THE PRESENCE OF UNINTENDED BIAS IN ARTIFICIAL LEARNING USED IN THE HIRING PROCESS

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

By

Nicholas Winans

March 28, 2022

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
Catherine D. Baritaud, Department of Engineering and Society

Artificial Intelligence and Machine Learning are among the most promising fields of research within computer science when measuring the capability to imitate a human's intellectual ability (IBM, 2020). Machine learning is the process of computers learning patterns in data and techniques are often modeled after how humans process the phenomena around them (McCulloch & Pitts, 1943).

With the help of machine learning, novel computing paradigms are being applied to many areas of computer science and real world scenarios. The introduction of machine learning into the field of cryptography and encryption is covered in my technical report. Under the guidance of Daniel G. Graham, a survey of how researchers are applying machine learning models to create encryption schemes is detailed. However, new ideas can introduce new problems, as well as reinforce existing problems. With the rapid utilization of machine learning, it is paramount to understand biases that we are introducing and how to account for those biases. The tightly coupled STS report focuses on biases, how they enter models and how to account for them. Specifically, this STS report focuses on the introduction of machine learning into the hiring process and the biases that have been discovered. It will do this by analyzing a case study of a single company’s foray into using machine learning in the candidate screening process. The report will first introduce the case itself, introducing what happened in the process of building a machine learning model. Then it will introduce the complications to the problem and why a solution is not readily available. Finally, the report will focus on the parties involved in the overall hiring process, their role and how bias can enter the system through shortcomings.

This topic is especially interesting as fourth-year students, who have recently entered the job market, will likely have to deal with many different machine learning models across all the applications submitted. Having to submit upwards of hundreds of applications, and barely

hearing anything back can be very discouraging. It often feels like there is not anyone actually looking at applications, instead just choosing to ignore it. While data isn't available for all companies hiring candidates, a glimpse into one large company can provide context for how others for the industry operate. Holding companies accountable for the biases they have will create a more equitable workforce.

MACHINE LEARNING IN THE HIRING PROCESS

Machine learning models are being used by companies in the talent sourcing, candidate engagement and prospective employee assessment and selection stages of the hiring process. (Bayern, 2020). This means that various companies are using machine learning models to find candidates to reach out to, communicate with candidates, and even narrow down candidates through screening processes (Bayern, 2020). A 2021 report showed that an estimated 75% of companies used automated systems in the hiring process, with that number jumping up to 99% for Fortune 500 companies (Fuller, Raman, Sage-Gavin, & Hines, 2021).

BIAS IN THE HIRING PROCESS

Many of the models that companies are using in the hiring process have been shown to be biased towards certain groups (Wiggers, 2020). For instance, take Amazon's foray into machine learning in the hiring process(Dastin, 2018). They started working on machine learning models back in 2014 for use in the candidate discovery process as part of an exploratory process, but realized a year later that their model showed gender-bias (Dastin, 2018). The problem with their

models, and many others, is that the data source to train the model was current employees (Bogen, 2019; Dastin, 2018). The tech industry has been predominantly male for the entire course of their training dataset, so the model learned to discriminate against candidates who included the name "women's" in their resume or went to all-women colleges (Dastin, 2018). In fact, the breakdown of Amazon's employee gender breakdown is shown below in Figure 1, highlighting just how large the gap is in technical roles.

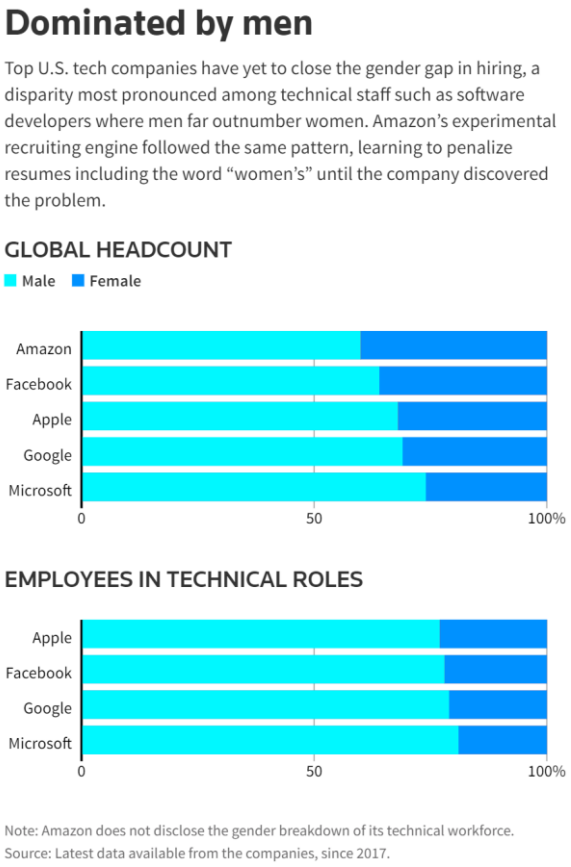

Figure 1: Dominated by men. A gender breakdown of top tech companies (Iriondo, 2018).

Ultimately, Amazon scrapped the project after attempting to remedy the bias, finding that there was no guarantee that the machine wouldn't produce other biased features in the future (Dastin,

2018). The incident highlights just how difficult it is to maintain a bias-free environment. The data points used to discriminate against candidates were their colleges and extracurriculars, not even their gender. Confounding variables can show up in many places in datasets so it is increasingly important for model creators to examine every variable for confounding effects, and work out how to cancel out those effects.

CONFOUNDING VARIABLES AND PROBLEMS

Companies can receive hundreds of thousands of applications for a single position (Bayern, 2020), so many companies turn to machine learning algorithms in some capacity (Fuller, Raman, Sage-Gavin, & Hines, 2021). There are already regulations in the space, as the Civil Rights Act of 1991 prohibits a preference for any group in the hiring process, minority or majority (Raub, 2018). As a result, Bogen reported in 2019 how many third-party companies who offer aptitude and personality assessments (machine learning models) for other companies highlight the steps they take to de-bias their models to ensure that they are within the guidelines of the law. Yet even this is not perfect, as remaining variables in datasets can serve as proxy variables to discriminate against. For example, consider Geronimus & Bound's research into how census-based data like Zip Codes can be a reasonable proxy variable for the socioeconomic status of an individual (1998). Even if discriminating variables are removed, companies can get a rough overview of candidates through proxy variables.

Many hiring algorithms attempt to find candidates similar to the high performers within the company, as seen with Amazon, but the process of determining high performers itself can be a deeply flawed and biased process (Bogen, 2019). While there is some regulation that covers the

field, the aforementioned Civil Rights Act of 1991 was written before the era of machine learning and is outdated. Indeed, further legislation also lacks at the federal level, with some individual states implementing their own legislation (Simonite, 2021). When companies and governments try to apply existing regulations to an evolved hiring process, cracks appear. The government did not foresee how much the internet would change the job application process, with the exponential increase in applications and resultant methods to deal with the volume.

HOW CAN COMPANIES STILL USE MACHINE LEARNING IN A LEGAL MANNER?

How can companies sift through hundreds of thousands of applications in a legal, bias-free manner when it seems that every step of the process has the opportunity to introduce bias? This paper highlights why previous implementations of machine learning models have failed by introducing bias, with the end goal of providing steps companies can take to maintain compliance with the law while using machine learning. Since it is clear that government regulations are outdated and companies have no initiative beyond the bare minimum, it is important to bring the bar to them (Bogen, 2019). By laying out where a company can inadvertently introduce bias in their models, the intention is to lay out a framework of what not to do when building a bias-free model.

Pymetrics gives job candidates mini games that attempt to measure qualities like "generosity, fairness, and attention" (Schellmann, 2021). Hiring companies can then match up candidate qualities with existing employee qualities. Importantly, Pymetrics paid Northwestern University to complete a third party audit of their algorithms to confirm their claims of no bias (Schellmann, 2021). Schellmann reports that the government definition of bias, known as the

four-fifths rule, involves making sure that if 100% of men are passing on to the next round, at least 80% of women need to also pass on (2021). This is a very binary solution to a nuanced problem, as it is only applied to broad categories like men vs women and black vs white, but not white men vs black women (Schellmann, 2021). This does not provide the level of granularity that deep issues like biased hiring patterns present. While Pymetrics ended up passing the four-fifths tests in its algorithms, per Schellmann, questions remain about the fairness of the algorithms (2021).

WHAT CAN BE LEARNED FROM PREVIOUS CASE STUDIES

No technological innovation is inherently positive or negative, it is how it is used that defines these connotations. Machine Learning algorithms are no different, and most of the negative press they get is because of biases that humans introduce (Silberg & Manyika, 2019). Often we train machine learning models to make decisions for us, like which candidates are right to hire for our company (Friedman & McCarthy, 2020). Silberg & Manyika showed how machine learning models are very sensitive to biases since humans have control over their design, data sets and implementations (2019). When we model datasets after the current status quo, it is destined to reinforce the status quo, and all the biases currently present (Silberg & Manyika, 2019). We saw this with Amazon's failed foray into the field, and there are no doubt countless other examples that did not get as much attention.

AN ACTOR NETWORK VIEW OF MACHINE LEARNING BIAS

Actor Network Theory is a useful tool to analyze the social relationships between both living and inanimate objects (Law & Callon, 1988). It helps in laying out the complex relationships between agents within a network, as well as where power is derived from within a network. By laying out the complex relationships, we can begin to analyze where and how biases are introduced, and move towards a more bias-free environment. In the Amazon case study, we see groups emerge, such as the company itself, the model, job candidates, government regulators and more. Each can have their own agenda, which quickly complicates their interactions and as a result, any diagram depicting the situation. Actor Network is a powerful framework to lay these interactions out in a digestible manner.

First, the machine learning models themselves have agency in how they predict outputs (whether to hire someone) given inputs (a person's qualifications and resume). The company that uses a model in the hiring process (Amazon) also has agency in how extensively the model is used. It is this interaction between the model and the company that gives the model power, as without a use case, a model has no power to make decisions.

The government and regulatory agencies have power as well. The Civil Rights Act of 1964, for example, prevents employers from discriminating based on "race, color, religion, sex or national origin" (Civil Rights Act (1964)). This includes company hiring processes and does not have any contingencies for machine learning. Thus, all models used in the hiring process to limit candidate pools or screen candidates must not discriminate against candidates for any of those reasons. Thus, the government exercises power over the companies implementing models and over the models themselves by limiting the results of a model.

Similarly, independent model auditors hold power over the use of models as well. They can work with government agencies to enforce regulations, but can also introduce their own standards that models have to follow. Independent model auditors have the power to tell companies that their model displays bias, but cannot restrict the use of such models or impose fines like the governments can. Yet independent auditors have other avenues, like appealing to the public, letting them know about a company's poor attention to biases within a model. Independent auditors exercise power over models due to the interactions with job candidates themselves, government agencies and companies seeking employees.

The subjects that comprise datasets the model is trained with also plays an important role. In the case of Amazon, the employees themselves were the dataset, as the company wanted to find employees similar to the ones they already hired. In this case, the employees don't exercise their own agency. Instead they give agency to the model, as well as the company who uses the model, allowing them to find additional employees like themselves.

Organizations that actually collect data also exercise agency over the process. The data collection organizations are responsible for choosing how to collect data, as well as ensuring that collection processes are followed. There are many areas where bias can enter datasets in this stage, like sampling non-representative groups of a population. The data collection organizations have complete control over this process, as well as communicating to any consumers of the data about the collection process. In the case of Amazon, they were the data collection agency as well as the consumer of the data. Yet, a clear communication breakdown happened when they trained the model on non-diverse employee data, and expected the model to output diverse results.

Next, job candidates themselves are also an important actor within the network. They do not have any agency over the model's results, but are most directly affected by the output of the

model. The company will eventually find a suitable candidate for an open position using the model, but the job candidate has their livelihood on the line, and thus have much more at stake when the model chooses an outcome. By extension of the model, the job candidate also has notable relationships with government regulators and the hiring company. As mentioned above, the government ensures that companies follow certain rules when seeking job candidates, as outlined in the Civil Rights Act of 1964.

The complex relationships between all these actors is highlighted in Figure 2 below. The web is very interconnected due to all the direct and indirect methods through which effects of an agent's actions are felt.

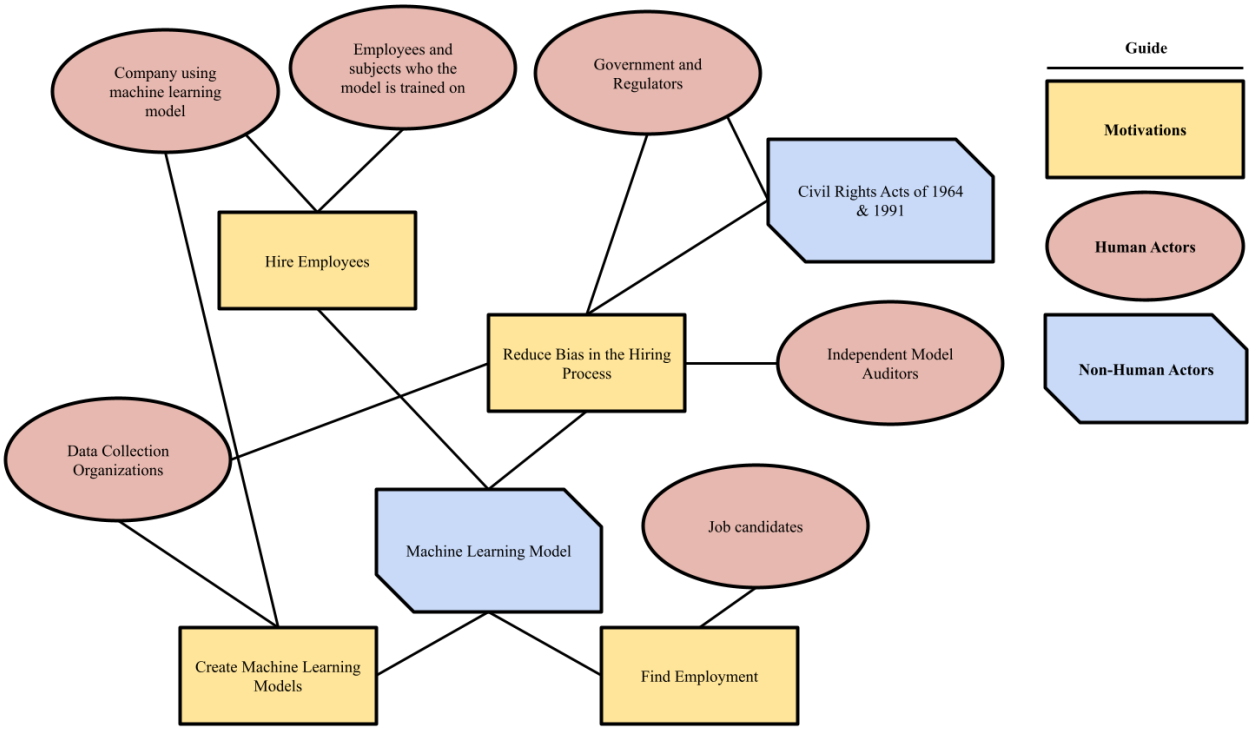

Figure 2: Machine learning in hiring actor network map. Depicts the different groups that interact with each other during model creation (Winans, 2022).

It is important to mention that while the model that Amazon used tended to bias against women, it was not doing this based on a datapoint saying that the candidate was a woman. As

mentioned previously, proxy variables can relay information that it was not meant to. Amazon's model was using colleges and extracurricular activities as proxies for gender due to the women-only nature of certain colleges. While this model was never actually used to filter candidates, it highlights how important it is for companies to fully understand the data they collect and extensively test for biases that may be present in the form of proxy and confounding variables.

Finally, we can also look at what the actors mean to each other within the network. To job applicants, machine learning models may represent an opaque step in the job application process - an automated one without humans to provide feedback. To companies seeking employees, it represents a solution to the problem of exponentially increasing numbers of applications to open positions. To model creators, it is a product they are selling; de-biasing a model could mean increased profit and more customers. To government regulators, models are an emerging technology that presents a tradeoff between regulation to protect constituents and a more laissez-faire attitude to promote innovation in companies. Machine learning models mean many different things to different actors, and only taking one into consideration leaves an incomplete picture.

AN ETHICAL THEORY PERSPECTIVE

In addition to an Actor Network perspective, we can analyze the case study of Amazon's machine learning experiment under the lens of ethical theories. Under utilitarianism, we can highlight the difference between expectations and reality with machine learning models. Ideally, models in the job seeking market should allow candidates and companies to find the right match,

as well as reducing bias in the process. However, in actuality, the models reinforce biases already present and remove trust and clarity.

By reinforcing biases, the machine learning models are violating the right of job seekers to seek employment and career opportunities. It would be morally permissible for companies to deny candidates jobs on the basis of qualifications, but external factors such as gender are not permissible as outlined in the Civil Rights Acts of 1964 and 1991. Similarly, biased candidate screening machine learning models prevent candidates from self-realization. Candidates all have different goals, and by broadly denying certain groups, the machine learning model is denying the self-realization that some may get from receiving this job offer.

Interestingly, when the model produces biased results, it also inhibits the company using the model from self-realization. If Amazon as a whole is viewed as an individual entity, it can be said that Amazon's goal is profits for its shareholders. When the machine learning model prevents the company from finding the best candidates, it prevents the company from self-realization. Flawed machine learning models inhibit all parties involved across multiple ethical theories.

MOVING FORWARD WITH MACHINE LEARNING IN HIRING

The accessibility of online job applications has increased the number of applications per open position dramatically. (Bayern, 2020). Machine learning is one of the most promising and realistic methods to deal with this problem. In order to fulfill its promises, however, we need to learn from previous transgressions and work towards a more fair and bias-free future.

Future work can investigate other fields with machine learning models for bias; the job finding industry is not alone in its use of machine learning. Indeed, not even Amazon is alone within the hiring industry, as companies like Pymetrics also insert machine learning into the hiring process (Schellmann, 2021). While this industry is a good place to start, it is important to keep a broad view of machine learning and reduce bias everywhere.

There is also an important distinction between meeting federal and state regulatory requirements and actually striving to eliminate bias from models. Federal regulations are lagging behind technological developments, leaving a gray area between legal and moral. Current regulations are failing both candidates and employers alike and research into appropriate regulations that still promotes innovation is well needed.

Further research should investigate other egregious implementations of machine learning across the plethora of applications. We should additionally investigate systems that seem to be fair and unbiased, like Pymetrics. This yields two main goals: (1) making sure that these systems truly do not bias against groups, and (2) learning from what these machine learning models do right so we can apply the same methods to new models.

The technological boom that enables machine learning models to become useful also necessitates an ethical boom. The consequences and realities of machine learning must be considered rather than simply the idealistic possibilities.

REFERENCES

Bayern, M. (2020, January 4). How artificial intelligence and machine learning are used in hiring and recruiting. *ZDNet.* Retrieved from https://www.zdnet.com/article/how-artificial-intelligence-and-machine-learning-are-used-in-hiring-and-recruiting/

Bogen, M. (2019, May 6). All the ways hiring algorithms can introduce bias. *Harvard Business Review*. Retrieved from https://hbr.org/2019/05/all-the-ways-hiring-algorithms-can-introduce-bias

Civil Rights Act of 1964 (1964). §7, 42 U.S.C. §2000e et seq.

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

Friedman, G., & McCarthy, T. (2020, October 1). Employment law red flags in the use of artificial intelligence in hiring. *American Bar Association.* Retrieved from https://www.americanbar.org/groups/business_law/publications/blt/2020/10/ai-in-hiring/

Fuller, J. B., Raman, M., Sage-Gavin, E., & Hines, K. (2021, September). Hidden workers: Untapped talent. *Harvard Business School*.

Geronimus, A. T., & Bound, J. (1998). Use of census-based aggregate variables to proxy for socioeconomic group: Evidence from national samples. *American Journal of Epidemiology*, *148*(5), 475–486. https://doi.org/10.1093/oxfordjournals.aje.a009673

IBM (2020) Machine Learning. *IBM Cloud Education*. Retrieved from https://www.ibm.com/cloud/learn/machine-learning

Iriondo, R. (2018, October 11). Dominated by men [Figure 1]. *Carnegie Mellon University School of Computer Science.* Retrieved from https://www.ml.cmu.edu/news/news-archive/2016-2020/2018/october/amazon-scraps-secret-artificial-intelligence-recruiting-engine-that-showed-biases-against-women.html

Law, J., & Callon, M. (1988). Engineering and sociology in a military aircraft project: A network analysis of technological change. *Social problems*, *35*(3), 284-297. https://doi.org/10.2307/800623

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics, 5(4),* 115-133.

Raub, M. (2018). Bots, bias and big data: Artificial intelligence, algorithmic bias and disparate impact liability in hiring practices comment. *Arkansas Law Review*, *71*(2), 529–570.

Schellmann, H. (2021, February 11). Auditors are testing hiring algorithms for bias, but there's no easy fix. *MIT Technology Review*. Retrieved from https://www.technologyreview.com/2021/02/11/1017955/auditors-testing-ai-hiring-algorithms-bias-big-questions-remain/.

Silberg, J., & Manyika, J. (2019, June 6). Tackling bias in artificial intelligence (and in humans). *McKinsey Global Institute*. Retrieved from https://www.mckinsey.com/featured-insights/artificial-intelligence/tackling-bias-in-artificial-intelligence-and-in-humans

Simonite, T. (2021, January 8). New York city proposes regulating algorithms used in hiring. *Wired*. Retrieved from https://www.wired.com/story/new-york-city-proposes-regulating-algorithms-hiring/

Wiggers, K. (2020, December 4). Researchers find that even 'fair' hiring algorithms can be biased. *VentureBeat*. Retrieved from: https://venturebeat.com/2020/12/04/researchers-find-that-even-fair-hiring-algorithms-can-be-biased/

Winans, N. (2022). Machine learning in hiring actor network map [Figure 2]. *STS Report* (Unpublished undergraduate thesis). School of Engineering and Applied Science, University of Virginia. Charlottesville, VA.