

The Ethical Fallout of Amazon's AI Hiring Tool

Author: Richa Sharma

Introduction

In 2018, Amazon, a major company in online sales and cloud computing, started using an artificial intelligence system for hiring new employees. The goal was to use advanced technology to improve the hiring process. However, using AI for hiring also raised concerns about data bias and unfair algorithms.

This essay will examine what happened with Amazon's AI hiring system in 2018 and discuss the issues of data bias and unfair algorithms when using AI for hiring. By looking closely at the ethical concerns and practical considerations of using AI for recruitment, this essay will highlight the importance of ensuring fairness and equality in decision-making processes that involve technology.

The Amazon's scandal

In 2018, Amazon introduced an AI-driven recruitment system to enhance its hiring practices and leverage innovative technology. According to Amazon's website, this initiative was intended to improve efficiency and streamline HR processes. These moves reflected Amazon's recognition of AI's potential to transform recruitment (Amazon.com, 2023).

The rollout of this AI-driven recruitment system was met with considerable anticipation and excitement within the company. Amazon viewed it as a bold step forward in optimizing its recruitment processes and attracting top talent globally. The system aimed to address long-standing challenges, such as the time-consuming nature of resume screening and the difficulty of identifying the most qualified candidates from a large pool of applicants (Yahoo.com, 2024). This move was particularly crucial for Amazon, given its rapid expansion and the need to fill thousands of positions across various departments and geographical locations.

As the implementation of the AI-driven recruitment system progressed, both opportunities and challenges emerged for the company. The system was designed to improve efficiency and accuracy in candidate selection. By automating the initial screening of resumes, Amazon aimed to allow recruiters to focus more on strategic aspects of the hiring process, such as assessing cultural fit and conducting interviews. This shift was intended to streamline routine tasks and enable more meaningful interactions with candi-

dates. Such automation aligns with broader trends in HR technology aimed at reducing administrative burdens and enhancing decision-making processes (Boden, 2019). Furthermore, Amazon hoped that by leveraging AI, it could identify unconventional talent that human recruiters might overlook, potentially diversifying its workforce and bringing in fresh perspectives.

However, alongside these opportunities, the implementation of the AI-driven recruitment system also presented several challenges that needed to be addressed. One of the key challenges was ensuring that the system's algorithms were unbiased in their evaluation of candidates. Concerns about algorithmic bias arose when it was discovered that the system inadvertently favored certain demographics over others (Women in Tech, n.d.). This revelation sparked internal debates about the ethical implications of AI in recruitment, with some employees questioning whether the technology was truly ready for such a critical task.

In October 2018, several reports, including one published by Computer Business Review (CBR Staff Writer, 2018), surfaced that Amazon's AI recruitment tool was biased against women. This bias reflected existing gender disparities in the data used to train the algorithms. The AI system had been trained on resumes submitted to Amazon over a ten-year period, most of which came from men, reflecting the male-dominated tech industry. Consequently, the algorithm began to penalize resumes that included the word "women's," as in "women's chess club captain," and downgraded graduates of two all-women's colleges.

Following these revelations, Amazon decided to retire the system. The decision was made after internal investigations and scrutiny from various stakeholders who highlighted the biases within the system (CBR Staff Writer, 2018). The internal reviews involved cross-departmental teams, including ethics boards, technical teams, and HR specialists, who conducted thorough assessments of the AI's performance. Despite attempts to reprogram the algorithms to be neutral, the biases proved too ingrained to eliminate completely.

This case exemplifies the broader ethical issues surrounding the use of AI in hiring, underscoring the importance of developing and implementing technologies that uphold principles of fairness and equality (O'Neil, 2016). The incident served as a wake-up call not just for Amazon, but for the entire tech industry, prompting a reevaluation of AI's role in high-stakes decisions that affect people's lives and careers. It highlighted the necessity of transparency, rigorous testing, and continuous monitoring of AI systems to prevent similar issues in the future. Amazon's experience has since influenced other companies to approach AI implementation with greater caution and a deeper understanding of ethical

considerations.

The bias against women in Amazon’s AI recruitment tool exemplifies the interconnected issues of data and algorithmic unfairness in AI ethics.

Data unfairness occurs when the datasets used to train AI systems are not representative of the broader population or reflect existing prejudices and inequalities. This can lead to discriminatory outcomes when the AI system makes decisions based on these biased datasets. In the case of Amazon’s AI-driven recruitment system, several instances of data unfairness contributed to its eventual failure.

As historical hiring data exhibited a preference for candidates from specific backgrounds or educational institutions, the AI algorithms would learn and replicate these biases, leading to the exclusion of qualified candidates from underrepresented groups. This data unfairness raised significant ethical concerns and undermined the system’s credibility as a fair and objective tool for candidate evaluation (ACLU, 2023). Research has shown that biased training data can perpetuate existing disparities, resulting in discriminatory outcomes in AI-driven decision-making processes (Barocas & Selbst, 2016).

The root of the problem with Amazon’s AI recruitment tool was the biased training data. The AI was trained on resumes submitted to Amazon over a ten-year period, most of which came from men. This historical data reflected the gender disparities prevalent in the tech industry, where men have traditionally held a larger share of technical roles. Consequently, the AI system learned to favor resumes that resembled those of past successful candidates, inadvertently penalizing those that did not fit this pattern (Bolukbasi et al., 2016). This led to the system downgrading resumes that included words like “women’s,” as in “women’s chess club captain,” while favoring resumes that used more masculine language (Dastin, 2018).

The data unfairness had a disproportionate impact on women applying for technical positions at Amazon. The AI system’s bias against resumes from female candidates perpetuated the underrepresentation of women in technical roles. This outcome is a clear violation of the ethical principles of beneficence, respect for persons, and justice. Beneficence requires that actions should promote good and avoid harm, but the biased system caused harm by unfairly excluding qualified female candidates. Respect for persons involves treating individuals with dignity and fairness, which was compromised when the system unfairly judged candidates based on gender. Justice demands equitable treatment and opportunity for all, which was not upheld in this scenario (Sweeney, 2013).

Another aspect of data unfairness was the lack of diversity in the data collection process. The data used to train the AI system did not adequately represent diverse demographics,

including various gender, racial, and ethnic groups. This homogeneity in the training data resulted in an AI system that was ill-equipped to fairly evaluate a diverse applicant pool. The system's inability to recognize and appropriately value the qualifications of candidates from different backgrounds further exacerbated the problem of bias (Barocas & Selbst, 2016).

Addressing data unfairness required a comprehensive approach that involved re-evaluating the data sources used to train the algorithms and implementing measures to mitigate bias. Techniques such as re-weighting training data, incorporating fairness constraints, and using diverse datasets can help reduce bias and promote equity (Zliobaite, 2015). As Amazon retired the product, this decision highlighted Amazon's recognition of the importance of diversity and inclusion in the recruitment process and its commitment to ensuring that the AI-driven system was calibrated to prioritize meritocracy and fairness. However, achieving true fairness in algorithmic decision-making remains a complex and ongoing challenge, requiring continuous monitoring and adjustment to minimize the impact of biases in the data (Hajian et al., 2016).

The case study of Amazon's AI-driven recruitment system served as a valuable lesson in the importance of ethical considerations in the development and implementation of AI technologies in HR. It underscored the need for organizations to critically assess their data sources and implement robust mechanisms to detect and mitigate biases. This case also emphasized the broader implications of AI fairness in maintaining public trust and ensuring equitable opportunities for all candidates (O'Neil, 2016). Ultimately, while AI holds significant potential to enhance recruitment processes, it must be developed and deployed with a strong ethical foundation to avoid perpetuating inequalities and to uphold principles of justice and fairness.

The unfairness experienced in the recruitment process with AI comes from the bias present in data and from the way we handle data through algorithms.

Algorithmic unfairness occurs when the algorithms used in AI systems produce biased or discriminatory outcomes, even if the data they are trained on is fair and representative. This form of unfairness can result from various factors, including the design of the algorithms, the way they process data, and the unintended consequences of their operational logic. In the context of Amazon's AI-driven recruitment system, algorithmic unfairness manifested in several significant ways, ultimately leading to the system's discontinuation. The design of the algorithms used in Amazon's AI recruitment system played a crucial role in perpetuating bias. These algorithms were engineered to optimize for specific outcomes, such as identifying candidates who resembled past successful hires.

This optimization process inadvertently encoded and reinforced existing biases within the hiring data. This phenomenon, often referred to as a feedback loop or recursive bias, contributed to the perpetuation of inequality within the recruitment process (Bolukbasi et al., 2016). For instance, if the algorithm placed a higher weight on experience from certain prestigious institutions, it could unfairly favor candidates from those backgrounds, disadvantaging candidates who did not fit this profile (Barocas & Selbst, 2016).

One significant aspect of algorithmic unfairness in Amazon's system was the feature selection and weighting process. The algorithm was likely trained to prioritize certain features from resumes, such as specific job titles, educational backgrounds, or keywords. This prioritization can lead to biased outcomes if the selected features disproportionately represent certain groups. An illustrative example of this bias could be seen in how the algorithm might favor candidates who attended Ivy League universities over those who attended community colleges, regardless of their actual qualifications or skills (Barocas & Selbst, 2016). In addition to data unfairness, Amazon's AI recruitment system also faced criticism for algorithmic unfairness. Even when trained on unbiased data, AI algorithms can still produce discriminatory outcomes due to the complexity of the underlying algorithms and the variables they consider. In the case of Amazon, the specific criteria used by the AI to evaluate resumes were not transparent, making it difficult to identify and address potential sources of bias. This lack of transparency raised concerns among both job seekers and industry experts, who called for greater accountability and oversight in the use of AI for recruitment purposes (BBC News, 2018). Algorithms often function as 'black boxes,' making it difficult to understand how decisions are made and to identify sources of bias. This opacity can prevent organizations from detecting and correcting unfair practices. In Amazon's case, the internal workings of the AI recruitment system were not transparent, complicating efforts to diagnose and address the biases that emerged (Pasquale, 2015). Algorithmic unfairness also resulted in disparate impact, where the AI system's decisions disproportionately affected certain groups of people, even without intentional discrimination. In the case of Amazon, the system's bias against female candidates for technical positions exemplified this issue. The AI tool systematically downgraded resumes with terms associated with women's activities and organizations, leading to fewer women being shortlisted for interviews. This outcome violated principles of justice and equality, as well as potentially exposing Amazon to legal challenges under anti-discrimination laws (Friedler, Scheidegger, & Venkatasubramanian, 2016). In response to the challenges faced by Amazon and other companies experimenting with AI in recruitment, alternative solutions have emerged. One such solution is HireVue, a platform that employs AI to analyze video interviews and assess candidates based on verbal and non-verbal cues. The objective of HireVue is to offer a more standardized and

objective approach to candidate evaluation, aiming to mitigate bias inherent in traditional resume screening processes. By scrutinizing factors such as facial expressions, tone of voice, and language use, HireVue purports to provide insights into candidates' skills, personality traits, and cultural fit.

While platforms like HireVue offer potential benefits in terms of objectivity and efficiency, they also raise concerns about privacy, fairness, and the risk of algorithmic bias. As with any technological advancement, it is imperative to meticulously evaluate the impact of AI-driven solutions on the recruitment process, prioritizing fairness, transparency, and ethical considerations (Hirevue, 2024).

Another company offering an alternative approach to traditional recruitment methods is Pymetrics. Pymetrics utilizes neuroscience-based games and exercises to evaluate cognitive and emotional traits of job candidates. These games are designed to measure skills such as attention, memory, and problem-solving abilities in a manner that is more engaging and less biased than conventional assessments. Pymetrics' platform analyzes candidates' performance in these games to generate a cognitive and emotional profile, subsequently matching candidates with job roles that align with their strengths and abilities.

By focusing on objective measures of cognitive and emotional skills, Pymetrics aims to mitigate bias and enhance the likelihood of finding the right fit between candidates and job roles. Analogous to HireVue, Pymetrics offers the potential for a more objective and data-driven approach to candidate assessment. Nevertheless, it also prompts questions regarding the validity and reliability of using gamified assessments as predictors of job performance, alongside concerns about privacy and fairness in algorithmic decision-making (Pymetrics, n.d.).

Amazon's experience with its AI recruitment system offers valuable lessons for other organizations deploying AI in human resources and beyond. First, it highlights the importance of critically examining both the data and the algorithms used in AI systems to identify and mitigate potential biases. Second, it underscores the need for ongoing monitoring and evaluation to ensure that AI systems remain fair and effective over time. Lastly, it emphasizes the ethical responsibility of companies to address algorithmic unfairness proactively, safeguarding against discrimination and promoting equal opportunities for all candidates.

Conclusion

Amazon’s experience with its AI hiring tool in 2018 teaches us a valuable lesson. They tried to use new technology to make hiring better, which was a smart idea. But things didn’t go as planned. The AI tool unfairly preferred men over women for jobs, which wasn’t fair or right. This shows us that even when we try to do something good with technology, like making hiring faster and better, we can accidentally cause problems if we’re not careful.

The main issue was that the AI didn’t treat everyone equally. It didn’t value the different skills and experiences of all candidates, especially women. This is wrong because everyone deserves a fair chance, no matter who they are. The AI’s bias came from the information it was given, which had more data about men than women. This led to unfair treatment, which goes against the idea of justice that says everyone should have equal opportunities.

What we learn from Amazon’s story is that when we use new technologies like AI, we need to be extra careful. We should make sure the AI understands and respects all people equally. This means giving the AI fair and balanced information, making sure we can understand how it makes decisions, and always checking to make sure it’s being fair. Companies like Amazon should focus on doing what’s right, respecting everyone, and being fair in their use of technology.

AI can be great for things like hiring, but we must use it responsibly. Amazon’s experience reminds us that technology should be used in a way that respects everyone and gives them a fair chance. It’s about ensuring that as we move forward with new tech, we don’t leave behind our values of fairness and respect for everyone.

Bibliography

- ACLU (2023) 'AI and the future of hiring: Ensuring fairness and equity', American Civil Liberties Union.
- Amazon.com (2023) 'AI-driven recruitment system', Amazon.com. Accessed 15 May 2024.
- Barocas, S. and Selbst, A. D. (2016) 'Big data’s disparate impact', *California Law Review*, 104(3), pp. 671-732.
- BBC News (2018) 'Amazon scrapped 'sexist AI' tool', BBC News.
- Boden, M. A. (2019) *AI: Its nature and future*. Oxford: Oxford University Press.
- Bolukbasi, T. et al. (2016) 'Man is to computer programmer as woman is to homemaker? Debiasing word embeddings', *Proceedings of the 30th International Conference on Neural*

Information Processing Systems, pp. 4356-4364.

CBR Staff Writer (2018) 'Amazon scraps AI recruitment tool that showed bias against women', Computer Business Review, October.

Dastin, J. (2018) 'Amazon scraps secret AI recruiting tool that showed bias against women', Reuters.

Friedler, S. A., Scheidegger, C. and Venkatasubramanian, S. (2016) 'On the (im)possibility of fairness', arXiv preprint arXiv:1609.07236.

Hajian, S., Bonchi, F. and Castillo, C. (2016) 'Algorithmic bias: From discrimination discovery to fairness-aware data mining', Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 2125-2126.

Hirevue (2024) 'How HireVue works', HireVue.

O'Neil, C. (2016) Weapons of math destruction: How big data increases inequality and threatens democracy. New York: Crown Publishing Group.

Pasquale, F. (2015) The Black Box Society: The secret algorithms that control money and information. Cambridge: Harvard University Press.

Pymetrics (n.d.) 'The science behind Pymetrics', Pymetrics.

Sweeney, L. (2013) 'Discrimination in online ad delivery', ACM Digital Library, 11(1), pp. 44-57. Women in Tech (n.d.) 'Gender bias in AI recruitment: Case studies and solutions'.

Yahoo.com (2024) 'Amazon's AI recruitment tool: a game-changer or a bias machine?', Yahoo.com, 15 May.

Zliobaite, I. (2015) 'A survey on measuring indirect discrimination in machine learning', Computing Research Repository.

AI Tool Logs

Instance	AI Tool used	Purpose	Section of the Essay
1	OpenAI ChatGPT 4.0	Improving language clarity	Introduction
2	OpenAI ChatGPT 3.5	Structuring the argument	Body of the essay (pg 2-6)
3	OpenAI ChatGPT 3.5	Organizing entries in Harvard style	Bibliography

Table 1: Table of AI Tool Logs