UCLA
UCLA Previously Published Works

Title

Artificial Intelligence and Oppression: Addressing Bias in Algorithmic Decision-Making

Permalink

https://escholarship.org/uc/item/9tk031zx

Author

Cesar, Dylan James

Publication Date

2025-03-01

eScholarship.org Powered by the California Digital Library
University of California

Artificial Intelligence and Oppression: Addressing Bias in Algorithmic Decision-Making

Cesar, Dylan J.[1]

[1]*Department of Civil and Environmental Engineering, University of California, Los Angeles, Los Angeles, California, 90095 USA*
**dy1anjc4747@g.ucla.edu*

Abstract: The integration of artificial intelligence into society offers multiple benefits to many groups of people. However, biases in training data and human inputs can reinforce existing systems of oppression, leaving certain groups marginalized. These systems, deeply rooted into society, also shape AI itself by introducing and perpetuating bias within its technologies. This paper examines how these biases influence AI's impact on oppressed groups, exposing critical flaws that often go unnoticed by the large majority.

INTRODUCTION

Artificial intelligence has advanced significantly in recent decades, but the concept of AI dates back thousands of years, with stories of artificial beings possessing human-like consciousness. The foundation of modern AI arose in the 1950s with the development of Logic Theorist by Newell, Shaw, and Simon. At the time, this was a groundbreaking program capable of solving mathematical theorems using symbolic reasoning (Reiff). Following the launch of Logic Theorist, advancements in deep learning were driven by increasing computer availability and funding from DARPA (Reiff). Still, early AI systems have limitations. Companies such as IBM and Microsoft later contributed to the revolution of AI by developing chess playing programs and speech recognition software. All these inventions paved the way for today's AI driven world. Today, AI technologies, ranging from chatbots like ChatGPT to machine learning algorithms and virtual assistants such as Apple's Siri, Amazon's Alexa, Google's Cortana, have become deeply integrated into everyday life.

AI's societal benefits are undeniable. It has the potential to drive economic growth and revolutionize industries such as education, housing, employment, and credit (Ajanaku). Ideally, AI should also contribute to formulating a more inclusive society. As technology continues to advance, improvements across various fields are expected. However, despite these advancements, AI falls short of achieving true inclusivity.

The dominant assumption is that AI systems are neutral and objective, designed to operate fairly without discrimination. Surely the developers of these technologies put a lot of thought into developing a system that is 'fair', or what reason would most individuals believe this to be the truth? A closer examination reveals a different reality, highlighting a system that reinforces existing biases and inequalities. Olga Akselrod, a racial justice activist, states, "rather than help eliminate discriminatory practices, AI has worsened them" (Akselrod). This is evident in fields such as policing, surveillance, healthcare, facial recognition, and other AI-driven systems, where these technologies have exhibited patterns of racial bias and discrimination against certain groups of people.

Diving deeper into AI's development exposes systemic biases against marginalized communities, hindering the technology's ability to represent society accurately. The group of focus for this examination will be women of color, specifically black women. This work is in partial fulfillment of the ENGR184 course using the blueprint curriculum in Ref.[1,2] and captured in Volume 3 [Carbajo]"

METHODS

The goal of this paper is to examine how AI technologies impact marginalized communities through biases in training data and human inputs that uphold various systems of oppression. A deeper investigation into various fields will expose how AI fails to equally represent all groups of people. Addressing these flaws is crucial for fostering a more just society. This analysis draws from Black feminist thought, focusing on how intersecting systems of oppression, such as race, gender, sexuality, and class, affect marginalized communities, including Black Women. Additionally, lived experiences and standpoint theory will be utilized to critically assess biases embedded in AI and their broader social implications.

RESULTS AND INTERPETATION

Artificial intelligence is trained on vast datasets and uses machine learning to identify patterns and correlations. The process, known as algorithmic decision-making, enables AI to make inferences and select the 'best options' far more quickly and efficiently than a human. However, significant pitfalls arise, particularly in predictive policing. Simply put, this is a strategy employed by law enforcement agencies that use AI to forecast potential crimes and their locations based on historical data. Unfortunately, this data is deeply flawed, as police have historically over-policed communities with lower socioeconomic status, leading to their disproportionate representation in criminal records. Systemic racism embedded within the police power structure introduces substantial bias into these datasets, which AI then absorbs and perpetuates. As Ashwini K.P., UN Special Rapporteur on contemporary forms of racism, racial discrimination, xenophobia and related intolerance, explains "when officers in overpoliced [neighborhoods] record new [offenses], a feedback loop is created, whereby the algorithm generates increasingly biased predictions targeting these [neighborhoods]" (*Racism and AI: "Bias from the past leads to bias in the future*). As a result, predictive policing not only reflects past biases but also reinforces and magnifies them, ensuring that the same communities remain over-surveilled and criminalized. This cyclical discrimination places a lasting burden on communities who have long been unjustly targeted, making it clear that predictive policing does not promote justice but instead upholds existing systems of oppression.

Black trans women face similar struggles due to deeply ingrained biases in policing. Gender-based policing patterns, rooted in stereotypes, disproportionately subject women and LGBTQ individuals to harassment by law enforcement (*Understanding Gender-Biased Policing*). Black trans women, at the intersection of multiple forms of discrimination (race, gender, and/or sexuality), experience compounded injustices within the legal system. Black individuals are stopped and overpoliced at significantly higher rates than their white counterparts, while LGBTQ individuals, especially those who are transgender, have historically been targeted and arrested at disproportionate rates (Mahowald). The systematic policing biases not only harm the current generation of black trans women but also set a precedent for future generations. As data from these discriminatory practices feed into predictive AI, the cycle of injustice perpetuates itself. The same communities are continuously targeted, reinforcing and amplifying existing prejudices. How can society break this cycle and implement meaningful reform? The persistence of these patterns demands urgent action, as true justice cannot exist within a system that repeatedly criminalizes the most vulnerable.

Facial recognition software is an innovative technology that enhances security and accessibility across various applications. Unfortunately, these systems frequently misidentify individuals with darker skin tones, particularly women. Consider the cases of Oprah Winfrey, Michelle Obama, and Serena Williams, all of whom were incorrectly classified as male by AI systems developed by IBM, Microsoft, and Amazon (Buolamwini). Studies have shown that these same systems exhibit an error rate of 34.7% for darker-skinned women, compared to just 0.8% for lighter-skinned men (Birhane). Ideally, facial recognition should accurately identify prominent female figures regardless of skin tone. However, technology reflects a reality skewed by bias. A deeper analysis of the data sets used to train these systems uncovers a troubling truth:

one unspecified dataset revealed that faces 75% of the faces collected for facial recognition belonged to white men, while 80% were categorized as lighter-skinned individuals, leaving women of color underrepresented at less than 5% (Buolamwini). How is it that facial recognition software struggles to identify all human beings equally? The answer lies in the complex interplay of social power structures, systemic bias, and the lack of diversity in training data, which are all issues that demand deeper scrutiny.

The healthcare field is arguably the most vital profession in society. Millions rely on modern medicine and technology not only to save lives but also to aid recovery, improve quality of life, and bring new life into this world. Yet, beneath its advancements lie deep-rooted social injustices that have long gone unnoticed. Medical technology is at the forefront of innovation and is expected to function effectively for all patients. Still, critical flaws continue to slip through the cracks, disproportionately affecting marginalized groups. Consider the pulse oximeter, a device designed to measure oxygen levels and pulse rates noninvasively. While it performs accurately on individuals with lighter skin, studies have shown that it frequently misreads blood oxygen levels in darker-skinned individuals (Andrews), leading to potential misdiagnoses and inadequate treatment. This issue highlights a broader problem in how artificial intelligence and technology are integrated into society. Biases in training data. Just as facial recognition struggles to identify darker-skinned individuals accurately, medical devices like pulse oximeters can produce life-threatening errors due to built-in biases. Recognizing and addressing these flaws is crucial to ensuring that technological advancements promote equity rather than perpetuate discrimination.

AI technologies are at the forefront of modern hiring practices, with around 70% of companies, and an astounding 99% of Fortune 500 companies, integrating AI into their recruitment processes (Akselrod). From an efficient standpoint, this shift makes sense. AI can swiftly analyze vast numbers of resumes, easily identify top candidates, and even personalize job recommendations, streamlining an often overwhelming process for job seekers. Sadly, these technologies are not immune to bias; instead, they risk amplifying historical patterns of discrimination. Research from UPENN reveals that black professionals receive 30 to 50% fewer callbacks when their resumes include racial or ethnic identity (Zapata). If AI is meant to foster equal opportunity, why for some groups does it continue to perpetuate unfair hiring practices? One key issue is the training data. If an algorithm is trained primarily on resumes from white males, it may struggle to recognize qualified candidates from underrepresented groups, such as black females. Moreover, historical biases are embedded in hiring data, influencing AI's decision-making in ways that disadvantage marginalized communities. Aditya Malik of Forbes warns that "AI, if not vigilantly curated, might misconstrue these patterns as indicators of incompetence, thus exacerbating the exclusion of qualified candidates from underrepresented backgrounds" (Malik). Unfortunately, discrimination in hiring is not new. Race, gender, and class have long been used to exclude capable individuals, deeming them 'incompetent'. AI, rather than eliminating bias, risks reinforcing these inequities by learning from flawed data and replicating unjust hiring practices.

Large language models (LLMs) are advanced machine learning systems capable of understanding and generating human language. Popular LLMs, such as OpenAI's ChatGPT and Google's Gemini, have revolutionized the way we interact with AI, providing rapid responses to complex questions. By February 2025, ChatGPT had amassed 400 million active weekly users (Singh), while Google Gemini reached 275 million users by December 2024 (Kumar), reflecting the growing reliance on these technologies since their respective launches in November 2022 and May 2023. Beneath their impressive capabilities, however, these models are not free from bias. Systemic prejudices embedded in their training data can shape their outputs in ways that reinforce discrimination. A study by Yifan Yang, Xiaoyu Liu & Furong Huang at the University of Maryland examined this issue by analyzing ChatGPT-3.5 Turbo's responses to patient data. Researchers removed racial and ethnic identifiers from medical records and reassigned them before prompting AI for analysis. The findings revealed higher

predicted death rates for Black patients compared to white patients, longer hospital stays for white patients than for Black, Asian, and Latino patients, and more severe diagnoses of non-cancerous diseases in Black patients (Yang). These results underscore the risks of racial bias in AI-driven LLMs, demonstrating how even state-of-the-art LLMs can perpetuate inequities if not carefully designed and monitored.

CONCLUSIONS

A careful examination of AI's integration into society reveals its many benefits that make life more convenient and efficient. As technology advances, artificial intelligence will continue to push humanity to new heights. Nevertheless, it is crucial to recognize that not everyone benefits equally from these advancements. Since AI heavily relies on training data that is often biased towards certain groups, this results in many being marginalized and systematically oppressed. As discussed in this paper, this issue is apparent in the fields of policing, healthcare, facial recognition software, large language models, and others that were not examined. The key question, then, is how can AI be modified to enhance equity for all groups of people rather than upholding existing systems of oppression? While this is a complex and nuanced issue that cannot be entirely answered from this examination, there is at least one important step society must take to enhance AI. Training data must be ensured to be more representative of diverse populations. AI datasets need to include individuals of different racial and ethnic backgrounds, genders, sexualities, and socioeconomic statuses to create technology that serves all people equitably. Achieving this will require significant time, effort, and structural changes, but the benefits will contribute to a more just and inclusive society, creating a true technology of the future. While policy change that reviews datasets for equal representation may be the solution to the problem in the long-term, this just isn't feasible in society currently. For the time being, people must urge corporations to modify datasets, pushing for change through publishing papers and formulating groups to inspire change. The path forward will be challenging, but we will ultimately be grateful for embarking on this journey.

REFERENCES

1) Lee, Ethan, et al. "Education for a Future in Crisis: Developing a Humanities-Informed STEM Curriculum." *arXiv preprint arXiv:2311.06674* (2023).
2) Y. Sergio Carbajo, Nurturing Deeper Ways of Knowing in Science, Issues in Science & Technology, 2025, v. 41, n. 2, p. 71, doi. 10.58875/jkrw4525
3) Z. Carbajo, Sergio. "Queered Science & Technology Center: Volume 3." (2025).
4) Ajanaku, Dayo. "How Artificial Intelligence Impacts Marginalized Communities." *How Artificial Intelligence Impacts Marginalized Communities – The Network*, sites.law.berkeley.edu/thenetwork/2022/01/26/how-artificial-intelligence-impacts-marginalized-communities/. Accessed 11 Mar. 2025.
5) Akselrod, Olga, and Cody Venzke. "How Artificial Intelligence Might Prevent You from Getting Hired: ACLU." *American Civil Liberties Union*, 15 Nov. 2023, www.aclu.org/news/racial-justice/how-artificial-intelligence-might-prevent-you-from-getting-hired#:~:text=AI-based%20tools%20are%20used%20throughout%20hiring%20processes%2C%20increasing,impacting%20whether%20or%20not%20you%20got%20the%20job.
6) Akselrod, Olga. "How Artificial Intelligence Can Deepen Racial and Economic Inequities: ACLU." *American Civil Liberties Union*, 3 July 2023, www.aclu.org/news/privacy-technology/how-artificial-intelligence-can-deepen-racial-and-economic-inequities.
7) Andrews, Erika, and Jason Kane. "The Problem with Pulse Oximeters Your Doctor Probably Doesn't Know." *NBCNews.Com*, NBCUniversal News Group, 19 Jan.

2025, www.nbcnews.com/health/health-news/pulse-oximeters-black-patients-blood-oxygen-doctors-rcna183199.
8) Birhane, Abeba. “The Unseen Black Faces of AI Algorithms.” *Nature News*, Nature Publishing Group, 19 Oct. 2022, www.nature.com/articles/d41586-022-03050-7.
9) Buolamwini, Joy. “Artificial Intelligence Has a Racial and Gender Bias Problem.” *Time*, Time, 7 Feb. 2019, time.com/5520558/artificial-intelligence-racial-gender-bias/.
10) “Chatgpt.” *ChatGPT*, OpenAI, chatgpt.com/. Accessed 16 Mar. 2025.
11) Kumar, Naveen. “How Many People Use Gemini (2025).” *DemandSage*, 1 Jan. 2025, www.demandsage.com/google-gemini-statistics/.
12) Mahowald, Lindsey. “Black LGBTQ Individuals Experience Heightened Levels of Discrimination.” *Center for American Progress*, 22 Aug. 2022, www.americanprogress.org/article/black-lgbtq-individuals-experience-heightened-levels-discrimination/.
13) Malik, Aditya. “Ai Bias in Recruitment: Ethical Implications and Transparency.” *Forbes*, Forbes Magazine, 13 Aug. 2024, www.forbes.com/councils/forbestechcouncil/2023/09/25/ai-bias-in-recruitment-ethical-implications-and-transparency/.
14) “Racism and AI: ‘Bias from the Past Leads to Bias in the Future’ | OHCHR.” *Racism and AI: “Bias from the Past Leads to Bias in the Future,”* www.ohchr.org/en/stories/2024/07/racism-and-ai-bias-past-leads-bias-future. Accessed 11 Mar. 2025.
15) Reiff, Nathan. “A Brief History of Artificial Intelligence (AI): From Turing to IOT.” *Decrypt*, Decrypt, 28 Sept. 2023, decrypt.co/resources/a-brief-history-of-artificial-intelligence-ai-from-turing-to-iot.
16) Singh, Shubham. “Chatgpt Statistics (March 2025): Number of Users & Queries.” *DemandSage*, 28 Feb. 2025, www.demandsage.com/chatgpt-statistics/.
17) “Understanding Gender-Biased Policing.” *Understanding Gender-Biased Policing: Police Misconduct, Domestic Violence, and Sexual Assault*, American Civil Liberties Union, www.aclu.org/sites/default/files/field_document/aclu_memo-understanding_gender_biased_policing-2016_0.pdf. Accessed 12 Mar. 2025.
18) Yang, Yifan, et al. “Unmasking and Quantifying Racial Bias of Large Language Models in Medical Report Generation.” *Nature News*, Nature Publishing Group, 10 Sept. 2024, www.nature.com/articles/s43856-024-00601-z.
19) Zapata, Dawn. “New Study Finds Ai-Enabled Anti-Black Bias in Recruiting.” *Thomson Reuters Institute*, 2 Aug. 2021, www.thomsonreuters.com/en-us/posts/legal/ai-enabled-anti-black-bias/.