Exploring AI Bias and Racism: Ethical Considerations

Tahsin Bin Reza
ID: 210041106
BSc. in CSE, IUT
tahsinbinreza@iut-dhaka.edu

Ahmed Shafin Ruhan
ID: 210041116
BSc. in CSE, IUT
ahmedshafin@iut-dhaka.edu

MD. Rifat Sarwar
ID: 210041134
BSc. in CSE, IUT
rifatsarwar@iut-dhaka.edu

MD. Nazmus Sadiq
ID: 210041139
BSc. in CSE, IUT
nazmussadiq@iut-dhaka.edu

***Abstract*—This term paper explores AI bias and racism, emphasizing ethical considerations. Examining algorithmic biases, developer roles, and real-world consequences, it proposes solutions such as diverse data collection, transparency, community engagement, and continuous monitoring.**

I. Introduction

In this exciting age of artificial intelligence (AI), technology's widespread impact is felt everywhere in our society. However, as decision-making processes are increasingly being made by algorithms, concerns about bias and discrimination have become an alarming issue. In this term paper, we're diving into the twisted world of AI racism where we'll be digging into how the biases caused in AI systems can lead to, and sometimes worsen, social inequalities. Sometimes people think of AI as a neutral tool, but it can unknowingly cause and extend the biases in the data it is trained on. As we explore the intersection of technology and ethics, it's important to dig into where AI racism comes from, how it shows up, and what kind of impact it has.

II. Motivation

Sometimes, programs that AI uses can unintentionally treat people unfairly, making existing inequalities even worse. This is more than just a technical mistake that affects real-life decisions and can harm certain groups of people. Biased datasets lead to the predictive technology also being biased. Furthermore, the deployment of the app before adequate testing based on social impact results in unintended consequences. In 2018, MIT student Joy Buolamwini wrote about her experience discovering that the facial recognition algorithms in her lab — used all over the world– couldn't detect Black faces. She even had to wear a white mask to get the computer to recognize her as a person. Similarly, users discovered in 2020 that Twitter's image-cropping tool constantly focused on white faces. AI robots trained on billions of images consistently identified women as "homemakers" and people of color as "criminals" or "janitors." When certain groups are systematically disadvantaged by biased algorithms, it limits opportunities for advancements in technology, maintaining a cycle of inequality. When individuals perceive that AI technologies discriminate against certain groups, it raises concerns about fairness and transparency, hindering the widespread acceptance and adoption of AI solutions across diverse populations.

A. Statistics

- One commercial tool had a 0.8% error rate for light-skinned males, but 34.7% error rate for dark-skinned females (Buolamwini & Gebru, 2018).
- COMPAS, an automated risk assessment tool used for criminal sentencing in several states, incorrectly labeled black defendants as future criminals at close to twice the rate as white defendants (Angwin et al., 2019).
- A healthcare algorithm responsible for 200 million people systemically prevented almost 30% of eligible black patients from receiving additional care by giving lower risk scores to black patients than white patients with equal diagnoses (Obermeyer et al., 2019).
- FinTech firms charged Latinx and African-American loan borrowers 7.9 and 3.6 basis points, respectively, more than equivalent White borrowers, costing a yearly extra $765 million in interest (Bartlett et al., 2019).

Therefore, we must understand and solve this problem so that AI aligns with the values of fairness and justice that we want in our increasingly automated world.

III. Proposed Methodology

Artificial Intelligence (AI) is not designed to be racist, but it can inadvertently perpetuate biases present in society. Mitigating bias in AI requires proactive measures across its lifecycle, including pre-design, design and development, and deployment stages.

A. Pre-design Stage

Addressing biases in the pre-design stage involves stakeholder engagement, strong evaluation processes, and representation-bias reduction. Diverse stakeholder involvement ensures comprehensive consideration of concerns and biases. Thorough evaluations with end-users and experts help identify and address biases early on. Representation-bias reduction entails diversifying datasets and considering broader societal contexts. Transparency and accountability mechanisms ensure clear decision-making processes, while continuous learning enables adaptation to evolving challenges.

B. Design and Development Stage

Engineers and scientists working on AI models must actively identify and correct biased algorithms. Performance op-

timization should not overlook fairness considerations. Techniques like counterfactual fairness can help mitigate bias by considering alternative scenarios. Creating a culture of questioning and challenging decisions during development is crucial. Standards for managing bias need to be established, including regular monitoring of AI systems to ensure fairness.

C. Deployment Stage

In the deployment stage, operators and decision-makers interact with AI systems, often leading to unintended biases. Poorly specified design decisions can affect statistical outcomes and human behavior. Bridging the gap between intended and actual uses is essential to ensure alignment with users' expertise and expectations. Techniques like counterfactual fairness and deployment monitoring help manage bias risks. Establishing standards for risk management tools is crucial for future activities.

IV. Conclusion

Mitigating bias in AI requires a multifaceted approach across its lifecycle stages. By addressing biases proactively and implementing fairness techniques, we can ensure AI systems align with societal values of fairness and justice.

References

[1] Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Proceedings of the 1st Conference on Fairness, Accountability and Transparency, 77–91. https://www.sciencepolicyjournal.org/article_1038126_jspg160205.html

[2] Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2019). Machine Bias. ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

[3] Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations. Science, 366(6464), 447–453. https://doi.org/10.1126/science.aax2342

[4] Bartlett, R., Morse, A., Stanton, R., & Wallace, N. (2019). Consumer-Lending Discrimination in the FinTech Era. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3359248

[5] Raikes, J. (2023, April 21). AI Can Be Racist. Let's Make Sure It Works For Everyone. Forbes. https://www.forbes.com/sites/jeffraikes/2023/04/21/ai-can-be-racist-lets-make-sure-it-works-for-everyone/?sh=2869b0fb2e40

[6] Buolamwini, J. (2018, June 21). Opinion — When It Comes to Gorillas, Google Photos Remains Blind. The New York Times. https://www.nytimes.com/2018/06/21/opinion/facial-analysis-technology-bias.html