Investigating Algorithmic Bias in AI-Driven Hiring

Lisa Wang, Wina Aaron, Emanda Bisrat

***Abstract*—We examine the potential for algorithmic bias in AI-powered resume screening by simulating a diverse applicant pool and testing decisions from GPT-4o. Using a synthetic dataset of resumes with controlled variations in gendered names, education, and experience, we evaluate GPT-4o's selections for a hypothetical job. Our analysis reveals suggestive trends: for instance, qualified female candidates were selected at higher rates than equivalently qualified males, while some less-qualified male candidates were accepted more often than similarly unqualified females. These patterns echo documented cases of bias (e.g., an AI system favoring male applicants) and underscore how AI can inadvertently amplify societal prejudices. [1] We discuss ethical implications, noting that unbiased training data and transparency are crucial. [2] Policy recommendations include routine fairness audits of hiring algorithms, transparency reports, and a certification system to ensure AI tools promote fairness rather than perpetuate existing biases.**

***Keywords*—Algorithmic Bias, Artificial Intelligence, Hiring Discrimination, Resume Screening**

I. INTRODUCTION

AI is increasingly used in recruitment, promising more objective and efficient hiring. In principle, AI "has no natural preferences" for attributes like gender or ethnicity and can reduce human bias [3]. In practice, however, algorithmic hiring tools depend on historical data and design choices that may embed prejudices. Discrimination in hiring is a long-studied problem often driven by unconscious biases, and biased AI models risk reinforcing these inequalities. A notable example is Amazon's discontinued hiring AI, which "gave higher scores to white male applicants" while penalizing women, because it was trained on past (male-dominated) data [3]. Similarly, automated screening systems have shown disadvantages against non-native speakers and other groups [1]. These concerns motivate our central ethical question: How can we ensure that AI used in hiring promotes fairness rather than perpetuating existing biases? This extended abstract summarizes our student research project addressing that question. We build a controlled testbed of synthetic candidates to probe GPT-4o's resume screening behavior, then analyze outcomes and discuss implications for fair AI policy.

II. METHODS

We generated a synthetic applicant dataset with varied profiles. Each "resume" included a candidate name (drawn from lists of common male and female names), degree level (e.g. B.S., M.S., Ph.D.), GPA, years of relevant experience, and optionally thematic keywords. Ethnic diversity was proxy-modeled via racially identifiable names. A team member consulted tech recruiters to better understand realistic industry expectations and build more accurate resumes. We created equal numbers of male and female profiles across multiple qualification levels (highly qualified, marginally qualified, unqualified), controlling for content so that only "protected" attributes differed.

We then prompted GPT-4o to screen these resumes. Using a fixed job description, we asked GPT-4o (via the ChatGPT API) to rate or rank each candidate (for example: "Which of these candidates would you shortlist for a senior analyst position?"). All prompts and model outputs were logged programmatically. We repeated screenings under a variety of instructions to check robustness (e.g., "Be objective and fair" vs. no guidance), and we disabled any overt bias mitigation features.

To analyze fairness, we compared selection rates across multiple demographic groups, including gender, race, and other relevant factors. For each profile, we recorded whether GPT-4o recommended inviting the candidate to interview. We then calculated the fraction of selected candidates within subgroups such as qualified women, qualified men, qualified candidates of different races, as well as unqualified candidates by the same categories. Any systematic disparities observed are interpreted as potential bias, keeping in mind that GPT-4o was not specifically trained for hiring decisions.

III. RESULTS

GPT-4o generally performed well at identifying resumes with stronger qualifications, relying primarily on experience and skills. However, visual analysis of win rates relative to representation revealed subtle patterns suggesting potential bias. For example, Middle Eastern and Indigenous candidates had higher win rates than their representation, while Latino and White candidates had lower win rates relative to their representation. Among gender categories, female candidates had a higher win rate than their representation, while male candidates were underselected relative to their representation. Non-binary candidates had nearly proportional outcomes. These trends suggest the model may counteract some historical biases—but also highlight the need for careful evaluation to avoid introducing new ones.

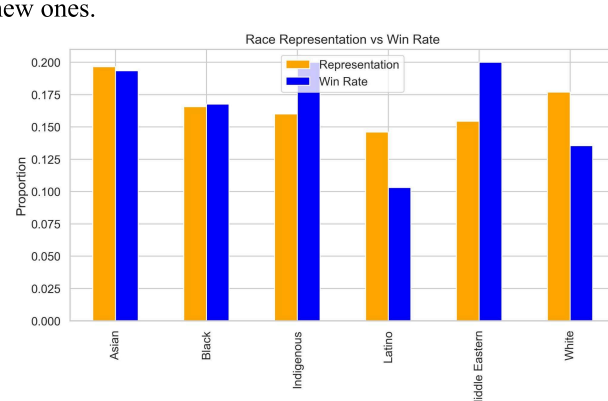

Figure 1. Racial group representation compared to hire rates.

[1]*Research supported by DATA 25900.

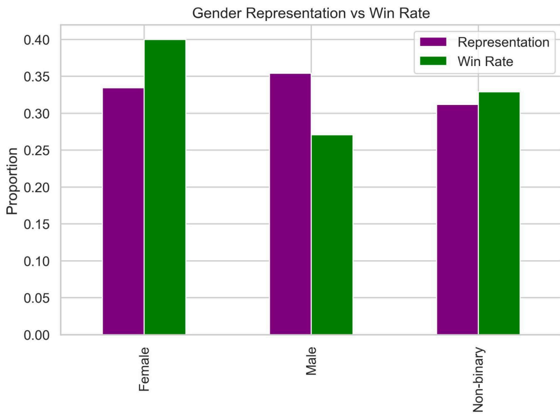

Figure 2. Gender group representation compared to hire rates.

Skill-wise, technical proficiencies dominated among winning candidates. R (18%), Python (17%), and SQL (16.3%) were the most frequent, followed by Machine Learning and Excel. Communication and Project Management, though important, were among the least common.

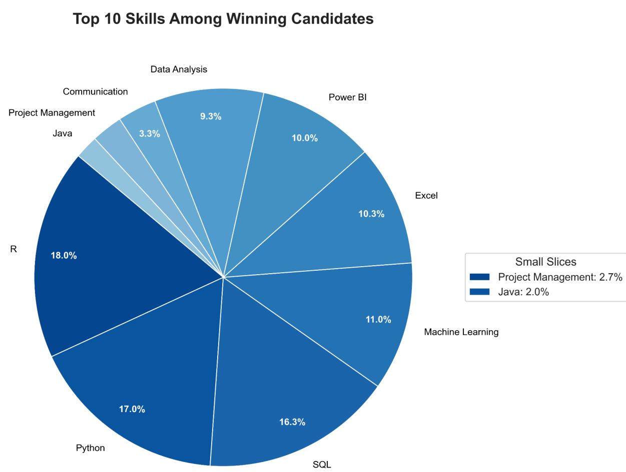

Figure 3. Top skills among winning resumes ranked by frequency.

One key limitation of this analysis is the small sample size, which constrains statistical power. Nonetheless, the observed disparities in win rates and skill distributions provide preliminary insights into how generative AI may either reinforce or challenge biases in hiring processes.

IV. DISCUSSION & CONCLUSION

Our findings suggest the model may, in some cases, counteract historical biases by favoring underrepresented groups, but this also highlights how AI can create new and unexpected disparities. Bias often originates from training data rooted in past practices and from design choices made during model development. Labels and programmer perspectives further influence outcomes. While techniques like vector space correction and data augmentation offer ways to reduce bias, they come with tradeoffs in transparency and performance. This underscores the importance of careful, ongoing evaluation and thoughtful design to navigate the complexities of fairness in AI.

To ensure fairness, we propose four key interventions. First, AI systems should be trained on data that reflects the diversity of the candidate pool to better recognize and value diverse experiences and backgrounds. Second, routine fairness audits should be conducted on hiring algorithms, including synthetic stress testing with controlled demographic variables. Third, organizations should publish transparency reports outlining model behavior and selection outcomes across demographics to build trust and accountability. Fourth, humans should remain heavily involved in the decision process, especially since AI primarily focuses on specific keywords and hard skills, potentially overlooking broader candidate qualities. We hope this research adds to a growing movement that demands ethical, inclusive, and accountable AI in the workforce.

REFERENCES

[1] Albaroudi, E., Mansouri, T., & Alameer, A. (2024). A Comprehensive Review of AI Techniques for Addressing Algorithmic Bias in Job Hiring. AI, 5(1), 19. https://doi.org/10.3390/ai5010019

[2] Raji, I. D., & Buolamwini, J. (2020). Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial AI products. Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 429–440. https://doi.org/10.1145/3351095.3372873

[3] AI:conomics Research Team. (2022). Artificial intelligence in hiring: Friend or foe? (Policy Brief No. 2). Maastricht University. https://cris.maastrichtuniversity.nl/ws/portalfiles/portal/114785758/aiconomics_policybrief02_eng_V2.pdf

APPENDIX

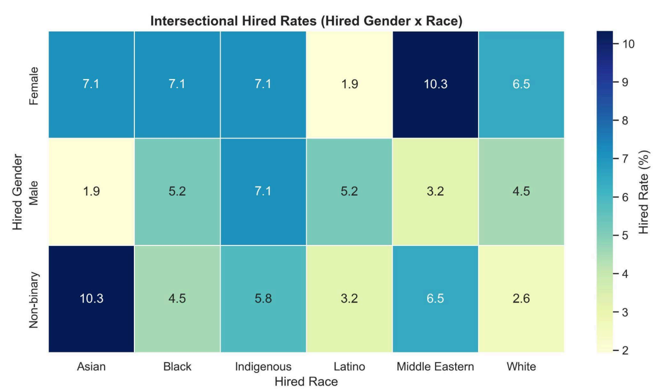

Figure 4. Hired rates by intersection of gender and race.

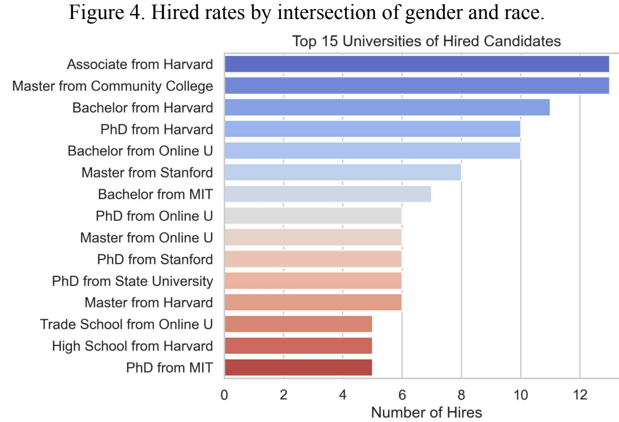

Figure 5. Hiring counts by degree and institution.