the current workforce, many fields still have significant discrepancies between men and women. Women dominate some occupations (teachers, nurses, and assistants) and are underrepresented in others (U.S. Department of Labor). Therefore, using a hiring algorithm based on current and past data can discriminate against the underrepresented gender in a field. For example, in high-paying and in-demand careers, such as STEM fields, women have historically and are currently underrepresented. This makes optimized models based on past data have bias against female applicants. In 1970, the percentage of women in STEM fields was only seven percent, and while this percentage has improved over the past fifty years, only twenty-seven percent of employees in STEM fields are currently women (Figure 1). Due to the underrepresentation of women in STEM, a machine learning hiring algorithm for a STEM position will have more employee data on men, so a woman applying to this position may be rejected by the model because it perceives men as better employees.

[1]While there are many gender identities beyond male and female, most data on this topic focus on binary gender identities, so this paper focuses on binary gender until the action items later in the text.



(Figure 1. "Percentage of women workers in science, technology, engineering, and math (STEM)." U.S. Department of Labor)

Within these STEM fields, there are two crucial professions involved in the developing algorithmic hiring models: computer scientists and mathematicians. Twenty percent of computer programmers and software developers in the U.S. are women, and forty-nine percent of mathematicians and statisticians are women (U.S. Department of Labor). Machine learning algorithms embed the opinions and choices of the programmers and mathematicians coding them (Raub). The lack of representation of women in the professions that develop machine learning algorithms, especially computer science, contributes to discrimination in the algorithm and the lack of transparency and attention to gender bias in the models.

COMPANY IMPLEMENTATION

How do companies implement these algorithms? Large companies, especially those with many computer programmers and software developers, attempt to create their own models for

[1]While there are many gender identities beyond male and female, most data on this topic focus on binary gender identities, so this paper focuses on binary gender until the action items later in the text.

algorithmic hiring. These tools are limited because the companies only have access to data from their own employees, and the goal of these models is to optimize and create the lowest possible margin of error, opening the door for discrimination. One example of a company that implemented its own hiring algorithm is Amazon. Amazon attempted to create its own hiring model with a historically male technical team. The computer analyzing Amazon's employee data observed that "females were promoted less frequently, tended to leave quickly, and got fewer raises," which made it conclude that men are better hires than women (O'Neil). The algorithm then went to "great lengths" to eliminate women applicants to Amazon technical roles by making a women's college on a resume a demerit, making typically male vocabulary a point in favor (O'Neil), and "downgraded resumes that included the word 'women's'" (Goodman) in its ranking process. This led to the algorithm replicating the existing pool of employees, which excluded nearly every female applicant to technical roles. Amazon is not alone in creating discriminatory hiring algorithms. Amazon just got caught. When companies formulate their own models, they prioritize optimization over all else, naively allowing the algorithms to sacrifice fairness.

Internally created hiring algorithms limit companies to their historical data. One way to work with a more comprehensive dataset is to have a third-party company create the model. A multitude of startups has emerged that focus on algorithmically screening candidates, and Cornell University conducted a study on eighteen vendors that create screening algorithms for companies. When examining these companies, the researchers focused on two main topics: validity and fairness. Validity ensures that the "outcome of a test should say something meaningful about a candidate's potential as an employee," ensuring the model performs as expected on candidates outside of the training data (Raghavan). Of the eighteen companies in the

[1]While there are many gender identities beyond male and female, most data on this topic focus on binary gender identities, so this paper focuses on binary gender until the action items later in the text.

case study, only two companies publicly conducted case studies to prove validation in their methods, and thirteen of the eighteen companies do not mention validation at all (Figure 2.). The researchers examine fairness in two ways: "disparate treatment and disparate impact" (Raghavan). Disparate treatment outlines that "it is illegal to explicitly treat candidates differently based on categories protected under Title VII" (Raghavan), which outlaws discrimination based on race, color, religion, sex, national origin, sexual orientation, and gender identity (U.S. Equal Employment Opportunity Commission). Hiring algorithms generally are not in danger of disparate treatment because they do not make explicit decisions based on the identities of candidates; instead, they make decisions based on the contents of a candidate's application. Disparate impact is much more dangerous for machine learning algorithms because it refers to cases where a seemingly "neutral process still produces substantially different outcomes for people that are correlated with legally protected attributes" (Wilson). The Uniform Guidelines 4/5th rule determines disparate impact (Raghavan), which requires that corporations accept minority candidates at least eighty percent of the rate the company accepts majority candidates. Out of the eighteen companies, only three ensure their models follow the 4/5th rule through debiasing methods (Figure 2.) The third-party companies programming hiring algorithms rarely reveal their methods for validation and fairness in the models they produce. This lack of transparency is concerning when these models determine which candidates get interview offers.

[1]While there are many gender identities beyond male and female, most data on this topic focus on binary gender identities, so this paper focuses on binary gender until the action items later in the text.

Vendor name	Assessment types [Features]	Custom? [Target & Training data]	Validation info [Validation]	Adverse impact [Fairness]
8 and Above	phone, video	S	–	bias mentioned
ActiView	VR assessment	C	validation claimed	bias mentioned
Assessment Innovation	games, questions	–	–	bias mentioned
Good&Co	questions	C, P	multiple studies	adverse impact
Harver	games, questions	S	–	–
HireVue	games, questions, video	C, P	–	4/5 rule
impress.ai	questions	S	–	–
Knockri	video	S	–	bias mentioned
Koru	questions	S	some description	adverse impact
LaunchPad Recruits	questions, video	–	–	bias mentioned
myInterview	video	–	–	compliance
Plum.io	questions, games	S	validation claimed	bias mentioned
PredictiveHire	questions	C	–	4/5 rule
pymetrics	games	C	small case study	4/5 rule
Scoutible	games	C	–	–
Teamscope	questions	S, P	–	bias mentioned
ThriveMap	questions	C	–	bias mentioned
Yobs	video	C, S	–	adverse impact

Table 1: Examining the websites of vendors of algorithmic pre-employment assessments, we answer a number of questions regarding their assessments in relation to questions of fairness and bias. This involves exhaustively searching their websites, downloading whitepapers they provide, and watching webinars they make available. This table presents our findings. The "Assessment types" column gives the types of assessments each vendor offers. In the "Custom?" column, we consider the source of data used to build an assessment: C denotes "custom" (uses employer data), S denotes "semi-custom" (qualitatively tailored to employer without data) and P denotes "pre-built." The "Validation?" column contains information vendors publicly provided about their validation processes. In the "Adverse impact" column, we recorded phrases found on vendors' websites addressing concerns over bias.

(Figure 2. Examining the Websites of Venders of Algorithmic Pre-Employment Assessments. Raghavan, Manish, et al. "Mitigating Bias in Algorithmic Hiring: Evaluating Claims and Practices." *Conference on Fairness, Accountability, and Transparency*, ACM, 2020)

Only one company out of the eighteen in the Cornell study follows the 4/5th rule and conducted a case study on validation in their algorithms: pymetrics (Figure 2.). Northeastern University performed a case study of pymetrics' methods to examine what steps the company takes to ensure accuracy and fairness in their models. pymetrics produces fairer and more transparent results than other third-party companies because it reverse engineers the model and implements mathematical debiasing constraints and objectives. pymetrics ranks candidates into three tiers: "Highly Recommended" typically for candidates that score above the 70th percentile, "Recommended" for candidates between the 50th and 70th percentile, and "Not Recommended" for candidates below the 50th percentile (Wilson). At each threshold, pymetrics tests the model

[1]While there are many gender identities beyond male and female, most data on this topic focus on binary gender identities, so this paper focuses on binary gender until the action items later in the text.

for fairness, ensuring that the hiring recommendations follow the 4/5th rule for all groups protected under Title VII before deploying the model (Wilson). pymetrics proves to be the best current option for algorithmic hiring at the screening stage. However, all algorithmic hiring companies and companies that implement these models internally must uphold or exceed the same standard of fairness to ensure that no hiring algorithms perpetuate discrepancies in the workforce.

RETHINKING PRIORITIES AND ENGINEERING INNOVATION PROCESSES

To ensure a fair model, companies must first rethink their priorities when formulating their hiring algorithms. Big data analytics is a “vehicle” that is only as good as the questions the engineers ask and the way they implement it (Bridgeford). Instead of asking the computer to produce the most optimal model, companies and their engineers must first require that the computer prioritizes fairness then produces the most optimal fair model. To ensure fairness is a priority, companies must implement three main changes when coding hiring algorithms.

First, the engineers writing the algorithm must mathematically ensure that the model follows the 4/5th rule. Ideally, engineers should add multiple constraints ensuring people from different identities are rated equally in the training data (Boyd). If “all of these notions of fairness cannot be fulfilled simultaneously,” engineers can add “multi-objective regularizations” that ensure the computer prioritizes minimizing the ranking differences between various identities over optimization (Köchling). After adding these factors, companies must ensure the recommendations for candidates the model has not yet seen follow the 4/5th rule. If the initial model fails this test, the engineers can adjust the constraints and objectives until the model fulfills the 4/5th rule for all protected groups. Second, companies must frequently update the model with more inclusive workforce data. Even with the best debiasing techniques, hiring

[1]While there are many gender identities beyond male and female, most data on this topic focus on binary gender identities, so this paper focuses on binary gender until the action items later in the text.

algorithms predict the success of future employees “based on current employees” (Raghavan) and utilize big data that mirrors “human biases” (Kennedy). As many fields in the workforce begin to have more representation and equitable hiring processes, engineers can utilize the increasingly inclusive data to produce increasingly fair models. Lastly, companies must ensure that the people implementing these models are from diverse backgrounds to increase fairness and transparency in the development process. While these steps will significantly improve fairness in algorithmic hiring, scholars must conduct more machine learning research to improve debiasing methods, and many next steps outside of the engineering design process can mitigate other fairness concerns on top of the design changes.

NEXT STEPS

Along with changes to the research priorities and the engineering innovation process in developing data-driven hiring algorithms, other next steps may improve the fairness of the algorithmic screening process. First, policymakers and lawmakers must implement legislation that holds companies accountable for discriminatory algorithms. Right now, companies want “plausible deniability” (O’Neil) and often use the “black box” in machine learning as an excuse for biased algorithms. No algorithm with discriminatory bias should be legal. Second, the mathematicians and computer scientists creating these algorithms must implement intersectional considerations. Specialists can currently ensure that different backgrounds are rated nearly equally in a model, but the algorithm must rate a person who falls into multiple underrepresented categories, such as a black woman in STEM, equally as well. Third, these algorithms must be made accessible to small companies. Even third-party companies with large datasets, like pymetrics, require their clients to have a large number of employees on which to train the model (Wilson). Because of this, small companies must rely on human screeners with individual bias

[1]While there are many gender identities beyond male and female, most data on this topic focus on binary gender identities, so this paper focuses on binary gender until the action items later in the text.

towards applicants. Although algorithmic hiring still need to improve, it still eliminates individual bias from the screening process, and all companies should find ways to decrease bias in their recruiting process.

CONSLUSION

Algorithmic hiring decreases costs, increases efficiency, and eliminates individual bias for companies in the recruiting process. Companies, however, build these algorithms in a society that embed systemic discrimination within the code. Without proper priorities and regulation, data-driven hiring processes perpetuate and amplify discrepancies in the workforce. Companies should only deploy data-driven hiring algorithms if they have undergone significant testing for bias against protected groups and follow the 4/5th rule as a minimum requirement. In addition, machine learning can potentially create fairer systems in applications far beyond algorithmic hiring. However, when these algorithms use people as their data, the people writing and regulating these models must ensure the model does not contain or perpetuate the flaws and discrimination within society.

[1]While there are many gender identities beyond male and female, most data on this topic focus on binary gender identities, so this paper focuses on binary gender until the action items later in the text.

Works Cited

Allal-Chérif, Oihab, et al. "Intelligent recruitment: How to identify, select, and retain talents from around the world using artificial intelligence." *Technological Forecasting and Social Change*, vol. 169, 2021, www.sciencedirect.com/science/article/pii/S0040162521002547.

Bridgeford, Lydell. "Experts Discuss Big Data's Effect on Hiring, Bias Claims." *HR Focus,* vol. 92, no. 9, 2015, pp. 4-6.

Boyd, Stephen. Personal Interview. 23 November 2021.

"Civilian Labor Force by Sex." U.S. Department of Labor, www.dol.gov/agencies/wb/data/lfp/civilianlfbysex.

Diaz, Jaclyn. "As AI Enters Hiring, Some Ask if It's Injecting New Biases." *Bloomberg Law,* 2019, news.bloomberglaw.com/daily-labor-report/ai-creeps-into-hiring-and-some-ask-if-its-injecting-new-biases.

"Employment and Earnings by Occupation." U.S. Department of Labor, www.dol.gov/agencies/wb/data/occupations

Goodman, Rachel. "Why Amazon's Automated Hiring Tool Discriminated Against Women." *ACLU*, 2018, www.aclu.org/blog/womens-rights/womens-rights-workplace/why-amazons-automated-hiring-tool-discriminated-against.

Kennedy, Elizabeth. "Can Data Drive Racial Equity." *MIT Sloan Management Review*, vol. 62, no. 2, 2021, pp. 9-11. *ProQuest*, www.proquest.com/docview/2479113313?accountid=14026.

Köchling, Alina, et al. "Highly Accurate, But Still Discriminatory: A Fairness Evaluation of Algorithmic Video Analysis in the Recruitment Context." *Business & Information*

[1]While there are many gender identities beyond male and female, most data on this topic focus on binary gender identities, so this paper focuses on binary gender until the action items later in the text.

Systems Engineering, vol. 63, 2021, pp. 39-54, link.springer.com/article/10.1007/s12599-020-00673-w.

O'Neil, Cathy. "Amazon's Gender Biased Algorithm is Not Alone." *Bloomberg*, 2018, www.bloomberg.com/opinion/articles/2018-10-16/amazon-s-gender-biased-algorithm-is-not-alone.

"Percentage of women workers in science, technology, engineering, and math (STEM)." U.S. Department of Labor, www.dol.gov/agencies/wb/data/occupations-stem?_ga=2.104888331.901278989.1638328337-277571395.1638328337.

"Protections Against Employment Discrimination Based on Sexual Orientation or Gender Identity." U.S. Equal Employment Opportunity Commission, www.eeoc.gov/laws/guidance/protections-against-employment-discrimination-based-sexual-orientation-or-gender

Raghavan, Manish, et al. "Mitigating Bias in Algorithmic Hiring: Evaluating Claims and Practices." *Conference on Fairness, Accountability, and Transparency*, ACM, 2020, dl.acm.org/doi/pdf/10.1145/3351095.3372828.

Raub, McKenzie. "Bots, Bias, and Big Data: Artificial Intelligence, Algorithmic Bias and Disparate Impact Liability in Hiring Practices." *Arkansas Law Review*, vol. 71, no. 2, 2018, pp. 529-570, heinonline.org/HOL/Page?handle=hein.journals/arklr71&id=549&collection=journals&index=.

Wilson, Christo, et al. "Building and Auditing Fair Algorithms: A Case Study in Candidate Screening." *Conference on Fairness, Accountability, and Transparency,* ACM, 2021, dl.acm.org/doi/pdf/10.1145/3442188.3445928.

[1]While there are many gender identities beyond male and female, most data on this topic focus on binary gender identities, so this paper focuses on binary gender until the action items later in the text.

"100 Years of Working Women." U.S. Department of Labor, www.dol.gov/agencies/wb/data/occupations-decades-100.

[1]While there are many gender identities beyond male and female, most data on this topic focus on binary gender identities, so this paper focuses on binary gender until the action items later in the text.