Can AI Discriminate Against Certain Patient Groups? An Academic Perspective on Algorithmic Bias in Digital Health

Rasit Dinc

Rasit Dinc Digital Health & AI Research

Abstract

Can AI Discriminate Against Certain Patient Groups? An Academic Perspective on Algorithmic Bias in Digital Health The integration of Artificial Intellig...

Can AI Discriminate Against Certain Patient Groups? An Academic Perspective on Algorithmic Bias in Digital Health

The integration of Artificial Intelligence (AI) into healthcare promises a revolution in diagnostics, treatment planning, and patient management. However, this transformative technology is not without its ethical and practical challenges. A critical question at the forefront of digital health discourse is: **Can AI discriminate against certain patient groups?** The answer, grounded in current academic research, is a resounding yes, and understanding the mechanisms of this bias is crucial for ensuring equitable healthcare delivery.

The Root of the Problem: Data Bias and Historical Inequity

AI systems, particularly those based on machine learning, are only as unbiased as the data they are trained on. The primary source of algorithmic discrimination in healthcare stems from the use of historical datasets that reflect and perpetuate existing societal and systemic inequities.

1. Unrepresentative Training Data

If a training dataset disproportionately features data from one demographic group (e.g., primarily white, male, or high-income patients), the resulting AI model will perform less accurately when applied to underrepresented groups (e.g., Black, Latinx, or low-income patients). For instance, an algorithm designed to predict patient outcomes might be trained on data where certain racial groups historically received less aggressive or lower-quality care. The AI, in turn, learns to associate these groups with poorer outcomes, not

because of biological factors, but because of the systemic bias in the care they received.

2. Proxy Variables and Social Determinants of Health

Another subtle form of bias arises when AI models use seemingly neutral variables that act as **proxies for race or socioeconomic status**. For example, an algorithm used by a major US health system to identify patients needing high-risk care was found to systematically assign lower risk scores to Black patients than to white patients who were equally sick. The model used healthcare costs as a proxy for health needs, but because Black patients historically incur lower healthcare costs due to reduced access to care, the algorithm incorrectly concluded they were healthier. This is a clear example of how an AI system can exacerbate existing racial disparities in treatment, as highlighted in numerous academic reviews on fairness in AI for healthcare [^1].

The Impact: Disparities in Diagnosis and Treatment

The consequences of biased AI are not theoretical; they manifest as tangible disparities in clinical settings:

Diagnostic Inaccuracy: *AI-powered diagnostic tools, such as those for skin cancer detection, have been shown to perform poorly on darker skin tones if the training images lacked diversity. Similarly, pulse oximeters, which use light to measure blood oxygen, have been found to be less accurate in patients with darker skin, a bias that can be amplified if AI models rely on these flawed measurements.* **Resource Allocation:** Algorithms used for resource allocation, such as scheduling follow-up appointments or prioritizing patients for specialized care, can inadvertently discriminate, leading to longer wait times or reduced access for certain patient groups.

Mitigating Algorithmic Discrimination: A Path to Equitable AI

Addressing this challenge requires a multi-faceted approach that spans data collection, model development, and regulatory oversight.

1. **Data Equity and Curation:** Developers must prioritize collecting and curating datasets that are truly representative of the entire patient population. This includes actively seeking data from diverse racial, ethnic, and socioeconomic groups. 2. **Fairness Metrics:** Moving beyond simple accuracy, AI developers must adopt and optimize for specific **fairness metrics** (e.g., equalized odds, demographic parity) to ensure the model performs equally well across different sensitive groups. 3. **Transparency and Auditability:** Healthcare providers and regulators must demand greater transparency into how AI models function. Algorithms should be auditable, allowing experts to identify and correct sources of bias before they impact patient care.

The ethical deployment of AI in medicine is paramount. It requires continuous vigilance and a commitment to addressing the historical biases embedded in our healthcare data. For more in-depth analysis on this topic, the resources at www.rasitdinc.com provide expert commentary

and further professional insights into the intersection of digital health, ethics, and AI governance.

Conclusion

The potential for AI to discriminate against certain patient groups is a serious ethical and clinical concern. It is not a flaw in the technology itself, but a reflection of the flawed, inequitable data we feed it. By embracing data equity, rigorous fairness testing, and transparent governance, the healthcare community can harness the power of AI to improve health outcomes for all, rather than widening the existing gaps.

**

[^1]: *Ueda, D., et al. (2023). Fairness of artificial intelligence in healthcare: review and recommendations.* BMC Medical Informatics and Decision Making*, 23(1), 1-15. [^2]: Mittermaier, M., et al. (2023). Bias in AI-based models for medical applications.* npj Digital Medicine*, 6(1), 1-10. [^3]: Norori, N., et al. (2021). Addressing bias in big data and AI for health care.* J Am Med Inform Assoc*, 28(12), 2742–2746.