Beyond Prejudice: Exploring AI's Impact on Fair Hiring Decisions

Chaymaa Dkouk El Ferroun
Universitat de Lleida
Lleida
chaymaadkouk674@gmail.com

ABSTRACT
Bias permeates various aspects of recruitment and hiring processes, including but not limited to ethnicity, gender, and socioeconomic background, leading to disparities and inequities in employment opportunities. This article narrows its focus to the issue of gender bias within hiring practices. Despite advancements in workplace equality, gender bias continues to hinder the advancement and representation of women in various industries and positions. Through an examination of existing literature and case studies, this article shows the multifaceted nature of gender bias in recruitment and hiring, highlighting its detrimental effects on organizational diversity and gender equality. Moreover, it explores the potential of artificial intelligence (AI) to mitigate gender bias in hiring processes.

1 Traditional Hiring Biases
In the modern workplace, traditional hiring methods have faced significant challenges and transformations. In the past, individuals seeking employment could often personally submit their resumes to enterprises, fostering a more direct and personal interaction in the hiring process. However, with the advent of digital platforms such as LinkedIn and InfoJobs, the landscape of recruitment has evolved dramatically. Job postings can now attract thousands of applicants, making the selection process more daunting and impersonal and causing for many job applicants to be ignored. Due to the large volume of applicants for open positions, manual selection processes have become impractical. Consequently, recruiters often rely on resume screening, a method prone to unintentional biases, including gender bias.

It has been found in a research article published by the Proceedings of the National Academy of Sciences [1] that upon doing the same task equally well, both men and women were twice more likely to hire a man than a woman. Other impactful findings include the results of a research conducted in Harvard and Princeton [2], which found that blind auditions increased the chances of women being hired by 25-46%.

Despite our best efforts to be fair, the truth remains that our hiring processes often inadvertently perpetuate inequality within our society. This systemic issue, though unintended, contributes to widening disparities and reinforcing existing inequalities. Acknowledging the existence of these biases is the first step towards mitigating their impact and fostering a more inclusive and equitable decision-making environment.

2 AI For Eliminating Bias
As already mentioned, bias can occur subconsciously, and it may be difficult to overcome. One promising solution for this problem is the usage of Artificial Intelligence (AI) to promote efficient and fair hiring processes.

The benefit of using AI lies in the fact that we can train it to meet our specific needs and evaluate it according to our standards so that we can remove bias. By standardizing the hiring process and trusting an algorithm to make the decisions, we can aim for a more ethical and fair assessment of job applicants.

Incorporating AI into our hiring processes isn't merely wishful thinking; it's supported by research indicating that it could yield significant positive outcomes for our society. This includes the studies by Sajjadi et al. [3] and Sühr et al. [4], which demonstrate that machine learning algorithms do not exhibit adverse effects based on gender and can even enhance the selection of female candidates. Specifically, Li et al. [5] found that some algorithms could increase the share of women selected by a 15% with respect to the decisions made by humans.

3 Methodology
To illustrate AI's potential to promote fairer hiring decisions regarding gender, this study employs a machine learning approach. Specifically, we train machine learning algorithm on the Kaggle ‘Job Placement Dataset’ [7]. The source code is available on GitHub1.

1 https://github.com/Chaymaa19/Job-Placement-ML.git
The dataset poses significant challenges, notably due to its substantial gender imbalance, with twice as many male samples as female. Furthermore, an analysis of placement decisions reveals a pronounced disparity in placement ratios between male and female students, with 70% of males placed compared to only 60% of females.

Overlooking these aspects could lead to the development of an algorithm that appears effective superficially but harbors gender bias, resulting in unfair outcomes for women. However, this bias can be quantified and mitigated through the application of both data and model selection bias mitigation techniques, as undertaken in this study.

4 Results

To assess the gender imbalance in a model, we've utilized various performance metrics. These metrics include overall accuracy, accuracy across gender categories (male and female), positive and negative rates, as well as true positive and true negative rates.

Initially, a baseline model was trained without employing any debiasing techniques. As expected from the initial examination of the data, this model exhibited a preference for placing males and not placing females. This bias is evident in Figure 1, where males have a higher placement rate, both in the points labeled as true_positive_rate and in the predicted placement rate. Conversely, females exhibit a higher rate of not being placed, as observed in both the true_negative_rate and predicted not placed rates.

Figure 1: baseline gender bias metrics.

To address this issue, we applied various debiasing techniques, ultimately finding that counterfactual augmentation coupled with gradient boosting yielded the best results. This approach achieved optimal performance while effectively mitigating gender biases.

Figure 2: gender distribution in the dataset.

In Figure 2, we demonstrate the impact of these debiasing techniques. By applying them, we successfully reduced the initial gaps observed between genders, resulting in a machine learning model that exhibits no gender preferences while maintaining high performance levels.

5 Conclusions

The study suggests that AI has significant potential to enhance inclusivity and fairness in hiring practices. By adopting AI technologies, organizations can prioritize meritocracy and diversity, fostering more representative workplaces. While further investigation is warranted, these initial results are promising, signaling a positive direction for future research and implementation.

REFERENCES