

Why do Artificial Intelligence based hiring systems, which include both the technology and the people that interact with it, result in hiring that might seem inequitable at times?

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

Pragnya Pandrate
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
S. Travis Elliott, Department of Engineering and Society

INTRODUCTION

In the past decade, Artificial Intelligence (AI) has been extensively used by employers to make hiring decisions. In fact, Amazon was one of the biggest organizations that utilized Artificial Intelligence to evaluate candidates in the hiring process although it resulted in controversy because of the results. The training dataset that they used gave a lower rating to resumes with the word “women” in them compared to those without the keyword. After further analysis, they concluded that since the tech industry is often male dominated, the dataset they used was biased towards resumes without certain female related keywords, unfairly disadvantaging this demographic (Kodiyan, A. A. , 2019). While Amazon stopped using this tool to evaluate resumes, it brought up the important question of how AI should be used in the hiring process. Clearly, AI is helpful in evaluating large numbers of candidates that companies such as Amazon receive on a regular basis, however the biased hiring decisions it can lead to need to be mitigated in order to use these tools to hire candidates in an equitable manner (Ghosh, 2017).

The bias against underrepresented demographics in AI hiring systems can be examined and potentially mitigated by using the Actor Network Theory to examine the role of different actors in the AI hiring system. The Actor Network Theory focuses on the role of entities that are human or non-human (herein referred to as actors) and how they interact in a network such that they affect social processes. The ANT framework posits that the connections between these actors are what results in scientific and technological developments. Rather than simply mapping the interactions between the different actors in the network, this framework aims to “map in which they [actors] define and distribute roles, and mobilize or invent others to play these roles” (Law & Callon 1988, p.285).

The ANT framework can be applied to AI hiring systems by examining the roles of algorithms, datasets, candidates, and hiring managers and how they propagate bias in hiring decisions. By analyzing

the effect of these actors on the technological development of AI hiring systems, the source of the bias can be pinpointed and potentially mitigated.

A BRIEF BACKGROUND ON AI HIRING SYSTEMS

As companies grow to larger levels than ever before, it has become increasingly important to recruit employees in an effective manner. Before the rise of AI, employers solved this problem by hiring more recruiters, expanding their Human Resources department, and utilizing their networks. However, with AI this process has been largely simplified as there are many systems that are able to process and recommend candidates to recruiters. This allows companies to be more economically efficient in their hiring process while also allowing more candidates to apply to various job positions (Johansson and Herranen, 2019).

There are two main aspects to how an AI hiring system is built: the algorithm and the datasets. Developers often create algorithms based on artificial intelligence and natural language processing to evaluate resumes and virtual interviews. At its core these algorithms rely heavily on training datasets in order to train the models to make accurate predictions about hiring decisions. For example, the training datasets that AI hiring systems are trained on might include data on which resumes passed the company's resume screening in the previous year. The system then uses this data to analyze which features are most prominent in these "successful" resumes and uses this information in order to predict which resumes should pass the screening. The issue arises when these training datasets are biased towards a certain demographic, resulting in predictions and hiring decisions that are also biased towards that demographic (Nugent, S. E., & Scott-Parker, S. , 2021).

Broadly, AI hiring systems can be split into two categories: candidate sourcing software and candidate evaluation software. Some of the most popular candidate sourcing systems that use AI include LinkedIn and ZipRecruiter. While the algorithms that these systems use are not widely documented,

LinkedIn does divulge some information about the technology that recruiters often use to source potential candidates in the first step of the hiring process. LinkedIn uses a search engine with data such as a query, job posting, or an ideal candidate profile to construct criteria that will ideally predict a well-qualified candidate to the recruiter. Then machine learning models (a subset of Artificial Intelligence) are used to rank candidates based on the generated qualification criteria (Li et. al., 2021).

On the other hand, candidate evaluation software is used to analyze candidates' qualifications based on online assessment questions, video interviews, etc. These software often use natural language processing and sentiment analysis in order to analyze interviewees' answers and predict if they would be a good match based on the recruiters' criterion for a "good hire". The factors that the AI model might use to predict the success of a candidate can vary largely from keywords that the candidate uses to body language in videos.

While AI hiring systems have been in wide use in the past decade, hiring decisions are still not completely automated. Often, these systems are augmented by the opinions of a hiring manager. For example, many systems involve AI at the resume screening stage which analyzes the applicant and the final hiring decision is made by the hiring manager who takes into account the AI prediction and any other significant factors they choose to consider.

ACTOR NETWORK THEORY AND ITS IMPACT ON TECHNOLOGY

The Actor Network Theory aims to define the roles and effects of different actors in a network and how this can lead to the development of technology. One of the key points the ANT framework makes is that even inanimate objects can be actors in a network, therefore actors need not be limited to humans. This widens the definition of what an actor can be and therefore networks can often include many actors. This definition leads to the main point of the framework – that connections between these actors can affect technology and social processes.

This theory can be contrasted with Technological determinism which states that technology influences society. From this theory, it follows that society and culture are defined by technological innovation. The theory also states that technological innovation follows a predictable path. In the context of this paper, this would mean that the development of AI hiring systems in and of itself led to bias in the hiring systems. In contrast, the ANT proposes that the effects of the connections between actors in a network can lead to technological and social changes. Viewing AI hiring systems with the ANT framework means that the actors and the network in which the system exists would first need to be defined.

For the purposes of this paper, the actors in the AI hiring network include algorithms, datasets, candidates, and hiring managers. According to this theory, the roles that each of these actors play in the network and the connections between them influence technological innovations such as the development of AI hiring systems. First, hiring managers influence the development of the algorithm because they are the main users of this system. Therefore, the algorithm is created to fulfill their needs in finding suitable candidates. Whether the algorithm is strict in its criteria for employees or more forgiving is dependent on the hiring manager's preferences. The next actor to consider is the dataset, which influences the algorithm. The dataset (in reference to the training data used to train the AI model) has a large influence on the algorithm's efficacy and bias. For example, a non-random, biased sample of resumes could adversely affect the training process and therefore the accuracy of the algorithm in its ability to predict qualified candidates to hire. Lastly, the potential candidates have connections to both the algorithm and the hiring manager. The algorithm is what determines whether the candidates will move on to the next round in the hiring process and the hiring manager also has decision making power over the candidates.

Another important part of ANT is that it states that since technology is a product of these social interactions between actors, it has the potential to shape social processes. ANT explains the fact that

technology is constantly evolving by assuming that the actors in the network are constantly shifting. As the network changes, the social interactions are affected, therefore resulting in the evolution of technology. This proves that consistently re-evaluating technology through the lens of the ANT is critical to identifying how technology such as AI hiring systems evolve.

The network of AI hiring systems is subject to the this framework as the technology is a byproduct of the social interactions between the actors that were previously defined. It also follows that if the network changes, the system changes as a consequence. For example, consider the hiring manager and their interaction with the algorithm. The algorithm is dependent on the hiring manager's preference on what they think is an ideal candidate and differ in their flexibility when it comes to these preferences. While this can result in a positive impact on the hiring efficacy of candidates, a biased hiring manager could also adversely affect this process. Even a minute change in the preferences of a hiring manager due to an unconscious bias could drastically affect the algorithm and the potential candidates in the hiring process in either a positive or negative way. This example of the extensive effects of the social interactions between the actors in the network shows that even small changes in the network can affect the social processes and the technology in the network. Consequently, this proves that the network defined for the AI hiring systems is in line with the main conclusions and assumptions made by the ANT.

THE IMPACT OF HIRING MANAGERS ON AI HIRING SYSTEMS

The extent to which hiring managers are involved in AI hiring systems is highly contested. AI was largely introduced to make the process of hiring more efficient, however it has been criticized for being depersonalized and even biased in some cases. Therefore, hiring managers were included as another actor to this network. As mentioned previously, when the network changes, the impact on the social processes involved is profound – which is what will be discussed in this section.

While the original intent of introducing hiring managers into this system was to make the process more personalized, it has led to inadvertent effects such as introducing unconscious biases which has an effect on the entire network. Unconscious biases have its roots in “mental shortcuts”, also known as heuristics, that many humans form in order to fill in gaps in the information they have. These heuristics can often be based on previous experiences and norms. Hiring managers are subject to these unconscious biases, just as other humans are, however, their role as an actor in the hiring network makes these unconscious biases even more important to examine.

One of the common biases that exists in the tech industry is the status quo bias which exists when people prefer things to remain as they are. This can manifest itself in the form of hiring managers preferring not to hire women or people of color in order to maintain the homogeneity of the tech industry. Even if AI is used in a hiring system to recommend qualified candidates from a diverse background, hiring managers that hold unconscious biases could weed out the applicants in a later stage. In an attempt to reduce bias by adding an actor to the network, it has actually resulted in increased unconscious biases.

Another issue with introducing hiring managers in a network that already involves AI powered hiring systems is the inconsistency it introduces. For example, the decisions between different hiring managers might vary or they might not make the same decision they made for previous candidates. Not only can this result in financial losses for the company they are recruiting for, it can also result in inequitable results for the candidates involved in the network.

The issue of biases and inconsistent decision making are clearly a result of hiring managers involved in the AI hiring system. Although it might mitigate biases the AI has, it can also introduce new unconscious biases and lead to inconsistent hiring decisions. Therefore, many have proposed leaning more on the algorithms involved in the network and incorporating AI to help reduce these unconscious

biases. Using an algorithmic approach to hiring will also lead to more consistent hiring decisions for potential candidates involved in the network. In fact, a study that involved interviews of 15 recruiters and HR professionals showed that hiring managers themselves felt that by including AI hiring systems they “saw a lot more diversity when it comes to what schools, work, where these participants or candidates were coming from.” (Li et. al, 2021). This shows that reducing the role of hiring managers in the network might actually help mitigate unconscious biases and make more informed and equitable hiring decisions.

THE IMPACT OF DATASETS ON AI HIRING SYSTEMS

Although reducing the role of hiring managers could lead to more equitable hiring decisions, there is no guarantee the AI system itself will make perfect hiring decisions. The algorithm that is developed is entirely dependent on the quality of the dataset and how the AI system is trained. A recent paper on data mining and its effects explains how data can be the core issue behind many AI systems.

In order to train the system properly there are two parts of data mining that need to be defined in the context of the problem: target variables and class labels. The target variable can be defined as the ideal result that the algorithm should identify. In this case, the target variable is be the ideal candidate that the hiring manager would hire. Class labels can be defined as different categories that are used to categorize all the possible values of the target variable. In this case, examples of class labels include job experience, interview skills, quality of resume, etc.

To be able to define a target variable properly, the problem must be specified. The problem at hand is hiring the ideal candidate without compromising equitable hiring standards. While it is easy to specify what the ideal candidate might be, different hiring managers might have different perspectives on what the standards are for equitable hiring. This is where hiring managers’ own biases might start to affect the problem specification. Consequently, if the problem is not specified correctly it can lead to the

incorrect target variable and therefore affect the algorithm's ability to identify good candidates equitably.

Similarly, choosing the proper class labels is also a vital part of data mining that could introduce unintended biases. For example, if predicted tenure is included as a class label in making hiring decisions then it could bias decisions in a way that would give certain socioeconomic classes an advantage over others (Barocas and Selbst, 2016).

After data mining and cleaning, the AI system must be trained with a training data set. The general machine learning problem is as follows: each training data point is labeled as a qualified candidate or not and then the algorithm must learn the similarities/differences between the qualified versus not qualified candidates. Learning these similarities and differences can help predict the potential success of future candidates. Although this algorithm is straightforward, it could lead to biased results in one of two ways: the training data might be biased due to data collection or the classifications of what is a qualified candidate could be incorrect.

Training data could be biased due to a variety of reasons including incorrect, non-random, or non-representative data collection. For example, if a hiring algorithm is trained with the data of resumes from only the highest ranked universities in America, then this could result in an unfair advantage to certain candidates. While going to a high ranked university can certainly mean that a candidate is more academically inclined and likely to succeed at the job, this is not necessarily true because college admissions are dependent on a variety of factors including socioeconomic factors. Therefore, what might seem like a representative data pool might not be if careful discretion is not used during the data collection process.

HOW THESE IMPACTS CAN BE MITIGATED

By framing AI hiring systems using ANT, it becomes clear that the actors that have the most impact are the hiring managers and data. When it comes to hiring managers, evidence shows that they actually add unconscious bias. Rather than introducing them as one of the main actors in network it might be beneficial to reduce their role. Some have argued that hiring managers should take part in increased training to recognize and mitigate their own unconscious biases, however this has proved to be an unreliable strategy (Houser, 2019). Rather than relying on the inconsistent decision making from hiring managers, increasing the AI algorithm's role in the system could result in equitable hiring decisions.

As mentioned before, AI hiring algorithms' equitability is highly dependent on data mining/cleaning and data collection. In order for both these processes to happen without unintended biases it is important to focus on the problem specification before implementing the algorithm. The product management/planning phase must be heavily involved before an algorithm is set in order to scope the problem of hiring in the context of different employers. Additionally, the algorithm must be constantly evaluated based on results no matter how carefully the data is examined because unintended biases can be very minute yet have a large effect.

CONCLUSION

The Actor Network Theory is beneficial in determining the effects of different actors in the AI hiring system. By including inanimate actors, such as data, their interactions with other actors such as the algorithms, candidates, and hiring managers can be considered. Evaluating the different actors and their relationships in the network showed that hiring managers and data were the most influential actors in the network. Therefore, efforts to mitigate bias should focus on examining the effects of these specific actors. While these are the current most pressing actors in the network, it is important to continue to use

ANT to examine AI hiring systems as actors are introduced and removed from the network to maintain equity and consistency in hiring decisions.

REFERENCES

Cresswell, K.M., Worth, A. & Sheikh, A. Actor-Network Theory and its role in understanding the implementation of information technology developments in healthcare. *BMC Med Inform Decis Mak* 10, 67 (2010). <https://doi.org/10.1186/1472-6947-10-67>

Cressman, D. A Brief Overview of Actor-Network Theory: Punctualization, Heterogeneous Engineering & Translation. *Simon Fraser University*. (2009). <https://summit.sfu.ca/item/13593>

Johansson J. and Herranen, S. The application of Artificial Intelligence (AI) in Human Resource Management: Current state of AI and its impact on the traditional recruitment process. *DiVA*. (2019). <https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1322478&dswid=-2370>

Manuel F. Gonzalez, Weiwei Liu, Lei Shirase, David L. Tomczak, Carmen E. Lobbe, Richard Justenhoven, Nicholas R. Martin, Allying with AI? Reactions toward human-based, AI/ML-based, and augmented hiring processes, *Computers in Human Behavior*, Volume 130, 2022, 107179, ISSN 0747-5632, <https://doi.org/10.1016/j.chb.2022.107179>.

Li, Lan and Lassiter, Tina and Oh, Joohee and Lee, Min Kyung. Algorithmic Hiring in Practice: Recruiter and HR Professional's Perspectives on AI Use in Hiring. 2021. Association for Computing Machinery. <https://doi.org/10.1145/3461702.3462531>

Barocas, Solon & Selbst, Andrew. Big Data's Disparate Impact. *California Law Review*. <http://dx.doi.org/10.15779/Z38BG31>

Black, J. S., & van Esch, P. (2020). AI-enabled recruiting: What is it and how should a manager use it?. *Business Horizons*, 63(2), 215-226.

Ghosh, D. (2017). AI is the future of hiring, but it's far from immune to bias. *Quartz at Work*, 17.

Houser, K. A. (2019). Can AI solve the diversity problem in the tech industry: Mitigating noise and bias in employment decision-making. *Stan. Tech. L. Rev.*, 22, 290.

Kodiyan, A. A. (2019). An overview of ethical issues in using AI systems in hiring with a case study of Amazon's AI based hiring tool. Researchgate Preprint.

Nugent, S. E., & Scott-Parker, S. (2021, September 25). Recruitment AI has a Disability Problem: anticipating and mitigating unfair automated hiring decisions.

<https://doi.org/10.31235/osf.io/8sxh7>