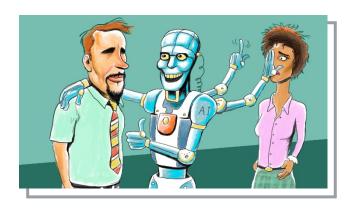
CASE STUDY

"The math-powered applications powering the data economy were based on choices made by fallible human beings. Some of these choices were no doubt made with the best intentions."— Cathy O'Neil

A.I. & Fairness: Beyond Blind Spot Bias?

Bias is the enemy of fairness. Humans tend to exhibit an in-group/out-group bias that causes them to—consciously or unconsciously—favor people who are like them over people who are not. This can cause unfair discrimination on many grounds (e.g., race, sex, age, religion, etc.) in many arenas (e.g., employment, housing, medical care, criminal justice, etc.). Many people and companies have turned to artificial intelligence (AI) tools in an attempt to improve the speed, efficiency, and objectivity of decision-making in these and other realms by replacing potentially prejudiced human judgment with unbiased machine judgment.

For example, sometime around 2014, Amazon began assembling computer models to review job applicants' resumes. Unfortunately, the company soon figured out that its new program was infected with a substantial gender bias. This probably should have been foreseeable. Amazon trained its models on resumes submitted to the company over a 10-year period and, unsurprisingly, the large majority of those applications came from men who held roughly 70% of the tech



jobs in the industry over that period. The model "taught itself that male candidates were preferable. It penalized resumes that included the word 'women's,' as in 'women's chess club captain.' And it downgraded graduates of two all-women's colleges." (Dastin). Amazon tried to fix it by making the program neutral to these terms. However, because Amazon couldn't ensure that the program would not figure out another way to similarly discriminate, it abandoned the project.

Google had a similar experience. Its engineers tried to teach an AI model what a "successful" candidate for a company tech job looked like by training it on data from Google's previous hiring decisions. Because that body of workers consisted disproportionately of male graduates of





CASE STUDY

highly prestigious universities, an "invisible bias was baked into the system from day one," (Hyer) replicating past mistakes. While Amazon suffered training data bias, Google's system was derailed by algorithmic bias. Like Amazon, Google abandoned this AI experiment.

And it's not just Amazon and Google that have been plagued with bias in developing and training AI technologies. Miranda Bogen of the Center for Democracy and Technology reports that "most hiring algorithms will drift toward bias by default." Furthermore, she says:

To attract applicants, many employers use algorithmic ad platforms and job boards to reach the most "relevant" job seekers. These systems, which promise employers more efficient use of recruitment budgets, are often ...predict[ing] not who will be successful in the role, but who is most likely to click on that job ad.

Unfortunately, research seems to indicate that generative AI, as embodied in ChatGPT and other new tools, is also plagued by racial and gender bias. One study tested three popular generative AI tools, including DALL E 2, and found that their "evident gender and racial biases ... were even more pronounced than the status quo...." (Zhou et al).

Many academics and recruiting firms are studying the reasons why AI has not lived up to its headlines by dramatically decreasing hiring bias. But so far, no solutions have emerged to eliminate it.

Case study written by:

Robert Prentice, J.D.

Department of Business, Government and Society McCombs School of Business The University of Texas at Austin





CASE STUDY



Discussion Questions

- 1. Do you think that Amazon and Google acted in good faith? Why or why not?
- 2. Would it change your mind regarding Google's good faith to learn that a study conducted about the same time found that Google Ads a tool used by third parties to find jobs discovered that "fake Web users believed by Google to be male job seekers were much more likely than equivalent female job seekers to be shown a pair of ads for high-paying executive jobs when they later visited a news website." (Datta et al.) Explain.
- 3. Hyer, which offers an employment app to connect potential employers with potential employees, suggests that firms use AI to do the heavy lifting of processing resumes and identifying promising candidates, but require that any final decisions be made by humans who keep in mind the potential for bias. Does this sound like a reasonable approach to improving hiring practices with AI? Why or why not?
 - a. Would this work to improve diversity, especially in light of a recent study finding that recruiters follow AI recommendations 85% of the time? (Alexander). Or would giving humans the final say in a hiring decision simply reintroduce all the bias that AI was supposed to eliminate?
- 4. Another firm (JobsPikr) recommends not only human oversight of the decision-making process, but also (a) utilizing diverse data sets for training AI models, (b) regular auditing of algorithms for bias, (c) building transparent AI models so that users can understand why they are making the recommendations that they are offering, and (d) implementing blind recruitment techniques that anonymize candidates' race, gender, and the like. (Alexander). Do these seems like sensible steps to you? Would they be adequate? Explain your reasoning.





CASE STUDY

- 5. The Algorithmic Justice League's Joy Buolamwini encourages companies creating AI tools in the hiring space to increase transparency by opening up the "black box" of how AI models are created so that their algorithms, data, and results can be audited for accuracy. Does this sound like a feasible idea? Why or why not?
- 6. Ethicists point out that tinkering with an algorithm in order to reduce its bias often also reduces its accuracy. How are firms to choose how best to balance those two important features of any AI model?
- 7. One multinational study (Vlasceanu & Amodio) found that greater national-level gender inequality was associated with more male-dominated Google image search results. The study also found that such biased search outputs guided the formation of gender-biased prototypes and influenced hiring decisions, creating "a cycle of bias propagation between society, AI, and users." Is this worrisome? Why or why not? How might such a cycle of bias be reformed?
- 8. In 2019, Google rolled out a new AI tool it called Bert. When prompted to consider 100 words in English (like "baby," "horses," and "money") it associated 99 of them more with men than with women. Only "mom" was identified with women. Professor Emily Bender remarked: "Even the people building these systems don't understand how they are behaving." How can those who conceive, build, and market new AI models best guard against bias?
- 9. Have you heard about the AI programs that are very competent at identifying pictures of white males, but not nearly as accurate at identifying pictures of dark-skinned people, particularly women? Does this sound like a problem? Why or why not? How should the criminal legal system, which obviously would love an effective AI tool to identify criminals, react to this bias in facial recognition tools?





CASE STUDY

Sources

Ifeoman Ajunwa, "Beware of Automated Hiring," New York Times, Oct. 9, 2019.

John Alexander, "Reducing Bias in AI Recruitment and HR Systems—Strategies and Best Practices," Oct. 29, 2024 at https://www.jobspikr.com/report/reducing-bias-in-ai-recruitment-strategies/.

Miranda Bogen, "All the Ways Hiring Algorithms Can Introduce Bias," <u>Harvard Business Review</u>, May 6, 2019, at https://hbr.org/2019/05/all-the-ways-hiring-algorithms-can-introduce-bias.

Krista Bradford, "Google Shows Men Ads for Better Jobs," (Sept. 29, 2023), at https://tgsus.com/diversity/google-shows-men-ads-for-better-jobs/.

Jeffrey Dastin, "Amazon Scraps Secret AI Recruiting Tool that Showed Bias against Women," <u>Reuters</u>, Oct. 10, 2018, at https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MKoAG/.

Amit Datta et al., "Automated Experiments on Ad PrivacySettings: A Tale of Opacity, Choice, and Discrimination," <u>Proceedings on Privacy Enhancing Technologies</u> 2015(1): 92-112 (2015).

Emilio Ferrara, "Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, and Mitigation Strategies," <u>Sci</u> 6(1), at https://doi.org/10.3390/sci6010003.

Megan Garcia, "Racist in the Machine: The Disturbing Implications of Algorithmic Bias," World Policy Journal 33(4): 111 (Winter 2016).

Adi Gaskell, "How Biased Google Search Results Affect Hiring Decisions," <u>Forbes</u>, Sept. 6, 2022.

Kimberly Houser, "Can AI Solve the Diversity Problem in the Tech Industry? Mitigating Noise and Bias in Employment Decision-Making," <u>Stanford Technology Law Review</u> 22:290 (2019)





LEADERSHIP

CASE STUDY

Hyer, "When Google's AI Hiring Tool Turned into a Diversity Disaster—And What HR Can Learn Today" (Oct. 13, 2024), at https://hyer.sg/when-googles-ai-hiring-tool-turned-into-a-diversity-disaster/.

Orly Lobel, The Equality Machine (2022).

Cade Metz, "Google Scraps AI Tool That Fosters Hiring Bias." New York Times, 2018. https://www.nytimes.com/2018/10/09/technology/ai-hiring-tool-bias.html

Cade Metz, "We Teach A.I. Systems Everything, Including Our Biases," <u>New York Times</u>, Nov. 11, 2019.

Cathy O'Neil, <u>Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy</u> (2017).

Natasha Singer, "Amazon is Pushing Facial Technology that a Study Says Could Be Biased," New York Times, Jan. 24, 2019.

Noah Smith, "Why AI-Driven Hiring Hasn't Delivered on Its Promise Yet." *Forbes*, 2020. https://www.forbes.com/sites/noahsmith/2020/01/14/ai-hiring-bias-problems/

Nitasha Tiku, "Google's AI Hiring Tool Failed to Live Up to Its Promise." *Wired*, 2018. https://www.wired.com/story/google-ai-hiring-tool-failed/

Madalina Vlasceanu & David Amodio, "Propagation of Societal Gender Inequality by Internet Search Algorithms," PNAS, Vol 119, No. 29 e2204529119 (2022), at https://www.pnas.org/doi/10.1073/pnas.2204529119.

Mi Zhou et al., "Bias in Generative AI," (2024), at https://arxiv.org/abs/2403.02726.

Algorithmic Bias: https://ethicsunwrapped.utexas.edu/glossary/algorithmic-bias

In-Group/Out-Group Bias: https://ethicsunwrapped.utexas.edu/glossary/in-group-out-group.