Examining Circumstances Around Discriminatory Machine Learning Hiring Models and Why Companies Have Permitted Their Use

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

Kyle Pecos
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
MC Forelle, Department of Engineering and Society

Introduction

Companies and businesses are always looking for new ways to increase their efficiency and cut operational costs with the development and creation of new technologies. One such technology that has undergone rapid development in recent years is the field of artificial intelligence (AI). By incorporating AI-related technologies into the workplace, companies have begun to replace human involvement in exchange for automated systems which are cheaper and supposedly more efficient (Schumann et al., 2020). Some companies currently using AI even estimate that by 2028 around 70% of their workforce will have AI assist automating or assisting with job tasks (Kong, 2023).

In these companies, the hiring process has begun to commonly use machine learning, a subset of AI, to automatically scan resumes and data from applications and potentially contribute to whether an individual will be hired or not. Machine learning attempts to teach programs how to learn information, detect patterns from provided data, and make decisions based on their training (Mahesh, 2020). As many executives and higher-ups in companies generally do not have backgrounds in machine learning, they may lack knowledge of its drawbacks and primarily focus on its reported benefits. Machine learning can be a powerful asset to companies if properly used, however, if precautions are not taken when building and training such models, these systems can make inaccurate and discriminatory decisions. Unrepresentative and skewed data can lead to bias forming within systems which can adversely affect the livelihoods of many, especially minority groups who commonly face this issue which is why I choose to research this topic further (Schumann et al., 2020).

Companies in the U.S. have been using inequitable datasets when training machine learning models due to executives lacking technical and ethical knowledge in the field and since there is an absence of financial or legal incentives to ensure these datasets are properly validated. To argue this point I will first provide a literature review of common types of biases, real-world examples of models discriminating in the hiring process, and the current U.S. initiatives to monitor the use of machine learning technologies. Then, I will document the methodology by which I conducted this literature review as well as give insight as to why I chose to investigate these certain topics of interest. Next, I will analyze my findings using the framework of machine ethics to understand the role of responsibility when developing these algorithmic hiring programs and examine a case study regarding a lawsuit against iTutorGroup. Finally, I will synthesize possible future steps that employees, developers, and lawmakers can take to help tackle this issue.

Literature Review

If a dataset is not validated for its completeness, quality, and diversity, a trained model might produce different types of biases with the most relevant ones being statistical, record, and structural bias. Statistical bias occurs when a given dataset does not accurately represent the larger population and commonly occurs since data pertaining to certain minority groups may be more difficult to gather (Costa et al., 2019). This issue can be reduced through injecting synthetic data that is based on pre-existing data into a database (Géron, 2019). Record bias forms when a training dataset is partially incomplete and missing values as it is unlikely a dataset will have all attributes for every data point (Costa et al., 2019). As some machine learning algorithms require all values to be filled, there are two primary methods of dealing with missing entries. The first is simply removing any data points with missing values whereas the second fills in missing data with a certain metric such as the average or median (Géron, 2019). Structural bias occurs when legitimate correlations still ultimately produce discriminatory bias and are the most difficult to prevent or rectify (Costa et al., 2019).

Pre-existing gender and race imbalances in certain industries can lead to machine learning models forming statistical bias and ultimately reinforcing a biased status quo. One such detailed instance of this occurring in recent years was with Amazon's applicant sorting model in 2014. The model was shown to be able to sort applicants more efficiently and timely manner, however, it also discriminated against female applicants as a result of being trained on biased company data (Dastin, 2018). The model was trained on resumes from the past ten-year timeframe with a large majority of these applications coming from men (Dastin, 2018). Additionally, the model was shown to increase ratings for "masculine" language that would generally be found on a male's resume (Dastin, 2018). As the programmers never properly scrutinized the biased training dataset, the project was ultimately deemed a failure and had to be scrapped.

Many companies have begun to utilize machine learning hiring algorithms, but few of them are transparent about how these models are validated and tested for bias if at all (Raghavan et al., 2020). A conducted study with eighteen vendors that utilized machine learning hiring algorithms demonstrated that very few vendors explicitly discussed concerns about mitigating possible bias, whether their models are validated, and whether the use of such systems can be justified (Raghavan et al., 2020). Studies have been conducted demonstrating ways to increase algorithmic fairness in machine learning models which include ensuring datasets are comprehensive and diverse, using external validity testing, and having a diverse range of developers (Sengupta et al., 2018).

In an attempt to monitor the use of AI and machine learning-related technologies in employment practices, certain initiatives and municipal legislation have been enacted. The U.S. Equal Employment Opportunity Commission launched an initiative in primarily ensures that employers do not break Title VII of the Civil Rights Act of 1964 which covers protection, “on the basis of race, color, religion, sex, or national origin” (Equal Employment Opportunity Commission, 2023b). However, the Commission has been seen to deal with other anti-discrimination laws as well as seen in the recently filed lawsuit against iTutorGroup, Inc. (Equal Employment Opportunity Commission, 2022). Their automatic application software was shown to reject certain applications based on their age which violates the Age Discrimination in Employment Act (Coyle, 2023). Additionally, certain states and municipalities have begun to draft laws that monitor the use of automated employment decision tools which encompass machine learning models (Rogers and Reed, 2021). New York City’s Local Law 144 of 2021 was recently passed in 2023 and requires all AEDT to undergo a yearly bias audit by an independent auditor (NYC Consumer and Worker Protection, 2023).

Based on my findings, I conducted my analysis using the framework of machine ethics which emphasizes the development of algorithms in programming that can make moral decisions in the social, technical, and political landscape (Allen et. al., 2006). Machine ethics applies a hybrid of primarily deontological and consequentialist theory in its rationale of responsibility (Tolmeijer et al., 2020). Developers should be aware of the effects their products might have on society and have the moral agency to take responsibility for the decisions that their models make. They should also strive for their models to act as artificial moral agents, programs that “honor privacy, uphold shared ethical standards, protect civil rights and individual liberty, and further the welfare of others” (Allen et. al., 2006).

Research Question and Methods

My proposed research question consists of the following: “What historical, economic, and political factors have contributed to companies using biased, inequitable datasets as part of training machine learning models for their hiring process?” To answer such a question, I have conducted an extensive literature review using primary legal documents and relevant secondary sources of machine learning theory with real-world application and a case study regarding the lawsuit against iTutorGroup. These hiring models directly affect the livelihoods of countless individuals who are looking for work and especially have the potential to negatively impact certain minority groups.

I first gathered secondary sources that detailed the differing common types of algorithmic bias and the methods by which they form. Following this, I examined studies documenting methods to alleviate machine learning bias and its applicability to machine learning hiring models. I hoped to determine how far the field of algorithmic fairness had currently progressed and whether proposed solutions could be adequately applied to machine learning hiring models. I then found secondary sources regarding the perception of machine learning models both from the average individual and those who have been educated in the field. I wished to survey commonly held misconceptions of the field for people such as a company executive who would likely not be aware of machine learning’s intricacies. Afterward, I sought out information on whether machine learning educated individuals had an additional background of taught ethics to guide their decision-making. Having knowledge of machine learning can only be so useful without a proper ethical understanding of the field’s relationship with the social.

Next, I examined primary and secondary legal sources detailing the current state of U.S. anti-discrimination laws on the federal, state, and local levels. I was curious to figure out if specifically tailored legislation for the use of machine learning in the field of hiring had been passed given its rise only recently occurring in the past decade. Subsequently, I searched for secondary sources detailing real-world scenarios of machine learning models discriminating during the hiring process and was able to find the cases regarding Amazon and iTutorGroup. I planned to examine the relationships between the parties of the developer, higher-ups, and the models themselves to get a better understanding of how the bias systems came to be.

I specifically chose to conduct a case study on the lawsuit between the EEOC and iTutorGroup as it was the most well-documented real-world example I could find. I primarily examined and scrutinized three main aspects of this case in my analysis using both primary legal sources and secondary sources from EEOC. The first was researching and discovering the criteria for the company to be held liable for discrimination and how the lawsuit came to be. The second was to examine how the affected parties were impacted and what recompense they requested. The third was to investigate how the company responded to the allegations and whether they took responsibility for failing to develop an artificial moral agent.

Using the framework of machine ethics as a guideline, I homed in on the ethical considerations and moral values for both programmers and executives when developing machine learning models. It remains important for both developers and executives to realize that these systems are not infallible and have the potential to make immoral decisions if precautions are not taken during development. Those who are aware of machine learning's potential downfalls should have the moral agency to push for thorough validation and testing in these hiring systems.

Analysis

Employees in companies have begun to place too much trust in machine learning models as a result of not understanding or knowing the intricacies of the field and its potential to produce bias. One study of students without a machine learning or computer science background found that many overestimated the capabilities of machine learning systems and assumed such models did not need additional human intervention (Long & Magerko 2020). This supports the notion that machine learning literacy is relatively low for people who do not work or have been educated in the field. An executive in a company without this foundational knowledge could have the potential to misconstrue the competence of machine learning hiring models which was seen in the case with Amazon. Higher-ups wished for the system to be absolute without the need for additional verification or screening once the model had produced its findings. The responsibility for the decisions made by the hiring model was pushed onto the model itself rather than the ones who developed it or advocated for its development and use (Dastin, 2018). This demonstrates a blind trust in the system rather than acknowledging the potentially fallible nature of machine learning models and accepting the moral agency that comes with employing these types of systems.

Even individuals who have been professionally educated in the field are not exempt from being unaware of the pitfalls of machine learning models. In a broad survey of the field of data science, 76% of respondents acknowledged the importance of including ethics in the education of the field (Saltz et al., 2019). However, a survey of the top fifteen data science programs in the U.S. found that only very few courses integrated ethics into their curriculum (Saltz et al., 2019). Without an additional focus on ethics during their education, it may be more difficult for students to apply their theoretical knowledge to complex real-world scenarios. One such scenario where missing this ethical background would be detrimental is in the case of filling in missing values when record bias error occurs. In the machine learning field, it is generally agreed upon that the most efficient and easiest way to solve this issue is to use a metric based on other existing data in the dataset such as the average or median (Géron, 2019). While this might be appropriate for numerical values in scientific research, it would not be for datasets used to train machine learning hiring models. Filling in missing data for minority groups using the proposed methods would not only misrepresent them but also reinforce structural bias. By following the general scientific knowledge without considering the ethical ramifications, the developer would fail to create a proper artificial moral agent.

As algorithmic fairness is still a newly developing field with uncertain claims of efficacy, even companies aware of ML's potential tendency to produce bias are not willing to shoulder the economic burden that proper validation entails. As publicly available relevant datasets are difficult to come by most of the methods found to reduce bias have to be supplemented with hypothetical synthetic datasets rather than ones that use real-world data (Mehrabi et al., 2021). Without being tested on more relevant datasets, there is no guarantee that these methods are as effective for reducing bias within machine learning hiring models with more research needing to be done. In addition, these methods can be economically costly and not always generalizable depending on the scenario and algorithm that the model utilizes (Sengupta et al., 2018). As introducing validation for machine learning hiring models remains an additional economic burden, the remaining primary motivation to prevent bias from forming would be a legal one for those bereft of moral agency that comes when employing these types of systems.

There is a significant lack of legal incentive or precedent for US companies to prevent bias from forming within their machine learning hiring models. As it currently stands, there are no federal laws that directly address the use of AI and machine learning related systems in employment hiring practices (Rogers and Reed, 2021). Though some states and municipalities have worked on pushing through anti-discriminatory AI hiring laws they unfortunately only apply to their respective regions and not all of them have been passed (Rogers and Reed, 2021). New York City's Local Law 144 of 2021 demonstrates an example of passable and thorough municipal legislation that could potentially inspire future federal legislation (NYC Consumer and Worker Protection, 2023). However, as this law was only enacted in 2023, more time must pass before an analysis can be done on its overall benefits. Without a set of tailored legal guidelines to "protect civil rights" (Allen et. al., 2006), companies are free to develop, train, and test their models by any means desired, some of which may have a higher tendency to form bias.

Some companies may argue that existing laws such as Title VII of the Civil Rights Act are already enough to incentivize preventing hiring models from discriminating, however, I argue that instead more specially tailored AI laws should be passed on a federal level (Kachra et al., 2023). Despite machine learning having been used to automate processes in employment for many years, the EEOC's lawsuit against iTutorGroup is the only settlement of its kind in which a company has been successfully sued for its use of AI in the hiring process (Coyle, 2023). While a lawsuit has recently been filed against the company Workday for a similar case of discrimination, it is uncertain whether it will be successful (Kachra et al., 2023). One study found that out of 1,672 employment discrimination cases, only 58% of them were settled which is below the general 70-80% range that other fields such as non-civil rights and tort cases have. Out of the 100 cases that were sent to trial only 32 plaintiffs were able to win or 2% of the entire whole (Eisenberg, 2015). These statistics demonstrate the difficulty of both settling and proving employment discrimination in the scenario that a case is taken to court. I believe that there should be more laws passed on a federal level ensuring some methods of quality control must be taken before such systems can be used as part of the employment automation process such as in the case of New York City's local law (NYC Consumer and Worker Protection, 2023).

EEOC vs. iTutorGroup

The lawsuit against iTutorGroup demonstrates a real-world example of the need for laws to ensure the validation of machine learning hiring models. iTutorGroup is a Chinese company that provides English tutoring services to students living in China and has hired thousands of employees from the U.S. to provide virtual online tutoring services. In 2020, Wendy Pincus submitted an application to the company; however, it was immediately rejected without explanation. Curious about this, the next day she resubmitted her application with the only change being a more recent date of birth and was offered an interview for the role (Equal Employment Opportunity Commission v. iTutorGroup, Inc., 2022a). This demonstrates an inherent difficulty when it comes to identifying discrimination within machine learning hiring models for the average individual. Due to the nature of machine learning systems, auditing proprietary information is generally difficult especially if it infringes on a user's or company's privacy (Raghavan et al., 2020). Without access to training data or the internal workings of the model, only the outputs of the model can be interpreted, which gives little insight into the decision-making aspect of the model. Most applicants would not attempt submitting multiple applications and more complex models might discriminate due to a combination of features rather than just one.

iTutorGroup's discriminatory model shows how minority groups are adversely affected by biased decisions. After additional testing, it was revealed that the automated application software was programmed "to automatically reject female applicants age 55 or older and male applicants age 60 or older." (Equal Employment Opportunity Commission v. iTutorGroup, Inc., 2022a) which affected over two hundred other applicants as well. This was a clear-cut case of discrimination based on age which violated the Age Discrimination in Employment Act (Coyle, 2023). As such, the affected parties of the case sought recompense for lost wages and hiring opportunities. Biased and discriminatory systems can act as barriers for marginalized groups struggling to work to make ends meet. This not only affected applicants and workers but also Chinese students who “lost the opportunity to learn English from highly qualified and experienced tutors” (Equal Employment Opportunity Commission, 2023a).

iTutorGroup’s response to these allegations demonstrates that companies will often deny claims of discrimination and refuse to take responsibility for failing to properly validate their systems. Even after the verdict, the company continues to deny “the EEOC’s allegations in their entirety and asserted numerous affirmative defenses.” (Equal Employment Opportunity Commission v. iTutorGroup, Inc., 2022b), claiming no wrongdoing. Even if the discrimination was unintentional, the company should ultimately be responsible for the decisions its hiring model makes. Instead of taking responsibility for this discrimination with the promise to improve or get rid of such biased systems, the company has chosen to halt all operations in the U.S. greatly harming those who were previously employed by them (Equal Employment Opportunity Commission, 2023a).

Conclusion

While there may not be one clear-cut all-encompassing solution to prevent biased hiring systems from being formed and used in the workplace, certain steps can be taken to help mitigate this ever-growing issue. Through education on both the technical and ethical side of machine learning hiring algorithms, it is hopeful that both executives and developers will develop a stronger sense of moral agency to ensure that these models do not become discriminatory. As the field of algorithmic fairness continues to develop, it is hopeful that further specialized studies regarding reducing algorithmic hiring bias will be conducted and produced. While ideally, one would hope that the moral arguments for reducing bias would convince companies, as demonstrated in my analysis, they are often more concerned about the financial and legal aspects of such systems. As such, the drafting and passing of stricter anti-discriminatory AI hiring laws would incentivize the development, validation, and proper testing of artificial moral agents.

As artificial intelligence and machine learning hiring models start to become more commonplace in the workforce, I hope that companies understand its shortcomings so that the hiring process can remain fair and impartial to all individuals. I can understand and acknowledge the perceived benefits of such systems in terms of economic value and importance, but this should not take precedence over their moral responsibility to those searching for work.

References

Allen, C., Wallach, W., & Smit, I. (2006). Why machine ethics?. *IEEE Intelligent Systems*, *21*(4), 12-17. https://doi.org/10.1109/MIS.2006.83

Costa, A., Cheung, C. Langenkamp, M. (2019). Technical Background. *Hiring Fairly in the Age of Algorithms*, 9-18. http://dx.doi.org/10.2139/ssrn.3723046

Coyle, B. M. (2023). Code and Prejudice: Regulating Discriminatory Algorithms. *Washington and Lee Law Review Online*, *81*(1), 57-60. https://scholarlycommons.law.wlu.edu/wlulr-online/vol81/iss1/1/

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

Eisenberg, T. (2015). Four Decades of Federal Civil Rights Litigation. Journal of Empirical Legal Studies, 12(1), 4-28. https://heinonline.org/HOL/P?h=hein.journals/emplest12&i=8

Equal Employment Opportunity Commission. (2022). EEOC Sues iTutorGroup for Age Discrimination. https://www.eeoc.gov/newsroom/eeoc-sues-itutorgroup-age-discrimination

Equal Employment Opportunity Commission. (2023a). iTutorGroup to Pay $365,000 to Settle EEOC Discriminatory Hiring Suit. https://www.eeoc.gov/newsroom/itutorgroup-pay-365000-settle-eeoc-discriminatory-hiring-suit

Equal Employment Opportunity Commission. (2023b). Select Issues: Assessing Adverse Impact in Software, Algorithms, and Artificial Intelligence Used in Employment Selection Procedures Under Title VII of the Civil Rights Act of 1964. https://www.eeoc.gov/laws/guidance/select-issues-assessing-adverse-impact-software-algorithms-and-artificial

Equal Employment Opportunity Commission v. iTutorGroup, Inc., 1 U.S. 1-8 (2022a). https://storage.courtlistener.com/recap/gov.uscourts.nyed.479565/gov.uscourts.nyed.479565.1.0.pdf

Equal Employment Opportunity Commission v. iTutorGroup, Inc., 24 U.S. 119-142., (2022b). https://www.bloomberglaw.com/public/desktop/document/EqualEmploymentOpportunityCommissionviTutorGroupIncetalDocketNo12/1?doc_id=X4663TFVFDF9ONAA8CMUJFPH3HR

Géron, A. (2019). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. (2nd ed.). O'Reilly Media, Inc.

Kachra A., Hilliard A., Gulley A., Wilson I., (2023). Lawsuits in the United States point to a need for AI risk management systems. OECD. https://oecd.ai/en/wonk/lawsuits-usa-risk-management

Kong, D., (2023). Over half of employees have no idea how their companies are using AI. CNBC. https://www.cnbc.com/2023/10/24/over-half-of-employees-have-no-idea-how-their-companies-use-ai.html

Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and Design Considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1-16. https://doi.org/10.1145/3313831.3376727

Mahesh, B. (2020). Machine learning algorithms-a review. International Journal of Science and Research (IJSR), 9(1), 381-386. https://www.ijsr.net/archive/v9i1/ART20203995.pdf

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A Survey on Bias and Fairness in Machine Learning. *ACM Computing Surveys (CSUR)*, *54*(6), 1-35. https://doi.org/10.1145/3457607

NYC Consumer and Worker Protection (2023). Automated Employment Decision Tools: Frequently Asked Questions. https://www.nyc.gov/assets/dca/downloads/pdf/about/DCWP-AEDT-FAQ.pdf

Raghavan, M., Barocas, S., Kleinberg, J., & Levy, K. (2020). Mitigating bias in algorithmic hiring: Evaluating claims and practices. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*, 469-481. https://doi.org/10.1145/3351095.3372828

Rogers, A. M., & Reed, M. (2021). Discrimination in the Age of Artificial Intelligence. *American Bar Association*. https://www.americanbar.org/groups/labor_law/publications/labor_employment_law_news/summer-2021-issue/discrimination-age-of-ai/

Saltz, J., Skirpan, M., Fiesler, C., Gorelick, M., Yeh, T., Heckman, R., D., Neil & Beard, N. (2019). Integrating ethics within machine learning courses. *ACM Transactions on Computing Education (TOCE), 19*(4), 1-26. https://doi.org/10.1145/3341164

Schumann, C., Foster, J. S., Mattei, N., & Dickerson, J. P. (2020). We Need Fairness and Explainability in Algorithmic Hiring. Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS), 1716–1720. https://par.nsf.gov/servlets/purl/10214110

Sengupta, E., Garg, D., Choudhury, T., & Aggarwal, A. (2018). Techniques to eliminate human bias in machine learning. In *2018 International Conference on System Modeling & Advancement in Research Trends (SMART)*, 226-230. https://doi.org/10.1109/SYSMART.2018.8746946

Tolmeijer, S., Kneer, M., Sarasua, C., Christen, M., & Bernstein, A. (2020). Implementations in Machine Ethics: A Survey. ACM Computing Surveys (CSUR), 53(6), 1-38. https://doi.org/10.1145/3419633