HUMAN RESOURCES

Why Amazon's AI Kept Rejecting Women (and What It Reveals)

Monisha Bichve, Linu George, Rajat Gera

Amity Business School, Amity University Maharashtra, Mumbai, India

Published November 2026

Abstract

Between 2014–2018, Amazon developed an AI recruitment tool. The aim was to streamline the process of reviewing candidate resumes by automatically scoring them between one and five stars. The ultimate goal was to find the best technical talent but instead became one of the most well known examples of machine learning picking up on human bias.

1 Introduction

Between 2014–2018, Amazon developed an AI recruitment tool. The aim was to streamline the process of reviewing candidate resumes by automatically scoring them between one and five stars. The ultimate goal was to find the best technical talent but instead became one of the most well known examples of machine learning picking up on human bias.

The model was trained using a decade of resumes, most of which were male dominated due to the existing gender disparity in the tech industry. It soon became apparent that resumes with the word "women's" were being marked down. This was first noticed in 2015 when Amazon realised that graduates from women's colleges were being devalued. The project was scrapped and Amazon admitted that the tool was never actually used to make real hiring decisions, but was looked at by recruiters.

This example is important because it shows that bias is rarely put into place deliberately. There was no intention to exclude women from the process. The discrimination was based on the data provided making it more difficult to identify and solve.

2 Background and context

This project started within the Amazon engineering hub in Edinburgh back in 2014. The team included around a dozen experts working together to develop this project. At the time machine learning was becoming a viable option and Amazon were looking to prepare their HR department for exponential growth. Since June 2015, Amazon have increased their global employee count by over 300%.

This team had incredibly high ambitions. They built 500 different computer models that would specialise within certain roles and areas. Within these they trained the system to understand approximately 50,000 words from previous resumes. The aim was to provide the user with 100 resumes and the system would return the top 5 to be hired. The system was never intended to fully replace humans but significantly reduce the amount of candidates for the recruiter to look through.

3 The problem: How bias manifested

By 2015, roughly a year into development, the team realized the system was not rating candidates in a gender-neutral way. The specific failures were concrete and disturbing:

- Resumes containing "women's" (such as "women's chess club captain") received lower scores
- Graduates of two unnamed all-women's colleges were downgraded
- The algorithm favored verbs more common on male engineers' resumes, such as "executed" and "captured"
- Technical skills common to all were given very little importance such as coding ability as this wouldn't help differentiate candidates.

The bias wasn't just based on gender. The issues with the data meant that many candidates weren't qualified for the jobs they were recommended. The results were almost random meaning the project wasn't viable.

4 Root causes

The training data was the problem. Amazon's models learned by observing patterns in resumes submitted over a ten-year period. Most came from men, a direct reflection of male dominance in tech. At the time, Amazon's own workforce was roughly 60:40 male-to-female. The algorithm did exactly what it was designed to do: it found patterns in the data and optimized for them. Those patterns happened to correlate with gender.

The model learned proxies, not gender itself. The system never "knew" a candidate was a woman. It learned that certain words, colleges, and writing styles appeared more often among candidates who had historically been hired. Since men had been hired more often, male-associated patterns became the signal for success. This is the insidious nature of algorithmic bias: it operates through correlations that no human explicitly designed.

The bias came from humans. One analysis stated that, "only modelled the recruiting process, as it had been applied by Amazon teams over several years. It therefore reproduced the biases that are inherently human". The system was not introducing any new form of discrimination but instead simply automated patterns that were already in place.

5 HR practices involved

This case highlights many components of the recruitment process;

- **Recruitment & selection:** The algorithm was designed to score the candidates.
- **Automated resume screening:** To reduce or replace manual screening.
- **AI based recruitment:** One of the first known attempts from a company to incorporate a ML into their hiring process.
- **Diversity and equal employment opportunity:** The biased issue caused the risk of legal issues and goes against Amazon's commitments of creating a diverse work environment.

It should be noted that Amazon recruiters did take the recommendations from the algorithm when looking for prospects. However, there were never any instances where they just used the rankings. This suggests that the biased recommendations still affected decision making in this aspect.

6 Remediation attempts

To overcome this issue Amazon's engineers attempted to remove bias. To do this they altered the programmes to stop them recognising bias within specific words (removing "women's" and the names of female colleges).

The fix did not restore confidence. The team could not guarantee that the machines would not devise other ways of sorting candidates that could prove discriminatory. Removing explicit indicators of gender does not eliminate the underlying learned associations. The model could find new proxies: different words, different patterns, different combinations that correlated with gender in ways no one had anticipated.

This is a fundamental limitation of surface-level bias mitigation. As one commentary put it, the algorithm was not "sexist by nature" but was "taught to be" through data that reflected historical inequities. Editing the inputs does not unlearn what the model has already absorbed.

7 Action taken and outcome

Amazon shut down the project. The Edinburgh team was disbanded by the start of 2017, roughly three years after work began, because executives lost hope that the tool could be made reliable.

Amazon stated that the tool "was never used by Amazon recruiters to evaluate candidates," though the company did not dispute that recruiters viewed the recommendations it generated. Some salvage occurred: a "much-watered down version" was reportedly used for rudimentary tasks like culling duplicate candidate profiles from databases.

The abandonment was not a public announcement. The project had been secret, and its failure only became known through the Reuters investigation in October 2018, based on interviews with five anonymous insiders.

8 Legal and ethical implications

Discrimination risk was real. Had the tool been deployed more extensively, its behavior—penalizing candidates based on gender-associated terms—could have constituted disparate impact under equal employment opportunity principles. The fact that the system was abandoned before full deployment limited Amazon's legal exposure, but the case demonstrates how easily AI tools can create liability.

Regulatory frameworks were not ready. In 2018, there was no specific law governing AI hiring tools. That has begun to change. New York City's Local Law 144, effective in 2023, now requires bias audits for automated employment decision tools and public summaries of selection rates by sex and race. The Amazon case is frequently cited as justification for such oversight.

Ethical responsibility remains contested. Amazon did not intend to discriminate. The bias emerged from data that reflected a discriminatory industry. This raises difficult questions: how far does corporate responsibility extend when the harm is produced by systems rather than people? The case suggests that "we didn't mean to" is not an adequate defense when the outcome is foreseeable and the data is known to be skewed.

9 Key learnings for HR

Training data is never neutral. Historical hiring data carries historical inequities. Any model trained on it will reproduce and potentially amplify those patterns.

Bias can be indirect. Removing explicit gender terms is not enough. The model may learn correlated proxies—verbs, college names, writing styles—that function as discriminatory signals.

Audits must be proactive, not reactive. Amazon discovered the problem roughly a year into development. A systematic bias audit before training, and ongoing testing during development, might have caught it earlier.

Human oversight is non-negotiable. Automated screening must remain explainable and subject to human review, especially for high-stakes decisions. The "human-in-the-loop" principle is now widely advocated, but it only works if humans understand what the system is doing and why.

Resume predictions aren't great. Some analysts argue that the real learning point here is that using resumes (either manually or automatically) isn't a good predictor for job success. This is made worse by the bias issue.

10 Rationale

The case is real life, documented and self corrected. The case was based on five genuine first-hand accounts taken from members of the team, and Amazon have not challenged the key findings of this case. It walks you through the whole process, from data gathering, model training, bias identified, attempt at fixing and abandoning the model.

The most important thing is that the bias was unintentional. No-one set out to create a sexist system. The discrimination came from a genuine attempt to automate using poor quality training data. This challenges the idea that technology is unbiased, and therefore is an excellent case study for anyone working within HR technology or AI governance.

11 Conclusion

This case study shows how easily discrimination can be replicated in algorithmic systems by accident. The model had been trained on male dominated resumes over a 10 year period and therefore decided that male related patterns were better. No-one programmed the model to favour males, however it was statistically rational based on the data provided.

Fairness in AI recruitment requires careful design, testing across different demographics, visibility into how the model is behaving and AI governance. Fixing bias after the fact is not a solid solution. Simply removing one proxy does not remove the learned association. In hindsight Amazon made the right decision to abandon the tool. This case shows how dangerous bias can be. Not only will there be legal issues and damage to reputation, but qualified candidates will be excluded due to the use of the tool.

References

[1] Reuters (via RNZ) (2018). Amazon scraps sexist AI recruiting tool.

[2] Reuters (via RTÉ) (2018). Amazon scraps secret AI recruiting tool that showed bias against women.

[3] BBC News (2018). Amazon scrapped 'sexist AI' tool.

[4] Communications of the ACM (2018). Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women.

[5] AssessFirst (2026). Lessons learned from Amazon's "sexist algorithm".

[6] PyPI (2026). aedt-audit: Bias-audit artifacts for automated employment decision tools (for NYC LL144 context).