SOCIAL EQUITY ANALYSIS OF MACHINE LEARNING-BASED HIRING TOOLS

A Research Paper submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Matthew Whelan

Spring, 2023

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

ADVISOR

Pedro A. P. Francisco, Department of Engineering and Society

INTRODUCTION

Machine learning is rapidly transforming the world around us, most notably in employment recruitment – 99% of Fortune 500 companies use automated hiring software (Hu, 2019, para. 1). The primary applications are Applicant Tracking Systems (ATS) and Recruiting Management Systems (RMS) which help organizations track the pipeline of applicants through the recruiting process and automate administrative tasks like candidate scoring, respectively. Alarmingly, 94% of companies agree that such RMSes vet out qualified candidates because they do not match the exact criteria set forth by the job description (Fuller et al., 2021, p. 3). Not just that, but there is a high potential for bias to develop in such systems that could discriminate against certain social groups due to learned characteristics of the algorithms.

This paper will attempt to answer the following research question: how can machine learning be used as an equitable hiring tool that minimizes bias? The modern world is employing machine learning to increase efficiency and lower costs across many different industries and applications. However, there are many examples of machine learning in hiring causing more harm than benefit by generating unfair results and reinforcing existing biases. So, for machine learning to be truly effective, it must be implemented responsibly to prevent inequity from developing and turning it into a tool for discrimination. To accomplish this, diversity is essential. Diversity should start at the top in machine learning teams and work its way into the actual training data and models hiring tools utilize. This paper will use the framework of Social Construction of Technology (SCOT) to understand the manifestation of bias in automated hiring systems and will offer an answer to the research question centered around the idea of diversity.

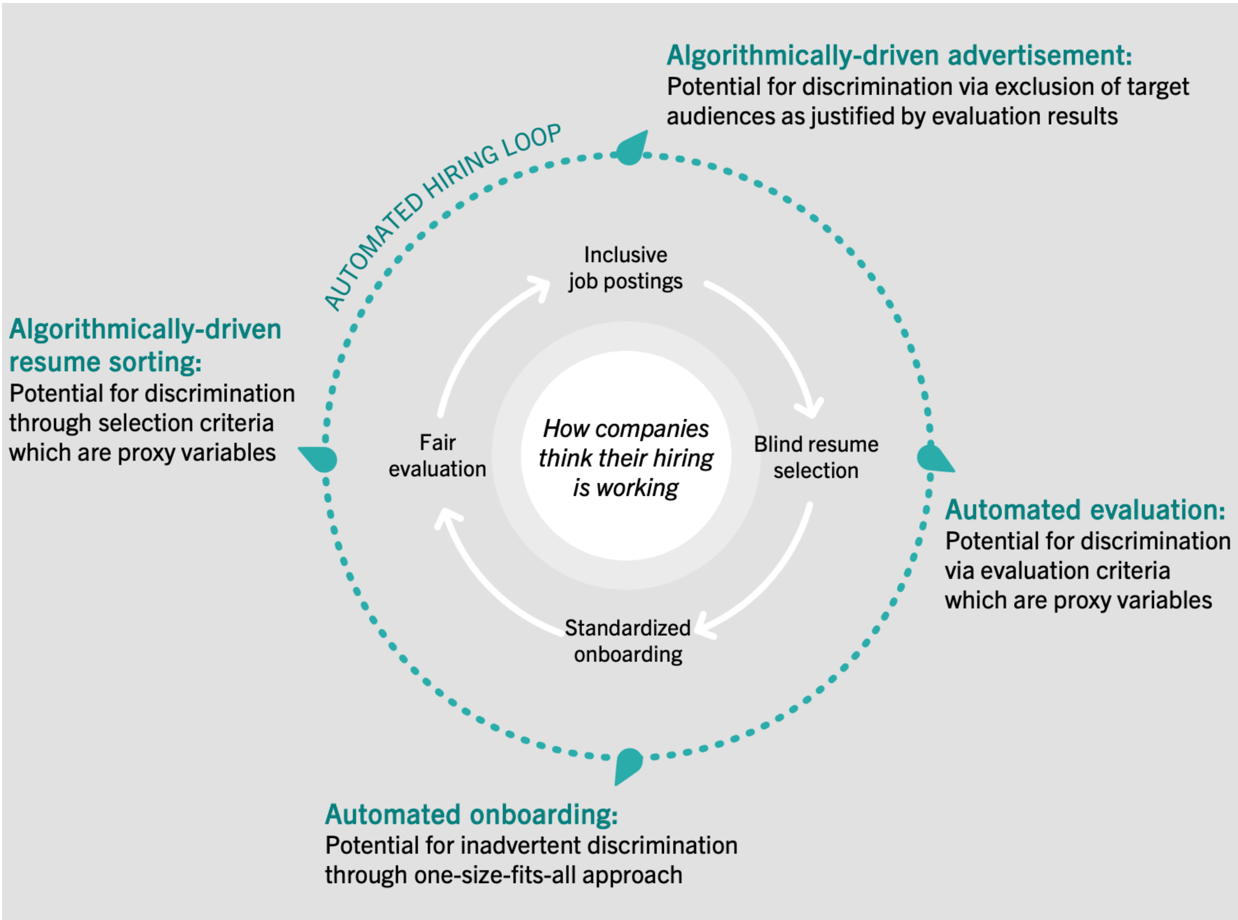
BACKGROUND AND SIGNIFICANCE

To understand how automated hiring systems function, it is first important to understand the process of machine learning. The first step is to gather and prepare training data, which is the information that the system will be trained on. This data is directly what the model 'learns' trends from, so it is imperative that this data is high quality and large enough to be sufficiently descriptive. This is arguably the most important step because a model absorbs characteristics of the data it is trained on, so if the data is biased against a certain subset of the population, the model will likely make similarly biased predictions that disadvantage that subset. The next step is to choose a specific machine learning model, which depends on the type of input data and the desired output. At this point, the model can train itself to find patterns, and over time, the developer can tweak the model by changing its parameters to try to deliver more accurate results.

In terms of hiring, there are algorithmic tools available for essentially every step of the hiring process. Recruiters can find candidates, and vice versa, through sourcing platforms like LinkedIn and Indeed. From there, companies use screening algorithms to analyze resumes, similar to how VMock and JobScan use algorithms to give feedback on resumes to help them pass through the company screening algorithm. Finally, there are also tools to automate the onboarding process to ensure new employees are welcomed into the company. The common strand behind these tools is they all rely on candidate data to infer how well the candidate will perform in the role. Figure 1 shows this so-called "automated hiring loop" where there are four general steps: job postings, resume selection, onboarding and evaluation. Each step is unique in its function and in its way to develop bias and discrimination based on its configuration and use case by the company.

Figure 1

Hidden Biases in Automated Employer Practices



Note: The automated hiring loop and its associated biases. Adapted from "Hidden Workers: Untapped Talent." By J. Fuller, et al, 2021, *Harvard Business School*.

Since bias can develop in a variety of ways in this automated hiring loop, it's clear that some action needs to be taken. For many applicants, the algorithm will not value unique experiences or qualities that don't specifically match the criteria set by the employer. This can lead to the denial of opportunity, and in the worst case, can lead to a continuous cycle of unemployment for deserving candidates. The worst part is that these candidates will never know how their information is used by employers, as these tools act as a "black box" that are completely hidden to the outside world. This is a pressing problem, as candidates should have the right to know how their information is being used, and how they are being evaluated if the process is mostly automated.

With their rise in prevalence, these tools have recently drawn some concerns whether they are discriminating against potential qualified candidates. The U.S. Equal Employment Opportunity Commission (EEOC) outlines certain protected characteristics to safeguard applicants and employees from employment discrimination. These characteristics, which include race, sex, disability status, and others, could be factors for some hiring algorithms, which is obviously illegal. To mitigate this, New York City passed a first-of-its-kind law in December 2021. This law requires that employers conduct an independent bias audit on the use of "automated employment decision tools," make the results publicly available, and disclose the data the tool is collecting (Moreno, 2022, para. 3). This law is the beginning of a growing wave towards regulation on algorithmic justice, as other states and Congress have introduced similar bills without being passed yet.

RESEARCH METHODS

This paper will use the literature review research method to analyze primary and secondary sources of machine learning's usage in modern hiring practices. Then, this paper will apply the Social Construction of Technology (SCOT) Theory to examine the roles various stakeholders have and the sociocultural embedding of these technologies. Finally, this paper will use this analysis to form constructive suggestions to reduce bias and improve machine learning hiring systems to yield a safe and just future.

RESULTS AND DISCUSSION

As stated earlier, the ATS and RMS are the principal components of automated hiring processes, and can either be sourced through a third party vendor or developed completely in-house. Many large corporations, like Amazon, opt for a completely in-house hiring pipeline, which includes an automated hiring tool that finds potential candidates that are a good fit for open roles. However, in 2018, Amazon shut down their hiring tool because it was discriminating against female candidates, penalizing the word “women” in resumes (Dastin, 2018, para. 6). The algorithm was trained on resumes submitted to the company over a 10 year period. The majority of these resumes came from men, which reflects the historical male overrepresentation in the tech industry. This shows how important training data is on algorithms; the hiring tool essentially learned that male candidates were preferable to female ones because of their abundance in the company’s past hiring.

One major counter argument to this finding is that women might have had less appealing resumes than men. Even if this is true, the algorithm still won’t give the next woman candidate a fair shot due to the history of women candidates in the past. Therein lies the whole problem with using learned features for an algorithm of this magnitude – one that can decide whether or not one remains unemployed or starts a new position. Some learned features should be important like previous employment history or academic success, but these should not be tied to protected characteristics. There’s another argument that these tools are simply reflecting what recruiters would have been doing, which is disproportionately hiring male applicants. In that case, what’s to say that this technology itself is inequitable, when in reality the whole system it reinforces is inequitable. This is a much hazier argument, as the bias would still fall on the employees making those choices which makes the company liable for discriminatory hiring practice lawsuits. However, automated tools can be corrected over time to remove certain prejudices, whereas there is obviously not a concrete way to remove prejudices from upper management in companies. If recruiters can input criteria or specifications for candidates, then the hiring tool is successful in its goal, but for all the wrong reasons. This would mean that human bias can infiltrate algorithms and reproduce itself in algorithm bias, although the actual role of recruiters in the automated hiring loop depends on the company. This can yield a wide range of outcomes, from recruiters being the main problem to algorithmic bias being the problem. Generally, the reality for many companies is somewhere in the middle, so a solution should look at both of these considerations. With regulation though, like the New York City law is proposing, there will be much more transparency to verify hiring decisions are made without factoring in protected characteristics. This would be a solution to this argument, and for both sides of the spectrum, as it will be much simpler to ensure that practices are balanced and not discluding one particular subgroup for automated hiring tools and recruiters' use of them.

Amazon's hiring tool sets a poor precedent for hiring algorithms, not only because they are rising in prevalence, but because these black boxes are all internal tools not visible to the outside world. In this way, algorithmic bias could be swept under the rug, without the knowledge of anyone, and thus used by companies to enforce inequitable hiring practices. Likewise, it is no wonder why New York City, with other states following suit, wants to break into the black box, so to speak, to see if these algorithms are making flawed recommendations. This effort is in good faith, as it is essential that machine learning, especially in applied to power systems like employment, are designed with equal representation of impacted groups. This might not be the case with many status quo algorithms though, as a recent study which found that a large portion of available online datasets are shown to not represent individuals equally across characteristics like sex and race. Buolamwini & Gebru (2018) found that many datasets, including National Institute of Standards and Technology (NIST) constructed datasets, represented lighter skinned people in around 80% of training examples, with darker skinned females only representing 4.4% of the overall dataset (p. 3). This is similar to the issue with Amazon's tools, as the training examples are simply not equal across all represented groups. This unequal representation propagates bias from the start of the training process, as models will overlearn features of the majority group and likewise not sufficiently learn features of the underrepresented groups.

Automated hiring tools have a challenging task due to the very limited set of information available on a resume. Due to this, algorithms will generally cling to certain proxy variables to represent a candidate's ability to fulfill job requirements, which was shown back in Figure 1. One example is employers defaulting to using college degrees as a proxy for a candidate's range and depth of skills. This comes as companies increasingly demand for applicants with college degrees; technology companies like Oracle, Intel and Apple require degrees in more than 90% of technology postings, much more than the national average of 52% (Langer, 2022, para. 10). This does not even account for the companies without a specific requirement for a degree, but it nevertheless shows the importance of a college degree in today's job market. Hiring algorithms have placed a high weight on degrees, whether that be a specification from the recruiter or not, but it might not even be for a good reason. Research has shown that non-graduates are nearly equal to graduates in various productivity and performance metrics and companies pay between 11% and 30% more for college graduates (Fuller, 2017, p. 2). So, it is very possible that the factors working inside of the black box of automated hiring algorithms are making inaccurate judgements, particularly for the component of college degrees.

Similar to how hiring algorithms will value certain proxy variables, research shows that these algorithms are quick to act on these proxy variables. Automated hiring tools will use a failure to meet some criteria (like a gap in full-time employment) as a basis for excluding a candidate from further consideration (Fuller et al., 2021, p. 3). This is precisely why many employers are wary of automated hiring tools' tendency to disqualify good candidates: they put too much emphasis on specific criteria instead of valuing unique experiences. Say one algorithm values college degrees because it learns that many recent new hires had degrees, then it will be much easier for that algorithm to automatically disqualify all non-graduates without any additional insight into them. Using one criteria as a basis for final decisions like this example can lead to generally biased results. Thus, these tools must take a holistic view of candidates, not valuing one variable too heavily over the rest to minimize the risk of bias.

Automated hiring tools represent complicated systems with many actors at play. A good start at interpreting this technology is through the Social Construction of Technology framework, which is a multidirectional view to analyzing the development of a technical artifact through its relevant social groups (Pinch & Bijker, 1987, pg. 28). Using this framework will show how relevant social groups shape automated hiring tools and likewise how these tools shape the same social groups. The largest social group is the set of applicants, which can be divided into past applicants (training data) and current/potential applicants. From Amazon's flawed hiring tool, it's clear that the profile of past applicants can have a large impact on how the model views future potential applicants. The set of potential applicants can change as the open positions and job listings shift, which relies on the actions of human resources executives overseeing such software. Human resource executives lead hiring managers, who actually use these tools to directly influence which applicants are moved along the process. The "designers" of this technology are the developers of the automated hiring tools. Developers use company culture and objectives to tweak the algorithm to find the set of applicants that best meet the criteria set out in the job listing, similar to changing parameters of a machine learning model. Looking back at the idea that the automated tools reflect the recruiter's intentions, if this is the case then the company culture and upper management also influence the recruiter's specification of hiring criteria. It's difficult to quantify the impact that recruiters' human bias has on automated hiring tools and their tendency to discriminate against certain applicants, but it's clear that recruiters have a very sizable role in the overall process.

All of these social groups have subtle relationships with one another, which thereby influence how automated hiring tools function. Hence, this framework reveals the relationships which makes seeing development of bias much easier. Human resource executives drive company culture and influence how hiring managers select new hires, which affects automated hiring decisions and where applicants end up. So, if a certain ethnic or racial group is hired at a much lower rate for a specific company, then automated hiring tools will systematically exclude such groups based on previous actions from the executives and hiring managers. This can be very damaging for a company long term if this cycle repeats and continuously decreases diversity.

There are various measures that companies can take to avert discrimination from coming out of automated hiring tools. The primary solution is one that has been an ongoing challenge for some time now: representation of minority groups in technology. As stated earlier, there is a historical overrepresentation of white males working in the technology industry. Among the top technology companies, only 2.5% of Google's workforce is Black while Meta and Microsoft are around 4%. Looking globally, only 22% of professionals in machine learning and AI are female (Howard, 2020, para. 5). Considering the role that machine learning plays for businesses, it is crucial that tools as impactful as automated hiring algorithms are created inclusively. The simplest way to do that is to ensure that the developer teams are inclusive and represent a diverse set of people. This is proven by a McKinsey study that shows that companies in the top quartile (25%) for gender diversity outperformed those in the fourth quartile by 25%; this number rises to 36% in the case of ethnic and cultural diversity (Dixon-Fyle et al., 2020, para. 5). Thus, it is clear that increasing diversity has a tangible impact on company success. With the coming regulation on automated hiring tools, companies should look inward and realize that diversity will reduce the bias these tools can produce and improve overall profitability.

Team diversity is not the only potential solution, but data and model diversity are as well. Data diversity was the reason why Amazon's hiring tool failed, and remains a problem as current datasets are not equally representing subgroups. In many cases, there are applicants from non-traditional backgrounds who firms do not have sufficient data on, so the algorithms will have trouble predicting their success and might just disregard them completely. To solve this problem, researchers at MIT added "exploration bonuses," to identify candidates the firm knows least about, in terms of educational background, work history or demographics. The researchers compared this exploration-based algorithm to the traditional static ones, and found that for the initial resume screening, the former passed along more than 3 times more Black and Hispanic applicants than the latter, as well as a 2.5 times better hire rate for candidates selected for an initial interview (Eastwood, 2020, para. 9). Using this method more candidates are given a chance, as the algorithm is designed to maximize quality without any preference for gender or ethnic diversity. This means that the Black and Hispanic candidates the algorithm selected were just as good as other candidates, because if not the model would've updated and learned to select fewer such candidates over time. An exploration-based approach like this one can greatly increase diversity of candidates selected by automated hiring tools by giving historically underrepresented candidates an actual chance in the hiring process.

The final call to change within this space is one that is likely inevitable for automated hiring tools – regulation. Although it is early in the process, the world has recognized that there must be some type of oversight to ensure that candidates are not discriminated against. The passed New York City law sets out a broad framework, including a mandatory bias audit from an independent party. Enforcing an audit is a step in the right direction because it imposes a general set of standards for these algorithms to eliminate egregious practices. Businesses have been required to issue audited financial statements for their stakeholders, because their internal practices appear as "black boxes" to the outside world. This is absolutely analogous to hiring algorithms, as audits will install the same accountability for unethical practices or unfair advantages occurring in the algorithm's "black box." Again, since it is very early in the regulation timeline, the specifics of algorithm bias audits are unclear, but hopefully there is eventually a common standard for conducting audits. Therefore, with increased regulation on algorithmic hiring tools, there will be a more level playing field for those seeking employment through bias reduction and transparency throughout the hiring process.

CONCLUSION

There is endless potential for automated hiring tools to revolutionize the job market and firms' hiring pipelines. However, left unchecked, algorithms can perpetuate the same biases and discrimination found in existing hiring practices. There are many past examples of algorithmic tools yielding biased results, so it is imperative that action is taken to prevent future discrimination. This paper has outlined three key areas where action can be taken to minimize bias and make automated hiring tools more equitable: diversity in developer teams, model and data diversity and algorithmic auditing. Each of these areas represent distinct ways to alter the status quo in the automated hiring space for the better of firms and applicants alike. Based on the social group interaction from the SCOT framework, with more diverse management using more inclusive tools with third party auditing, the likelihood of certain subgroups being excluded from the hiring process is significantly lower. Improving automated hiring practices is one step towards a more equitable, inclusive and safe future.

REFERENCES

Buolamwini, J., & Gebru, T. (2018, February 4). *Gender shades: Intersectional accuracy disparities in commercial gender.* MIT Media Lab. Retrieved October 11, 2022, from http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf

Dastin, J. (2018, October 10). *Amazon scraps secret AI recruiting tool that showed bias against women*. Reuters. Retrieved October 11, 2022, from https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

Dixon-Fyle, S., Dolan, K., Hunt, D. V., & Prince, S. (2020, May 19). *Diversity wins: How inclusion matters*. McKinsey & Company. Retrieved March 14, 2023, from https://www.mckinsey.com/featured-insights/diversity-and-inclusion/diversity-wins-how-inclusion-matters

Eastwood, B. (2020, September 30). *Exploration-based algorithms can improve hiring quality and diversity*. MIT Sloan. Retrieved March 14, 2023, from https://mitsloan.mit.edu/ideas-made-to-matter/exploration-based-algorithms-can-improve-hiring-quality-and-diversity

Fuller, J., Raman, M., et al. (October 2017). Dismissed By Degrees. Published by Accenture, Grads of Life, Harvard Business School.

Fuller, J., Raman, M., Sage-Gavin, E., Hines, K., et al (September 2021). Hidden Workers: Untapped Talent. Published by Harvard Business School Project on Managing the Future of Work and Accenture.

Howard , A., & Isbell, C. (2020, September 21). *Diversity in AI: The invisible men and women*. MIT Sloan Management Review. Retrieved December 1, 2022, from https://sloanreview.mit.edu/article/diversity-in-ai-the-invisible-men-and-women/

Hu, J. (2019, November 7). *99% of Fortune 500 Companies use Applicant Tracking Systems*, Jobscan, https://www. jobscan.co/blog/99-percent-fortune-500-ats/.

Langer, C. (2022, February 11). *Skills-based hiring is on the rise*. Harvard Business Review. Retrieved from https://hbr.org/2022/02/skills-based-hiring-is-on-the-rise

Moreno, J. (2022, August 29). New York City AI Bias Law Charts New Territory for Employers. Bloomberg Law. Retrieved from https://news.bloomberglaw.com/daily-labor-report/new-york-city-ai-bias-law-charts-new-territory-for-employers

Pinch, T. J. & Bijker, W. (1987). The Social construction of facts and artifacts. In The Social construction of technological systems: New directions in the sociology and history of technology. Cambridge, MA: MIT Press.