The Algorithmic Mirror: Does AI Create Bias in Medical Treatment?

Rasit Dinc

Rasit Dinc Digital Health & AI Research

Published: February 21, 2022 | AI Diagnostics

DOI: 10.5281/zenodo.17998018

Abstract

Artificial Intelligence AI is revolutionizing healthcare, offering powerful tools for diagnostics and personalized medicine. While promising a future of more...

The Promise and Peril of AI in Healthcare

Artificial Intelligence (AI) is revolutionizing healthcare, offering powerful tools for diagnostics and personalized medicine. While promising a future of more efficient and accurate care, a critical question arises: **Does AI create bias in medical treatment?** The answer is complex. AI does not invent bias; instead, it acts as an algorithmic mirror, reflecting and often amplifying the deeply entrenched historical and systemic biases present in the healthcare data it is trained on [1]. Understanding this mechanism is crucial for ensuring equitable care.

The Mechanisms of Algorithmic Bias

Bias is encoded into AI systems at various stages of development, primarily through two key mechanisms:

1. Data Collection and Labeling Bias

The quality and diversity of the training data are paramount. If clinical datasets lack diversity, the resulting model will be inherently biased against underrepresented groups, such as women, racial minorities, or individuals from lower socioeconomic backgrounds [2].

Underrepresentation: *Historical over-reliance on data from white, male, and affluent populations means the model's performance suffers significantly when applied to diverse patient groups.* **Proxy Variables:** A notorious example involves the use of variables like healthcare spending as a proxy for health need. A widely used algorithm for managing high-risk patients was found to systematically assign lower risk scores to Black patients than to white patients with the same health conditions [3]. This occurred because the model learned from historical data reflecting unequal access and lower spending on Black patients, leading to fewer being flagged for high-risk care management programs.

2. Model Development and Evaluation Bias

Even with diverse data, the way models are built and evaluated can introduce or amplify bias.

Flawed Metrics: *Developers often prioritize overall accuracy across the entire patient cohort. This "whole-cohort" optimization can obscure poor performance in smaller, underrepresented subgroups. A model might achieve high overall accuracy but only 70% accuracy for a specific racial group, leading to systematic misdiagnosis or inappropriate treatment [4].* **Real-World Deterioration:** The Epic Sepsis Model serves as a real-world example. Its performance was found to deteriorate differentially across patient groups in clinical settings, demonstrating a sample selection bias that required continuous monitoring and re-calibration to maintain fairness [5].

The Impact on Clinical Decision-Making

Biased AI systems perpetuate and amplify existing health disparities. A biased algorithm can lead to systematic errors in care:

Biased Outcome	Affected Population	Real-World Implication
Underestimation of risk	Racial minorities	Delayed or denied access to high-risk care management programs.
Diagnostic inaccuracy	Women, specific ethnic groups	Missed diagnoses where symptoms present differently across demographics.
Inappropriate treatment	Underrepresented groups	Suboptimal dosing or treatment plans due to the model's limited exposure to their data.

Furthermore, clinicians relying on AI for decision support may develop **automation bias**, over-relying on the system's output without critical evaluation, thereby unknowingly propagating the algorithmic error [6].

Strategies for Mitigating Algorithmic Bias

Addressing this challenge requires a multi-faceted approach across the entire AI lifecycle:

1. **Data Equity and Diversity:** Actively curating diverse, high-quality datasets that accurately reflect the patient population is essential. This includes incorporating socioeconomic and environmental factors rather than relying on flawed demographic proxies [4]. 2. **Fairness-Aware Model Design:** Developers must employ **subgroup analysis** and **bias-centered optimization metrics** to ensure equitable performance. Techniques like explicit statistical debiasing and Explainable AI (XAI) are crucial for identifying and correcting hidden biases [7]. 3. **Regulatory Oversight and Transparency:** Clear regulatory frameworks are needed to mandate bias reporting and fairness testing before clinical deployment. Ongoing, real-time monitoring systems are necessary to detect and quantify bias as the model interacts with real-world data [5].

For more in-depth analysis on this topic, the resources at [www.rasitdinc.com] (https://www.rasitdinc.com) provide expert commentary and further insights into the ethical and technical challenges of digital health and AI.

Conclusion: Towards Equitable AI in Medicine

AI's potential to revolutionize medicine is immense, but its success is contingent on our commitment to fairness. The biases in medical AI are a direct reflection of historical healthcare inequities. By implementing rigorous, fairness-aware development and deployment strategies, we can move beyond simply mirroring the past. The ultimate goal is to build AI systems that not only improve health outcomes but do so equitably for every patient.

References

[1] Norori, N. et al. (2021). Addressing bias in big data and AI for health care. *International Journal of Medical Informatics*, 154, 104568. [2] Obermeyer, Z. et al. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447-453. [3] Cross, J. L. et al. (2024). Bias in medical AI: Implications for clinical decision-making. *PLOS Digital Health*, 3(11), e0000651. [4] Hasanzadeh, F. et al. (2025). Bias recognition and mitigation strategies in artificial intelligence for healthcare: A systematic review. *npj Digital Medicine*, 8(1), 1-12. [5] Labkoff, S. et al. (2024). Recommendations for AI-enabled clinical decision support. *Journal of the American Medical Informatics Association*, 31(11), 2730-2738. [6] Abdelwanis, M. et al. (2024). Exploring the risks of automation bias in healthcare: A systematic review and Bowtie analysis. *Digital Health*, 10, 20552076241243547. [7] Ueda, D. et al. (2023). Fairness of artificial intelligence in healthcare: review and recommendations. *Journal of Medical Systems*, 47(1), 1-15.
