



New Study Shows AI Resume Screeners Prefer White Male Candidates: Your 5-Step Blueprint to Prevent AI Discrimination in Hiring

Insights

11.11.24

A new study reveals that popular AI-based resume screening tools often favor White and male candidates, showing that resumes tied to White-associated names were preferred 85% of the time. The University of Washington Information School study found that AI resume screening may inadvertently replicate – or even amplify – existing biases against certain demographics, and the implications are clear: without targeted safeguards, AI tools may lead to discriminatory hiring practices, even if inadvertent. This Insight breaks down the study's findings, explains why they matter for HR professionals, in-house counsel, and employers, and provides five practical steps you can take to avoid AI bias at your workplace.

Shocking Results From AI Hiring Study

The University of Washington study, presented at the October Association for the Advancement of Artificial Intelligence/Association for Computer Machinery (AAAI/ACM) Conference on AI, Ethics, and Society, highlights the biases present in several leading open-source AI resume-screening models.

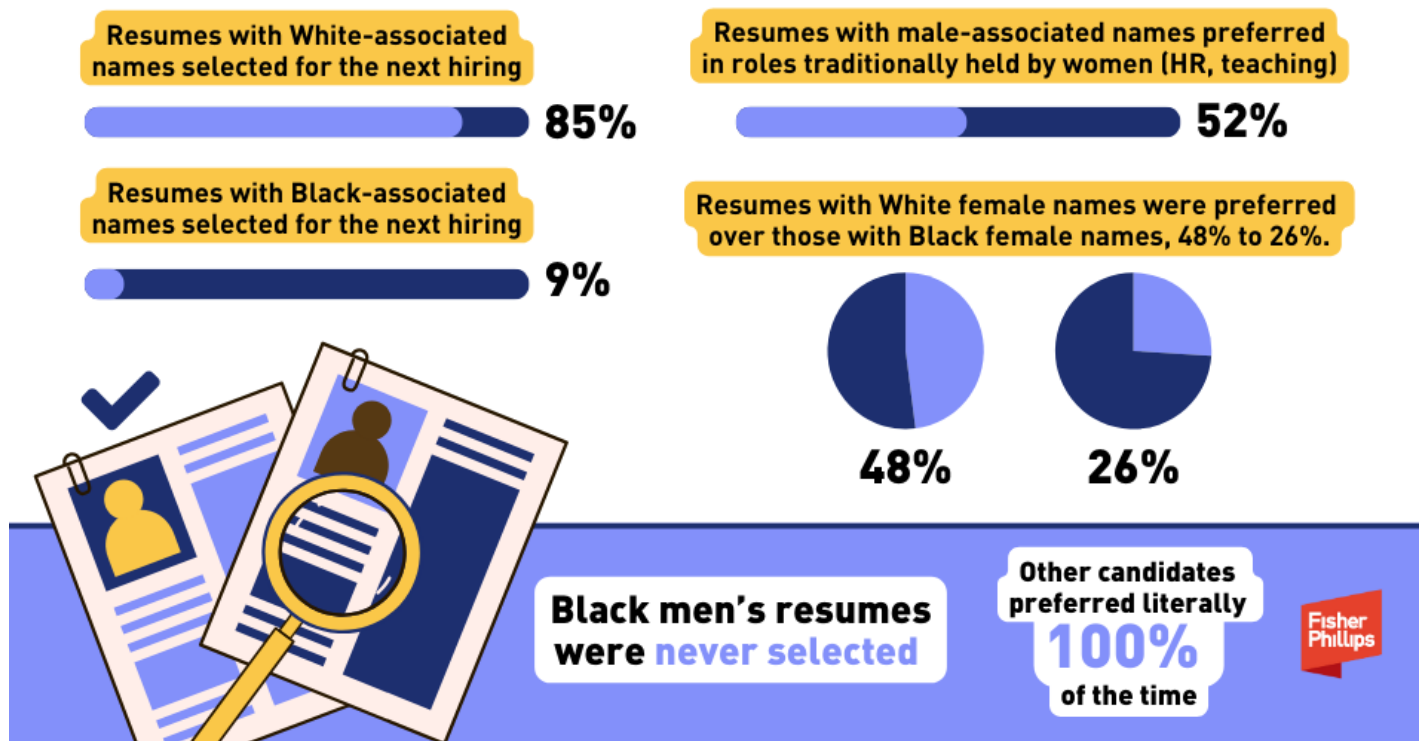
Researchers Kyra Wilson and Aylin Caliskan audited three popular large language models (systems trained on a general corpus of language data rather than data specific to hiring tasks) using 554 resumes and 571 job descriptions, with over 3 million combinations analyzed across different names and roles. They then changed up the resumes, swapping in 120 first names generally associated with individuals who are male (such as John, Joe, Daniel), female (Katie, Rebecca, Kristine, etc.), Black (Kenya, Latrice, Demetrius, etc.), or White (Hope, Stacy, Spencer, etc.). They then submitted the resumes for a series of positions ranging from chief executive to sales workers.

The results were stark:

- Resumes with White-associated names were selected 85% of the time for the next hiring step, while resumes with Black-associated names were only preferred 9% of the time.
- Resumes with male-associated names were preferred 52% of the time, even for roles with a traditionally high representation of women – like HR positions (77% women) and secondary school teachers (57% women).

- Resumes with White female names were chosen over those with Black female names, by a margin of 48% to 26%.
- Black men faced the greatest disadvantage, with their resumes being overlooked 100% of the time in favor of other candidates.

STARTLING RESULTS FROM AI HIRING STUDY



“Garbage In, Garbage Out”

Researchers attribute these biased outcomes to the data used to train the AI models. AI systems inherently mirror the patterns present in their training datasets. When these datasets are drawn from sources with historical or societal inequities, the AI system is likely to replicate or even amplify those inequities, leading to biased decision-making. For instance, if past hiring data or generalized language data is skewed to favor White or male candidates due to longstanding social privileges, the AI model will absorb and perpetuate this bias in its recommendations, leading to unfair preferences for candidates from those groups.

Moreover, because the underlying training data often incorporates a wide array of data from sources that may include biased language or assumptions, AI models trained on such data inherit biases that may not be immediately visible but become evident in their outputs. This highlights the vulnerability of AI use in applications, like hiring, where fairness is critical.

When we feed in “garbage” – data influenced by historic inequalities – we get “garbage” out, meaning decisions that are inherently flawed or biased. Consequently, AI tools that lack sufficient

meaningful decisions that are inherently flawed or biased. Consequently, AI tools that lack sufficient attention to diversity and equity in their training data risk becoming automated gatekeepers of discrimination, systematically disadvantaging qualified candidates from underrepresented groups.

Ultimately, this “garbage in, garbage out” effect underscores the need for companies to scrutinize both the data used in AI systems and the models themselves. Without interventions to correct these biases, companies risk reinforcing the very inequalities that AI was initially meant to help eliminate.

Here’s Why Employers Should Care

The stakes are high for employers adopting AI when hiring. While the technology promises efficiency, the potential for biased outcomes carries significant legal, ethical, and reputational risks.

- Discriminatory hiring practices can lead to **costly lawsuits or government investigations** into your workplace practices.
- The EEOC and other agencies have already said that **existing laws cover so-called AI discrimination**, so you shouldn’t stick your head in the sand waiting for new protections to be enshrined into the law before acting.
- Even **unintentional discrimination** can lead to an adverse finding, so you won’t be able to fully defend your organization by saying you didn’t mean to discriminate.
- The government has also already indicated that employers can’t simply point the finger at their **AI vendors and hope to deflect responsibility**. If your technology ends up discriminating against applicants or workers, agencies have said that your company could be on the hook.
- By screening out candidates who don’t fall into the White male category, you will lose out on an **array of diverse, qualified talent** that can strengthen your organization.
- The **negative publicity** that could result from accusations of AI bias could be costly. Your recruitment and retention efforts could take a real hit, as could your reputation among clients and customers, if you are outed as an employer that uses AI technology that discriminates against applicants.

5 Best Practices Checklist to Avoid AI Discrimination in Hiring

As the study notes, simply removing names from resumes may not address the issue. Resumes contain various elements, such as educational institutions, locations, group memberships, and even language choices (e.g., women often use terms like “cared” or “volunteered,” while men might use “repaired” or “competed”) that can indicate an individual’s race or gender, potentially perpetuating biased outcomes.

The good news is that there are some simple steps you can take to minimize the chance of AI bias at your workplace. Here are five best practices you should consider implementing at your organization.

1. Regularly Audit AI Models for Bias

Conduct regular audits of AI tools to detect racial, gender, and intersectional biases. Collaborate with third-party evaluators or leverage in-house resources to regularly examine screening outcomes and ensure fairness in the hiring process.

2. Consider Data Transparency and Fairness

Ensure that the data used to train your models is balanced and representative, and prioritize models with built-in transparency. Ask vendors the right questions so they can demonstrate their clear efforts to minimize biases in training datasets.

3. Avoid Over-Reliance on Automated Decisions

“Automation bias” is a real cognitive bias which can lead people to trust AI decisions more than human judgment. This can reinforce biased outcomes when hiring, as your managers may rely heavily on AI-driven recommendations, believing them to be “better” than anything they could do on their own. Address these risks head on. Make sure to integrate – and mandate – human oversight into AI decisions. Train your hiring managers to use AI tools as supplementary rather than primary decision-makers.

4. Adopt Inclusive Job Descriptions and Screening Criteria

Ensure your job descriptions and selection criteria are neutral and inclusive. Remove unnecessary criteria that may unfairly disadvantage certain candidates, such as rigid experience or narrowly defined skills that may unfairly disadvantage certain applicants. Remember, if you input these unintentional biases into your AI systems, they will spit biased results right back at you.

5. Implement Data-Driven Safeguards

Regularly analyze the outcomes of your AI screening tools, comparing demographic results to identify and address any biases. Use this data to make necessary adjustments to screening algorithms or criteria.

5 Best Practices Checklist to Avoid AI Discrimination in Hiring

1

REGULARLY AUDIT AI MODELS FOR BIAS

Check for racial, gender, and intersectional

biases to ensure fairness in hiring.

2

CONSIDER DATA TRANSPARENCY AND FAIRNESS

Ask the right questions of vendors so they can demonstrate their efforts to minimize biases.

3

AVOID OVER-RELIANCE ON AUTOMATED DECISIONS

Mandate human oversight, training your hiring managers to use AI tools as supplementary decision-makers.

4

ADOPT INCLUSIVE JOB DESCRIPTIONS AND SCREENING CRITERIA

Ensure your job descriptions and selection criteria are neutral and inclusive.

5

IMPLEMENT DATA-DRIVEN SAFEGUARDS

Regularly analyze AI outcomes, comparing demographic results to identify and address biases.

Fisher
Phillips

Conclusion

We'll continue to monitor developments in this area and provide the most up-to-date information directly to your inbox, so make sure you are subscribed to [Fisher Phillips' Insight System](#). If you have

questions, contact your Fisher Phillips attorney, the authors of this Insight, or any attorney in our [AI, Data, and Analytics](#) Practice Group.

Related People



Amanda M. Blair

Associate

212.899.9989

[Email](#)



Karen L. Odash

Associate

610.230.2165

[Email](#)

Service Focus

Employment Discrimination and Harassment

AI, Data, and Analytics

Trending

AI Governance Hub