Bias and Discrimination in Hiring Algorithms

A Research Paper Submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Sciences
University of Virginia Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree
Bachelors of Science, School of Engineering

Justin Tran

Spring 2024

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
Dr. Gerard Fitzgerald, Department of Engineering and Society

Introduction

Machine learning (ML) is a subset of artificial intelligence (AI) that is most closely associated with the capability of machines to imitate human thinking. Arthur Samuel, widely considered the father of machine learning, described it as “the field of study that gives computers the ability to learn without explicitly being programmed.” (Brown, 2024) The deployment of ML applications is becoming more common every day. It is seen in facial recognition, healthcare, fraud detection, and predictive analytics. It is growing so fast, that the ML market size in the U.S. was valued at $4.74 billion in 2022 and is expected to grow from $6.49 billion in 2023 to $59.30 billion by 2030 (Growth Analysis, 2023). ML, in theory, exhibits value by mitigating human biases and facilitating autonomous processes. Often, however, practical deployment of these algorithms reveals deeply embedded biases that are often left undetected until significant harm occurs.

Bias in ML emerges when algorithms produce outcomes that unfairly discriminate, often human and or societal prejudices. This bias typically originates during the data collection and training phases, where flawed or unrepresentative data leads to skewed algorithms. Deploying biased algorithms in critical processes, intended to enhance efficiency, can inadvertently harm marginalized groups. Instead of eliminating human biases to ensure equitable outcomes, these systems reinforce existing disparities, illustrating the need for increased scrutiny and correction in the development of ML technologies to genuinely support fairness and inclusivity.

Correctional offender management profiling for alternative sanctions, or COMPAS, was a product utilizing ML to predict the likelihood of reoffence by criminals. It was soon discovered that the algorithm was falsely labeling black defendants as future offenders at twice the rate of their white counterparts (Mattu et al., 2016). In 2021, there was a report that mortgage lenders were 80% more likely to deny black applicants, 40 percent more likely to turn down Latino applicants, 50 percent more likely to deny Asian/Pacific Islander applicants, and 70 percent more likely to deny Native American applicants while using ML to screen potential borrowers. (Martinez, 2021). In facial recognition technology, accuracy in recognizing white males can reach as far as 99% while error rates for recognizing black women can be upwards of 35% (Hassanin, 2023). Scenarios such as these illuminate the unintended biases present in ML technologies, putting into question the process of their development and deployment which often result in inequitable treatment across marginalized demographics.

It is evident that the identification and mitigation of bias within ML algorithms presents a significant challenge. Especially when the deployment of these technologies can affect a person's very livelihood, it is paramount that the sources of error are examined, and the effects of said errors are documented. By mitigating the damages done to both demographics that are most at risk of marginalization and the developers of said ML applications, fairness is a commonplace that can easily be achieved.

In this paper, the presence, effect, and origin of bias within ML, specifically within hiring practices utilizing resume screening technology, will be examined. To do this, a literature review will be conducted to document and explore cases of when bias embedded within hiring algorithms resulted in the discrimination of certain racial and gender groups. Additionally, to illustrate the connections anywhere between the job applicants, resume screening software, and developers, the actor-network theory (ANT) will be deployed. Using this framework, a deeper understanding of the impact of the design and implementation of ML software in hiring practices will be made. With this analysis, the extent to which the integration of ML in hiring algorithms either contributes to or mitigates gender and racial discrimination within modern employment practices will gradually be made clear.

Background

Where it was once common practice to simply walk into a business, fill out a job application, and drop off a resume, filling out a job application is made significantly easier and more accessible for job seekers in this era of high connectivity via the internet. Instead of physically filling out a job application and having to print out copies of resumes, the average job seeker is able to simply create an account on LinkedIn and press the “Easy Apply” button on any job posting that piques their interest regardless of the time and location. This shift has had major benefits for both applicants and employers alike as applicants are applying to jobs quicker and easier, and employers are filling openings on their teams with qualified and competent candidates. With that being said, this monumental shift in the recruitment process between applicants and employers does not come without its own shortcomings.

The recruitment process is a daunting task, one that requires many man-hours and meticulous steps in order to source the best candidates possible for the task at hand. Typical steps hiring managers or human resource (HR) representatives take during this recruitment process include but are not limited to resume screenings, screening calls, assessments, and interviews. With corporate job postings attracting an average of 250 applicants (https://zety.com/blog/hr-statistics), recruitment processes amongst hiring managers and recruiters have evolved in order to accommodate the mass flux of applications coming from online job postings. To facilitate a smooth and efficient process for identifying talent, the implementation of AI has been adopted into many facets of the recruitment process for most companies. So much so, that it is now considered mainstream for HR professionals or recruiters to utilize some form of AI during the recruitment process. 56% of HR organizations already use AI for talent acquisition and for the third year in a row, implementing and improving technology, especially AI, is a top priority for HR teams (Delaney, 2020) (Turner, 2023).

The term for this evolution recruiting has been coined as “AI recruitment.” It is the use of AI during recruitment to automate, streamline, and simplify various aspects of the recruitment workflow (Leoforce). In doing so, recruiters are able to focus on building closer relationships with candidates, improve hiring strategies, and eliminate unconscious bias in talent acquisition. Additionally, with the help of AI, recruitment is made easier and more efficient as recruiters can now personalize and automate hiring strategies simultaneously across various channels such as job boards and advertisements. Some of the ways the implementation of AI is making an impact on the recruitment landscape are seen in resume screening software, digitalized interviews, and personalized chatbots.

By utilizing AI-based tools to source and connect with candidates, a process that would traditionally take 14 hours per week can be cut significantly. With this shift in the recruitment process that has AI as the main driver, it is important to discuss the inadvertent consequences that may arise such as gender and racial bias deeply embedded within these AI-based tools.

STS Framework

To better understand the impact of ML and AI on hiring practices, we will be using the actor-network theory, a sociological framework, to understand the actors that play a role in forming the recruitment process. Central to ANT is the principle of generalized symmetry the notion that society is formed through an ontology of relationships among diverse entities, called "actors," which can include humans, non-human entities, and technological artifacts (Nickerson, 2024). John Law, a major proponent amongst others of ANT, argues that these actors engage in dynamic relationships that shape and are shaped by the network (Law, 2007). Understanding the complex networks and relationships between these actors will provide an understanding of how they shape social and technical realities. By applying ANT as a framework to examine the network of human and non-human actors in the recruitment process, intersections within this vast web of interactions can be identified as points of intervention. By normalizing the effects of human intention with the influence of non-human entities, a holistic understanding of how biases are embedded and perpetuated. By treating ML as an active participant in the network, ANT can reveal how certain demographics are systematically disadvantaged and provide insight into ways to address and mitigate such biases. To begin this investigation, one must first map the involved actors within the network. Concerning ML and hiring practices, this includes:

- Job Applicants: The individuals submitting their resumes, affected by the software's biases.
- Employers and HR Professionals: Those who decide the quality and use of the software in the recruitment process.
- Developers and Designers: The individuals and teams who create and maintain the software.
- Resume Screening Software: The algorithms that analyze resumes, including their programming logic and data they're trained on.
- Societal Norms and Biases: The broader cultural and social beliefs about race and gender along with the recruitment process as a whole.

- Regulatory Frameworks: Laws and guidelines that govern employment practices and anti-discrimination policies.

The next step of this exploration is to then map the relationships between each of the actors. Each actor has a specific interaction with another, which eventually results in the massive web of interactions that lays the foundation of the network. For example, an obvious relationship can be drawn between the developers and the resume screening software that is deployed. The resume screening software used by the company is developed by a team or individual. Another relationship that can be drawn from the list of actors can be the job applicant and the employers. The job applicant chooses to apply for a job because of the culture and work done at the company that is set by the employer. HR professionals and regulatory frameworks are another one. The regulatory frameworks in place guide HR professionals in doing their jobs in a fair and equitable way, which ultimately influences how they perform their talent acquisition duties. These are just a few examples of relationships that can be made between the list of actors within this network. The beauty of ANT is that for each actor, their influences on another actor are endless, which is why Law depicts networks while using ANT as constantly dynamic, always relying on each other to actively form reality. Changes in technology, law, societal attitudes, or practices among job seekers and employers, big or small, could potentially shift the entirety of the network, potentially reducing or exacerbating biases.

The next step in utilizing ANT would be understanding actor roles and influences within the system. Consider the role each actor plays in the network and their influence on outcomes. ANT suggests that actors do not operate in isolation, their roles and the power they wield are defined through interactions. For instance, developers wield significant power through their ability to encode criteria into the resume screening software, potentially embedding their own biases and preconceptions into the algorithm. Employers influence the system by their selection of software, which would reflect and perpetuate their organizational culture and biases. Legal and regulatory bodies can act as counterbalances to biases within organizations but are also influenced by broader societal norms and lobbying from wealthy stakeholders. By taking this step, we begin to get closer to the bigger picture and understand why biases are allowed to slip through the cracks and realize themselves within society.

The next and final step of using ANT would be to identify points of intervention. By having mapped out the ontology of actors and interactions alike, problematic sections within this vast web of interaction and influences between actors can be identified so that interventions can be made to reduce bias. These might include developing more inclusive software design practices that actively seek to minimize bias. Using ANT in this way provides a holistic view of the recruitment process, highlighting how technologies like resume screening software are not merely tools but active participants in a network that shapes and is shaped by, human actors and societal forces. This approach can help uncover the subtle ways biases are embedded and perpetuated within the recruitment process, offering a foundation for developing more equitable practices.

Because there is a plethora of ways to combine each actor and examine the interactions and influences of each, for simplicity's sake we will explore an example between two actors. For instance, let us consider the relationship between the job applicant and the resume screening software. The obvious connection between the two actors is that the job applicant fills out the job application and provides a resume while the resume screening software scrubs the resume for data and determines whether or not to move forward within the hiring process with the candidate. Now to get a better understanding of this minute section of the entire system, let us consider each of their roles and influences. The applicant's role in this scenario is to provide a resume that is able to be properly ingested by both the software and the recruiter should the resume make it past the initial screening. The software's role in this scenario is to read the resume, process the information provided by the job applicant, and provide ratings and insights on the candidate based on its algorithm. Depending on the criteria and the data that was used to train the algorithm, factors such as successful and unsuccessful performances from current employees, skills, experience, and qualifications can play a crucial role in determining if the candidate is turned away immediately or shortlisted for an interview. Additionally, should the resume screening algorithm utilize continuous machine learning, then successful candidates from the past will also impact the approval rate of sequential job applicants.

Now we must understand how each of these roles from each of these actors might influence each other. Because applicants know they must first pass the resume screening process that is likely to be done by some form of AI, they will adjust their resume accordingly to best fit the idea of the "perfect resume." This means applicants are adding impressive numerics to quantify achievements, tailoring each resume to include technical buzzwords from the job description, and using strong action verbs to exude confidence and competence. The software on the other hand will continually adjust itself upon each candidate depending on whether the applicant was a successful hire or not. If the resume is passed along for the next step of the recruitment process and eventually hired, the algorithm will recognize the resume as a template for other candidates. One actor in this case can not survive without the other. They are both benefiting from each other and shaping the outcomes of each other within even this smaller-scale system within the whole network.

Now we know each of their roles and how they influence each other, the next step is to create meaningful insights relating to the issue at hand. Although the AI-based resume reading tool in theory should eliminate human bias, bias may be exacerbated. If the historical data used to train the algorithm contains biases, such as a predominance of successful candidates being from a certain demographic, then the algorithm is likely to learn and replicate these biases. For example, if most past hires are from a particular gender or racial background, the algorithm might prioritize candidates from these groups. In an applicant's endeavor to "beat the system" and pass the resume screening portion of recruitment, who is to say they will not lie on their resume or simply add messages in white text to bypass the screening algorithm?

From these insights, interventions can be made in order to prevent bias from being embedded deep within the software and deployed for use. Due to the nature of ML, oftentimes time the inputs and data going into the training model are known and the end product is functional, however, what goes on in between is often "black box" as the machine is responsible for its training. The internal workings of algorithms while being allowed to learn are often not transparent or understandable to users or even to some developers. One such solution could be to enhance transparency around how resume screening software works and its automated tracking systems. Prioritize understanding what takes place within the black box in order to provide an algorithm truly without bias. Another solution could be shifting organizational cultures and practices altogether to prioritize diversity and inclusion beyond just improving the screening process. That way improving fairness and equality within the recruitment process landscape may extend well past merely fixing software.

ANT has profound implications for the study of socio-technical systems and the understanding of social phenomena. By using the ANT's principle of generalized symmetry treating humans and non-humans as equals, it allows for a more nuanced analysis of how technology, objects, and ideas influence human behavior and social organization. ANT's focus on the dynamic nature of networks and the process of translation highlights the fluidity of social structures and the active role that actors play in shaping them. Moreover, the concept of performativity challenges conventional notions of objectivity and invites a reexamination of how knowledge and facts are produced.

In summary, the actor-network theory provides a robust framework for analyzing the interplay between human and non-human actors within complex networks. It challenges traditional distinctions between society and technology, suggesting that they are co-produced through interactions among diverse actors. By focusing on the relationships and processes that constitute networks, ANT offers valuable insights into the construction of social reality and the dynamics of socio-technical systems.

Literature Review

Fairness in hiring practices has long been a subject of interest. In a study by the Society for Human Resources Management, 1,500 people from all 50 U.S. states in 2022 participated in an experiment to understand the influence of ethnic-sounding names on job candidacy. It was revealed that names perceived to be white were considered at rates twice as high as candidates perceived to be black (Abel, 2023). Additionally, because real-world hiring managers typically spend less than 10 seconds reviewing resumes during initial screening processes, it would make sense that mental shortcuts, including prejudices and racial stereotypes, would be used in order to access job applicants. In the same study, it was discovered that when the time was limited, participants would discriminate against black-sounding names 25% more than if time restrictions were not imposed (Abel, 2023). It is easy to see why the implementation of ML algorithms in the hiring process may seem advantageous in minimizing bias by eliminating human decision-making as a whole. Still, often this is not the case.

In 2014, Amazon's ML specialists began working on an application to digest resumes with the hopes of producing a list of only the top most qualified candidates. To do this, they created an algorithm that would scrape information off resumes and compare them using patterns found on resumes sent to Amazon over the past 10 years (Dastin, 2018). Amazon appeared to have developed an ideal system for narrowing down resumes to the best candidates. This would have been the case had the algorithm not exhibited preference towards male applicants.

What initially appeared as an ideal solution quickly became problematic, as it was discovered that Amazon's algorithm was inadvertently favoring male candidates. This bias stemmed from the algorithm's reliance on resume data sent to Amazon over the past decade, which predominantly featured male candidates. Consequently, the system learned to prefer male applicants, penalizing female candidates and allowing for the introduction of bias into the automated hiring processes.

The oversight of gender bias in this AI-driven hiring algorithm reflects a broader issue within the composition of the tech industry. The predominantly male developers of Amazon failed to recognize the implications of using male-centric data sets, inadvertently allowing their own biases into the technology. This issue is further perpetuated by the significant gender disparities within the tech industry itself, where top companies demonstrate a clear underrepresentation of women in technical roles. In technical roles across prominent tech companies such as Apple, Facebook, Google, and Microsoft, gender representation varies, with Microsoft having the lowest percentage of female employees at 19%, while Apple leads slightly higher at 23% (Huang, n.d.). The remaining companies fall within this range, illustrating the ongoing challenge of gender diversity and inclusivity in tech. Such disparities not only highlight the sector's gender imbalance but also illustrate the need for diverse perspectives in technology development to prevent perpetuating existing societal biases.

This scenario exemplifies how slight oversights on data procurement for ML training can lead to biased AI systems, especially in automated hiring processes. In fact, reports indicate that 70% of all companies and 99% percent of Fortune 500 companies already use AI-based and other automated tools in their hiring processes (Venzke et al., 2023). The difficulty experienced by a leading company like Amazon implies that other organizations might also need help to achieve fairness in their automated hiring practices.

Conclusion

In summary, this thesis has explored the complex relationship between ML algorithms and bias in hiring practices, revealing how even well-intentioned technologies can inadvertently perpetuate discrimination. Through the lens of ANT, we've examined the intricate networks involving human and non-human actors, highlighting the role of data, algorithms, and societal norms in shaping employment outcomes. Despite the potential for ML to offer unbiased decision-making, cases like Amazon's hiring algorithm demonstrate the embedded biases within these systems. The necessity for vigilant, ongoing scrutiny and correctional measures will ensure that AI in hiring not only streamlines recruitment but also facilitates fairness and inclusivity. By addressing the origins and impacts of bias, this thesis seeks to further discuss areas of improvement not only in ML but in hiring practices as a whole.

References

Brown. (2024, February 6). Machine learning, explained. MIT Sloan School of Management. https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained

U.S. Machine Learning Market Share | Growth Analysis [2030]. (2023, May). Fortune Business Insights. https://www.fortunebusinessinsights.com/u-s-machine-learning-ml-market-107479

Mattu, J. A., Larson, J., Kirchner, L., & Mattu, S. (2016, May 23). Machine bias. ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Martinez, E., & Kirchner, L. (2021, August 25). The secret bias hidden in mortgage-approval algorithms. The Markup. https://themarkup.org/denied/2021/08/25/the-secret-bias-hidden-in-mortgage-approval-algorithms

Hassanin, N. (2023, August 23). Law professor explores racial bias implications in facial recognition technology. Ucalgary News. https://ucalgary.ca/news/law-professor-explores-racial-bias-implications-facial-recognition-technology

Delaney, J. (2020, November 19). 2021 Hiring Trends—Automation and Artificial Intelligence. LinkedIn. https://www.linkedin.com/pulse/2021-hiring-trends-automation-artificial-jessica-delaney/

Turner, J. (2023, October 17). What Will HR Focus on in 2024. Gartner. https://www.gartner.com/en/articles/where-hr-will-focus-in-2024

AI recruitment guide: All HRs need to know about AI in recruiting. (n.d.). Leoforce. Retrieved April 7, 2024, from https://leoforce.com/guides/guide-to-ai-recruitment-for-hr-to-understand-ai-recruiting-tools/

Nickerson, C. (2024, February 13). Latour's Actor Network Theory. Simply Psychology. https://www.simplypsychology.org/actor-network-theory.html

Law, J. (2007, April 25). Actor Network Theory and Material Semiotics. Heterogeneities. http://www.heterogeneities.net/publications/Law2007ANTandMaterialSemiotics.pdf

Abel, M. (2023, September 27). Viewpoint: How a person's name can prompt hiring discrimination. SHRM. https://www.shrm.org/topics-tools/news/employee-relations/viewpoint-how-persons-name-can-prompt-hiring-discrimination

Dastin, J. (2018, October 11). Insight—Amazon scraps secret AI recruiting tool that showed bias against women. Reuters. https://www.reuters.com/article/idUSKCN1MK0AG/

Huang, H. (n.d.). Dominated by men. Reuters. Retrieved February 7, 2024, from https://fingfx.thomsonreuters.com/gfx/rngs/AMAZON.COM-JOBS-AUTOMATION/010080Q91F6/index.html

Venzke, O. A., & Cody. (2023, August 23). How artificial intelligence might prevent you from getting hired. American Civil Liberties Union. https://www.aclu.org/news/racial-justice/how-artificial-intelligence-might-prevent-you-from-getting-hired