Fairness and Bias in Job Hiring

Yuhan Chen, Jinqiu Fan, Yating Yang
Department of Computer Science, University of Southern California, Los Angeles, California 90007, USA
(Dated: May 3, 2024)

This paper investigates biases in algorithmic hiring using two datasets, revealing significant gender biases in recruitment decisions. We evaluated two debiasing strategies: excluding gender data and augmenting datasets. Both methods effectively reduced bias, enhancing fairness metrics like Statistical Parity and Equalized Opportunity. The exclusion of gender information slightly impacted model accuracy, while data augmentation improved it, demonstrating that careful interventions can promote fairness without compromising performance. Our findings highlight the importance of ongoing monitoring and adaptation in AI-driven recruitment to ensure fair and equitable employment practices.

Keywords: Fairness, Bias, Algorithmic fairness, Machine learning, Job hiring, Recruitment, Diversity, Discrimination, Equal opportunity, Ethical considerations, Human resources

I. INTRODUCTION

Inequities and biases are presented as big challenges in the recruitment practices, no matter what tools companies utilize. These biases, whether conscious or not, can stem from cultural stereotypes, personal prejudices or systemic issues within organizations. According to research, inequity in the hiring process is increasing year over year. This is likely due to the emergence of artificial intelligence, where companies replace part of the hiring process by recommender algorithms. The use of algorithmic hiring has brought about the digitization of information, making candidate ranking more inclined to value similarity of Job Description and candidate's resumes at the expense of a person's characteristics and potential abilities. If not carefully designed and managed, these technologies have the potential to perpetuate the existing biases in recruitment.

In the algorithmic hiring process, bias can be presented in terms of gender, race, color, and personality. Gender stereotypes can arise because algorithms are trained on historical data that reflects past hiring decisions that favor one gender over the other. Racial discrimination may arise also because of training on historical data that favor certain groups of people. Color prejudice is closely related to racial prejudice, but can also extend to prejudice against individuals in racial or ethnic groups based on skin color. Word choice, tone shifts, and facial expressions can be used to determine candidates' personality, which may lead to algorithmic bias in personality. Ensuring transparency and fairness in AI recruitment tools is critical since hiring decisions are based on results from the algorithm.

In this paper, we plan to explore the unfairness and bias present in hiring practices, a crucial concern in contemporary employment landscapes. We will discuss the various origins of these biases, determine whether bias exist in recruitment process and how bias influence hiring decisions, and also evaluate bias mitigation methods. This paper will also highlight the complex interplay between human decision-making and modern technological algorithmic hiring tools in recruitment processes.

II. RELATED WORK

To better understand the issue of bias in job hiring processes, it is essential to examine the existing literature on the subject. This section aims to provide a comprehensive overview of the various factors that contribute to biased hiring practices and their implications. The review will cover the current state of algorithmic hiring systems, the sources of bias in these systems, the impact of biased data and models on recruitment outcomes, and the strategies proposed to mitigate bias. Through an exploration of the existing literature, this section will highlight the key factors contributing to bias in job hiring processes and provide insights into the potential consequences of these biases.

Factors such as biased datasets, discriminatory features, and flawed model design contribute to algorithmic bias in hiring systems. Chen's review paper [5] analyzed the reasons for algorithmic recruitment discrimination, including biased datasets and feature selection, which can manifest as gender, race, color, and personality biases. Kumar et al.'s survey [1] explored the recommendation algorithms used in recruitment systems, the datasets employed, fairness definitions, and evaluation metrics. They found that the Earth Mover's Distance and Mann-Whitney U test were commonly used for evaluating Demographic Parity, while True Positive Rate Parity was used for assessing Equal Opportunity. Notably, the authors highlighted the scarcity of open-source datasets in this domain, with most research relying on data.

In the multidisciplinary survey "Fairness and Bias in Algorithmic Hiring" [2], the authors examined the current state of algorithmic hiring systems, the fairness

and bias concerns, and strategies to address bias. The paper gathered information from various literature sources, datasets, and hiring systems, analyzing how these systems could contribute to biased outcomes. It also highlighted methods to mitigate bias by integrating insights from computer science, human resource management, philosophy, and law.

Given the sensitivity of this topic, although our research showed a lack of publicly available datasets, we were still able to find recruitment dataset - "Utrecht Fairness Recruitment dataset" [7], as well as one real job applicants' dataset in the recruitment domain - "Employability Classification of Over 70,000 Job Applicants" [9]. These datasets allow for an more in-depth investigation of bias in job hiring processes.

III. METHODOLOGY

A. Analysis 1

1. Dataset

The "Utrecht Fairness Recruitment dataset" contains the recruitment decisions of companies for over 500 candidates, the sensitive attribute: gender, age, and nationality, as well as attributes related to recruitment decision such as university grade, programming experience, international experience for each candidate. This dataset would provides information on types and patterns of bias and discrimination in hiring practices. We plan to use visualization and modeling to see whether biases exist in the dataset regarding gender, age, nationality and sports in recruitment.

2. Data Exploration

For this dataset, we focused on detecting whether there is gender bias in job hiring. In order to dig deeper into the recruitment data, we performed cross-analysis on sensitive attribute such as gender and their likelihood of being accept by the company.

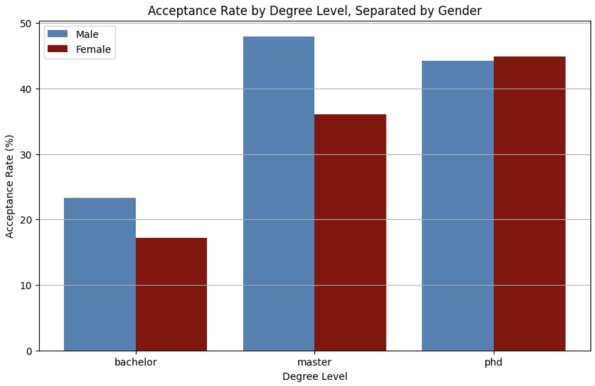

FIG. 1. Acceptance Rate by Degree Level, Separated by Gender

These numbers point to clear systemic biases against women in the hiring process, at least at the bachelor's and master's degree levels. With equivalent educational qualifications, women are being passed over for roles that their male counterparts are securing at much higher rates. The reasons behind this are complex, but likely stem from deep-rooted societal biases, inadequate professional networks and mentorship for women, and perhaps even discrimination during the interview process itself.

Interestingly, the one bright spot is that there appears to be more gender parity for those with doctoral degrees. This could suggest that ultimately, elite educational credentials and highly specialized expertise can help override gender biases during hiring to some degree. Or it may simply reflect the relatively small sample sizes of PhD holders.

Regardless, these statistics lay bare an unfortunate reality - despite decades of progress, gender discrimination remains deeply entrenched in many sectors when it comes to hiring and career advancement, particularly early on. Clearly much more work needs to be done by employers, policymakers and society at large to finally achieve true equal opportunity regardless of gender. The costs of failing to fully capitalize on the talents of over half the workforce are much too high.

3. Model

We first drop the gender attribute in the data, to examine the impact of this sensitive attribute on the decision tree's performance compared to the original data. By excluding the gender attribute, we can determine whether the decision tree model exhibits disparities in its predictions or outputs with and without the gender attribute. Comparing the model's performance under this gender-removed scenario to the real-world outcomes enables us to detect any concerning discrepancies that may

suggest bias related to gender in the recruitment process.

In this study, we also analyzed bias in the predictions using measures of Statistical Parity and Equal Opportunity. Statistical parity was measured by comparing the mean predicted outcomes between two groups defined by sensitive attribute which is gender in this case. Furthermore, we calculated the statistical equal opportunity, which measures discrepancies in the true positive rates between different gender groups. This metric is critical for understanding if both groups have equal chances of being correctly identified by the model.

4. Result

TABLE I. Comparison between ratio of being accepted

Model	Ratio of being accepted
Data without Gender	0.31606
Original Data	0.299242

As we can see in the Table I, the proportion predicted by the Decision Tree model without gender attribute(0.311) is slightly higher than the actual proportion of being accepted(0.299). This indicates that gender attribute influences the employment acceptance rate, indicating that the decision might be biased toward gender.

TABLE II. Evaluation Result

	Gender
Statistical Parity	-0.1206
Equalized Opportunity	0.1343

The resulting value (0.13434880722114761) in equal-opportunity implies a difference in treatment between the gender groups, with the gender group of male having a higher true positive rate. This suggests a bias in the model's performance, favoring male group over the other when it comes to predicting positive outcomes. We will further analyze how to mitigate bias in the next section.

B. Analysis 2

1. Dataset

The "Employability Classification of Over 70,000 Job Applicants" dataset contains a comprehensive collection of information regarding job applicants, their respective employability scores, and recruitment decision. It has sensitive attributes: age, gender, nationality, as well as attributes related to recruitment decision such as education level, programming experience, and previous salary. This dataset would provides information on types and patterns of bias and discrimination in hiring practices.

We checked for missing values in the data and found out the feature "HaveWorkedWith", indicating the programming languages applicants have worked with before, has 63 missing values, so we decided to remove this feature. For simplicity of analysis, we converted the "Country" attribute to "Continent", with categories: "Europe", "NorthAmerica', "Asia", "South America", "Australia", and "Others". Since the attribute "YearsCodePro" which describes how long the applicant has been coding in a professional context is taken out since it offers similar information as "YearsCode". We plan to use visualization and modeling to see whether biases exist in the dataset regarding gender, age, nationality in recruitment.

2. Data Exploration

There are three sensitive attributes in the dataset: gender, age, and nationality. Since nationality does not imply race and we do not know whether the data was collected in a particular country or not, we decided to focus on detecting potential bias in gender and age.

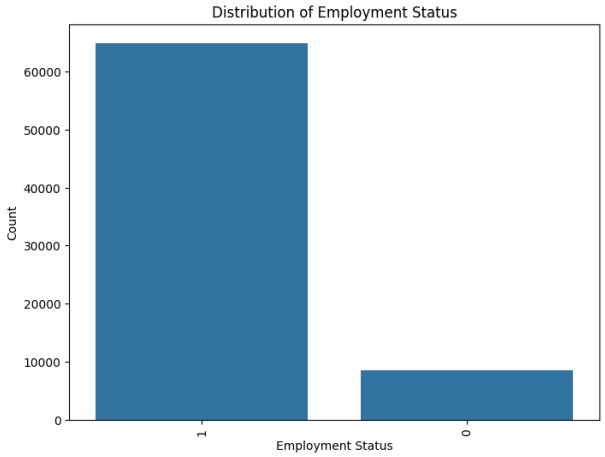

FIG. 2. Distribution of Employment Status

First, we explored whether there is potential gender bias in hiring decision. In the dataset, among all the job applicants, 93% are male, only 5% are female, and the rest 2% are non-binary. By plotting all the applicants' gender data vs. their employment decision, as shown in Figure 3, we can see that the number of male applicants who were hired surpassed those who didn't get hired, whereas it is quite the opposite for female. By calculation, there are 54% of male applicants who were hired; 53% for non-binary; only 45% for female. This observed

difference by 9% between male and female applicants strongly implies potential bias in gender when making hiring decisions.

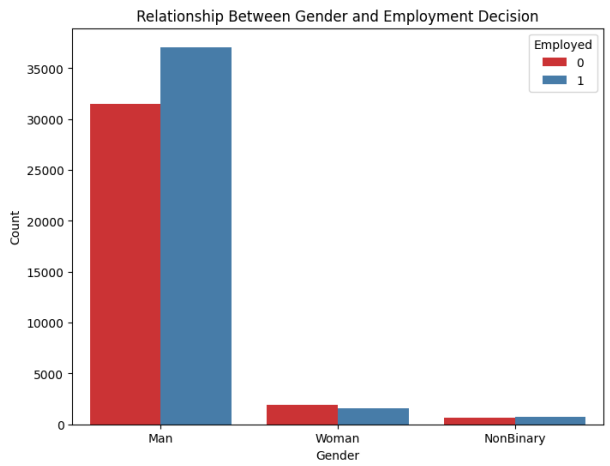

FIG. 3. Relationship Between Gender and Employment Decision

Second, we took a close look at the relationship between age and employment decision. As shown in Figure 4, we can see that the ratio of the number of applicants who were hired to whose who weren't hired is higher for applicants younger than 35, than those older than 35. There are 54.7% of applicants younger than 35 who were hired, whereas the hiring rate is only 51.6% for applicants older than 35. Therefore, age could also be a potential bias in the recruitment process.

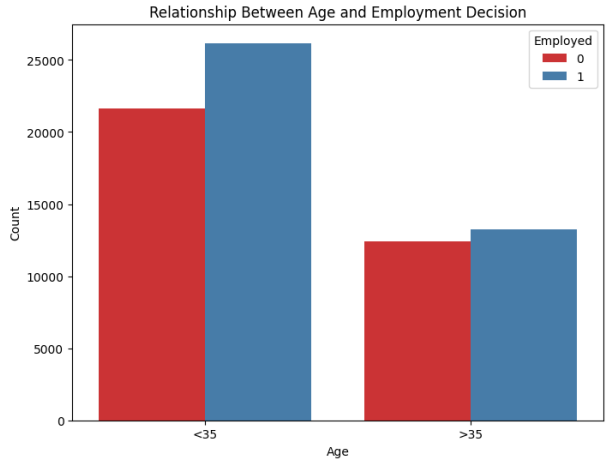

FIG. 4. Relationship Between Age and Employment Decision

As shown in this section, there are potential bias in both gender and age in the recruitment process. Next, we will use model to further analyze the influence of these two sensitive attributes to hiring decisions.

3. Model

We developed prediction models for recruitment decisions via decision trees to understand better how bias in dataset may influence the fairness of the recruitment decision prediction model, and how to mitigate bias. We first built the original decision tree model using all the attributes and evaluated the model using accuracy and two evaluation metrics: statistical parity and equalized opportunity. Then we explored two ways to mitigate bias in prediction: 1. excluding the protected feature; 2. data augmentation. We evaluated the bias mitigation techniques by computing the two fairness evaluation metrics and accuracy, and comparing them with the original model result. So we have three types of model:

Baseline Model: The baseline model is the decision prediction model trained using all the attributes.

Model Excluding the Protected Feature: This approach selectively excludes the protected feature when training the model to avoid its potential influence on recruitment decision prediction. This method allows the prediction to be independent of the protected feature so that we can better understand the influence of a specific sensitive attribute on the model's predictions.

Model with Data Augmentation: This model was developed by using date augmentation to counteract the effect of the protect feature. For every data in the original dataset, a new data sample with the opposite value of protected attribute is added. The synthetic is used for augmenting the original dataset. And the model is trained using the concatenated data (synthetic data + original training data).

4. Result

The findings are summarized in the following table:

	Age	**Gender**
Statistical Parity	0.026964	0.110484
Equalized Opportunity	0.007053	0.075860

TABLE III. Comparison between Sensitive Attributes

	Baseline	**Without Gender**	**Augmentation**
Accuracy	0.708797	0.708755	0.710941
Statistical Parity	0.110484	0.109773	0.110327
Equalized Opportunity	0.075860	0.054154	0.056495

TABLE IV. Comparison of models

Our analysis commenced with the calculation of two key fairness metrics: Statistical Parity and Equalized Opportunity, for two sensitive attributes—Age and Gender. The findings are summarized in Table III.

Analysis of Sensitive Attributes: Table III presents the fairness metrics for Age and Gender. The Statistical Parity for Age (0.026964) is considerably lower than for Gender (0.110484), indicating that Age as an attribute contributes minimally to bias within the dataset. Similarly, the measure of Equalized Opportunity for Age (0.007053) is much lower compared to Gender (0.075860), reinforcing the notion that Age does not significantly influence the fairness of the recruitment decision model. Based on these observations, it was determined that Age does not contribute substantially to bias. Therefore, we focuses on Gender attribute for subsequent model training aimed at debiasing method.

Comparison of Models: Following the initial analysis, we focused on Gender as the primary sensitive attribute for debiasing the predictive models. Three models were compared: Baseline, Without Gender, and Augmentation. The results are shown in Table IV.

The Baseline model, which included all attributes, demonstrated an accuracy of 0.708797, a Statistical Parity of 0.110484, and an Equalized Opportunity of 0.075860. Upon excluding the Gender attribute from the model (Without Gender), the Statistical Parity improved slightly to 0.109773, and Equalized Opportunity saw a more noticeable improvement to 0.054154, indicating a reduction in bias at the cost of a very minor reduction in accuracy (0.708755). The Data Augmentation model, which involved adding synthetic data to counter the protected attribute's bias, showed further enhancements in fairness metrics with a Statistical Parity of 0.110327 and an Equalized Opportunity of 0.056495. Interestingly, this model also showed a slight increase in accuracy (0.710941), suggesting that data augmentation may help in balancing fairness without compromising the model's predictive performance.

Summary: The examination of fairness metrics for Age and Gender indicates that Gender plays a more significant role in bias within the dataset used for recruitment decision modeling. The efforts to mitigate bias through the exclusion of Gender and data augmentation have shown promising improvements in fairness metrics, with minimal impacts on accuracy. These results highlight the effectiveness of the applied debiasing methods in enhancing model fairness.

IV. DISCUSSION

In the exploration of debiasing methods for recruitment decision prediction models, our analysis focused on the performance of decision trees across different configurations—specifically, the baseline, excluding the protected feature, and data augmentation models. The purpose of these configurations was to evaluate how the intentional manipulation of input data could influence both the fairness and the accuracy of recruitment decisions.

Bias Mitigation Efficiency: Our findings suggest that both debiasing strategies, excluding the protected feature and data augmentation, had measurable impacts on improving fairness metrics. For instance, the exclusion of the protected feature (Gender) in the model slightly reduced the Statistical Parity difference from 0.110484 to 0.109773 and significantly improved the Equalized Opportunity from 0.075860 to 0.054154. This indicates that removing the gender attribute from the decision-making process can decrease the likelihood of biased outcomes, particularly in terms of opportunity equality.

Similarly, the data augmentation approach, which involved supplementing the dataset with synthetic entries designed to counteract the biased representation of the protected attribute, showed improvements. The Statistical Parity was slightly better than the baseline at 0.110327, and Equalized Opportunity was also enhanced to 0.056495. This suggests that augmenting the data to represent a more balanced view of protected attributes can effectively reduce bias.

Bias and Accuracy Relationship: An essential aspect of deploying debiasing techniques is understanding the relationship between bias and model accuracy. In our analysis, the baseline model, which included all attributes, achieved an accuracy of 0.708797. When the gender attribute was excluded, the accuracy slightly decreased to 0.708755, illustrating a negligible impact on the model's performance despite the modification for fairness. In contrast, the data augmentation model showed a slight increase in accuracy to 0.710941. This increment, although small, suggests that augmenting the training data can potentially improve the model's predictive capabilities alongside enhancing fairness.

V. CONCLUSION

In this study, we undertook a detailed analysis of potential biases in recruitment practices using the Utrecht Fairness Recruitment dataset and Job Applicants dataset. For the Job Applicants dataset, after identify existing biases in hiring decisions, we further explored and evaluated various methods to mitigate these biases effectively.

Key findings: Our investigations revealed significant gender-related biases in the recruitment data, where

women were less likely to be hired compared to men with equivalent qualifications. This bias was present in both datasets, and had a significant influence in recruitment decision prediction. These findings highlight the persistence of gender discrimination in hiring practices, despite advancements in workplace equality.

Debiasing Efforts: To address these biases, we implemented two primary debiasing strategies: excluding gender from the decision-making model and augmenting the dataset with synthetic data to ensure a balanced representation of genders. Both methods proved effective in reducing bias, as reflected by improvements in fairness metrics such as Statistical Parity and Equalized Opportunity. The "Without Gender" model demonstrated that removing sensitive attributes could decrease unfair bias without significantly impacting the accuracy of the model. On the other hand, the "Augmentation" model showed that introducing synthetic data could not only decrease bias but also slightly improve the model's accuracy. This suggests that thoughtfully designed interventions can achieve fairness while maintaining or even enhancing the model's performance.

VI. FUTURE DIRECTIONS

The results of this study affirm that debiasing techniques can be effectively integrated into predictive models to mitigate bias. Both methods employed—excluding protected features and data augmentation—demonstrated their respective strengths in enhancing fairness metrics without substantially compromising accuracy. This observation underscores the potential for thoughtful data manipulation strategies to produce more equitable outcomes in automated decision-making processes.

It is crucial for practitioners to weigh these approaches carefully, considering both the ethical implications and the practical outcomes of their models. As machine learning continues to play an integral role in decision-making, the pursuit of both fair and accurate models remains a paramount challenge and necessity. For future work, we can try the following methods:

Continuous Monitoring: Regular audits of recruitment AI systems to ensure they do not develop or perpetuate biases.

Broader Data Consideration: Expanding the dataset to include more diverse demographics and intersectional analysis to cover a wider range of potential biases.

Legal and Ethical Compliance: Ensuring that AI-driven recruitment tools comply with evolving legal standards and ethical norms related to employment and anti-discrimination laws.

REFERENCE

All code above can be checked here.

1. Kumar, D., Grosz, T., Rekabsaz, N., Greif, E., & Schedl, M. (2023). Fairness of recommender systems in the recruitment domain: An analysis from technical and legal perspectives. *Frontiers in Big Data, 6*. `https://doi.org/10.3389/fdata.2023.1245198`

2. Fabris, A., Baranowska, N., Dennis, M. J., Hacker, P., Saldivar, J., Borgesius, F. Z., & Biega, A. J. (2023, September 25). Fairness and bias in algorithmic hiring. arXiv.org. `https://arxiv.org/abs/2309.13933`

3. Kubiak, E., Efremova, M. I., Baron, S., & Frasca, K. J. (2023). Gender equity in hiring: Examining the effectiveness of a personality-based algorithm. *Frontiers in Psychology, 14*. `https://doi.org/10.3389/fpsyg.2023.1219865`

4. Employment, fairness at work, and Enterprise. GOV.UK. (n.d.). https://www.gov.uk/government/publications/the-report-of-the-commission-on-race-and-ethnic-disparities/employment-fairness-at-work-and-enterprise

5. Chen, Z. (2023, September 13). Ethics and discrimination in artificial intelligence-enabled recruitment practices. Nature News. https://www.nature.com/articles/s41599-023-02079-x

6. Guardian News and Media. (2022, May 11). Finding it hard to get a new job? robot recruiters might be to blame. The Guardian. https://www.theguardian.com/us-news/2022/may/11/artitifical-intelligence-job-applications-screen-robot-recruiters

7. Institute, I. (2022, October 5). Utrecht Fairness Recruitment Dataset. Kaggle. https://www.kaggle.com/datasets/ictinstitute/utrecht-fairness-recruitment-dataset/data

8. University of Pittsburgh gender inequality research lab. (n.d.). https://www.girl.pitt.edu/gen-pacs-data

9. AyushTankha. (2023, July 10). 70K+ job applicants data (human resource). Kaggle. https://www.kaggle.com/datasets/ayushtankha/70k-job-applicants-data-human-resource