

The Need for a Feminist Approach to Artificial Intelligence

Christo El Morr

School of Health Policy and Management, York University, Canada
elmorr@yorku.ca

Abstract (This Is the Abstract Head Style)

Artificial intelligence (AI) presents immense potential and significant challenges concerning algorithmic bias. This paper explores how feminist theory provides a critical lens for understanding and addressing algorithmic bias's root causes and impacts. The historical context of systemic discrimination reveals how power imbalances have shaped data collection and analysis, leading to biased datasets that perpetuate inequalities through AI systems. The "black box" problem further obscures these biases, amplifying discriminatory outcomes in various domains. Feminist interventions, particularly intersectional feminism, offer a framework for uncovering how algorithmic bias interacts with multiple forms of oppression. Feminist data science challenges traditional methodologies and advocates for transparency, accountability, and diversity in AI development. Critiques of techno-solutionism highlight the need for broader societal change alongside technical fixes. By embracing a feminist approach, we can envision and work toward a future where AI technology is used for social justice, inclusivity, and collective liberation.

Introduction

The rapid development of AI is changing our world in ways we cannot yet grasp. While AI presents incredible advantages, it also introduces a significant challenge – algorithmic bias. Algorithmic bias refers to the systematic discrimination in AI systems that arises from biased data or algorithms, which leads to AI systems often exacerbating social inequalities. Feminist theory (Delmar 2018), which examines power dynamics and systems of oppression, has the potential to provide insights into bias sources and effects. Specifically, Intersectional feminism explores how various forms of oppression—such as race, gender, and class—intersect to produce complex inequalities (Crenshaw 2021).

Historical Context and Power Dynamics

The origin of algorithmic bias can be found in centuries of systemic discrimination (Bohren et al. 2022). Throughout

history, power imbalances embedded in social structures (Royce 2022) have influenced the way data is collected and analyzed. Data is believed to be unbiased and impartial while it mirrors power dynamics (Miceli et al. 2022). The gathering, organization, and interpretation of data are fundamentally influenced by political factors and dominant ideologies. Hence, existing datasets can be biased, as they might reflect the viewpoints and preferences of those in power. Such biased datasets can form the basis for AI algorithms that perpetuate existing inequalities. By being trained on biased datasets, AI systems inherit and magnify past prejudices, creating a dangerous cycle where historical discrimination continues to affect present-day results.

The Black Box: Amplification of Inequalities

Many AI systems work like mysterious "black boxes," where we can't see how the algorithms and decision-making processes actually work (Chaudhary 2024). This lack of transparency makes it hard to find and fix any biases that are built into systems that are based on these algorithms. The hidden nature of some AI algorithms can hide unfair practices, making it hard to implement accountability measures and mitigate algorithmic bias (Smith 2021). Algorithmic bias makes existing inequalities even worse and causes unfair outcomes in different areas. Biased algorithms in hiring can keep gender or racial disparities going. In criminal justice, tools for assessing risk can unfairly target people from marginalized communities (Garrett & Rudin 2023). In healthcare, algorithms can mean unequal access to care or different treatment options (Smith 2021). And in finance, biased algorithms can make economic inequality worse (Cohen et al. 2023; Rai 2020).

Feminist Interventions and Critiques

Feminist theory presents crucial insights and criticisms in addressing algorithmic bias. Particularly, intersectional feminism offers a framework for comprehending how

various forms of oppression, including gender, race, class, and sexuality, intersect and compound within algorithmic systems (Carastathis 2014; Hunt 2018). By applying an intersectional perspective, we can expect to unveil the intricate ways in which algorithmic bias perpetuates and magnifies existing inequalities (Figueroa et al. 2021). Feminist data science can challenge conventional methods of data collection, analysis, and interpretation (D'ignazio & Klein 2023). It can advocate for increased transparency, accountability, and diversity. By adopting feminist principles, feminist data scientists can foster a more just and inclusive environment.

Critiques of Techno-Solutionism

The belief that we can solve complex social problems using technical solutions persists in academic environments (Sætra & Selinger 2023). Such “Techno-solutionism” is not a good approach to facing algorithmic bias as this narrow view ignores the fact that technology is not neutral and that addressing algorithmic bias requires broader societal changes (Marinelli 2023; Sætra & Selinger 2023). A feminist lens is important to face “Techno-solutionism” as it advocates for tackling the underlying factors of inequality, such as systemic discrimination and power disparities, instead of depending solely on technical solutions.

Envisioning Equitable Technological Futures

A feminist lens within the artificial intelligence landscape can help dismantle oppressive systems and establish a fair and just world. It will help prioritize inclusivity and diversity in AI development, guaranteeing that the viewpoints and backgrounds of marginalized communities are reflected throughout the process. If marginalized communities are involved in the AI-based systems’ design process, the resulting AI-based systems can be expected to be more equitable. In this regard, fostering transparency and accountability in constructing reliable AI systems is crucial. Finally, a feminist lens can help to envision different scenarios where AI is used for social justice and collective freedom (Figure 1). By imagining these scenarios, we can challenge prevailing narratives and power dynamics, and spark meaningful changes toward a liberating approach to technological innovation.



Figure 1. Steps to Incorporate a Feminist Lens AI Design

Conclusion

It is important to tackle algorithmic bias in AI. Feminist theory methods offer great insights into where this bias comes from and its effects, stressing the importance of addressing power imbalances, historical backgrounds, and systemic inequalities. By adopting a feminist framework for AI innovation, we can advance toward a tech ecosystem that actively works for equitable outcomes while remaining cognizant of the inherent tension between AI's potential for empowerment and oppression.

References

Bohren, J. A., Hull, P., & Imas, A. 2022. Systemic discrimination: Theory and measurement.

Carastathis, A. 2014. The concept of intersectionality in feminist theory. *Philosophy compass*, 9(5), 304-314.

Chaudhary, G. 2024. Unveiling the Black Box: Bringing Algorithmic Transparency to AI. *Masaryk University Journal of Law and Technology*, 18(1), 93-122.

Cohen, S. N., Snow, D., & Szpruch, L. 2023. Black-box model risk in finance.

Crenshaw, K. 2021. Demarginalizing the intersection of race and sex: a black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. *Droit et société*, 108, 465.

D'ignazio, C., & Klein, L. F. 2023. *Data feminism*. MIT press.

Delmar, R. 2018. What is feminism? In *Theorizing feminism* (pp. 5-28). Routledge.

Figueroa, C. A., Luo, T., Aguilera, A., & Lyles, C. R. 2021. The need for feminist intersectionality in digital health. *The Lancet Digital Health*, 3(8), e526-e533.

Garrett, B. L., & Rudin, C. 2023. The Right to a Glass Box: Rethinking the Use of Artificial Intelligence in Criminal Justice. *Cornell Law Review*, Forthcoming, Duke Law School Public Law & Legal Theory Series(2023-03).

Hunt, G. 2018. Intersectionality: Locating and critiquing internal structures of oppression within feminism. *Philosophy: Feminism*, 121-138.

Marinelli, R. (2023). *Technology and Sustainable Development: The Promise and Pitfalls of Techno-Solutionism*. In: JSTOR.

Miceli, M., Posada, J., & Yang, T. 2022. Studying up machine learning data: Why talk about bias when we mean power? *Proceedings of the ACM on Human-Computer Interaction*, 6(GROUP), 1-14.

Rai, A. 2020. Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science*, 48, 137-141.

Royce, E. 2022. *Poverty and power: The problem of structural inequality*. Rowman & Littlefield.

Sætra, H. S., & Selinger, E. 2023. The Siren Song of Technological Remedies for Social Problems: Defining, Demarcating, and Evaluating Techno-Fixes and Techno-Solutionism. *Demarcating, and Evaluating Techno-Fixes and Techno-Solutionism* (September 19, 2023).

Smith, H. 2021. Clinical AI: opacity, accountability, responsibility and liability. *Ai & Society*, 36(2), 535-545.