Title

Training Data and Algorithmic Bias in Healthcare AI: A Standpoint Theory Perspective

Author

Zahoui, Sarah

Publication Date

2026-03-16

Training Data and Algorithmic Bias in Healthcare AI: A Standpoint Theory Perspective

Sarah Zahoui[1]

[1]Electrical & Computer Engineering Department, University of California, Los Angeles

*sarahz25@g.ucla.edu

Abstract: This paper examines bias in healthcare AI using standpoint theory, showing how demographic imbalances in training datasets can produce unequal diagnostic performance and highlighting the need for more representative data in medical AI development.

INTRODUCTION

The use of Artificial Intelligence (AI) has been on the rise across numerous healthcare domains, with 71% of hospitals reporting the use of AI in 2024 for predicting patient health trajectories, identifying high-risk patients, scheduling, and hospital operations [5]. Similarly, 66% of physicians reported the use of AI [4] for diagnosis assistance, writing visit notes and medical charts, creating discharge instructions and care plans, translation services, and more [8]. This rapid adoption has been accompanied by increasing financial investments as well. Global venture funding of AI startups in the medical field approximately reached $10.5 billion in 2024, showcasing the investors' confidence in AI technologies [6]. Additionally, the global AI healthcare market was valued at $39 billion last year and is projected to exceed one trillion dollars by 2034 [7]. These statistics reinforce the development of AI technologies and reliance in the upcoming decade.

As artificial intelligence continues to become integrated in healthcare systems, it is crucial to ensure its accuracy and safety, especially in its applications in decision-making processes. In applications such as diagnostic imaging and risk prediction, AI directly influences disease diagnosis and treatment. Therefore, AI-based technologies and programs carry significant implications for patient outcomes and mortality rates.

This cautionary perspective isn't uncommon. In fact, a survey conducted in 2022 by the Pew Research Center found that 60% of Americans are apprehensive about their medical providers using Artificial Intelligence in their practice [17]. In a more recent survey by the same organization, conducted in 2024, 56% of adults still didn't think that the impact of AI in the U.S. "over the next 20 years would be very or somewhat positive" [18]. This shows that the public opinion is more anxious than excited or trusting of this change, which presents more reason to correct any fault and improve AI systems before further development.

AI systems use clinical datasets and patient electronic records to generate predictions as they recognize patterns and relationships between medical symptoms and conditions. Standpoint theory leads us to understand how patient records aren't equally representative of different ethnicities, and prompts us to analyze the consequences of that. "This work is in partial fulfillment of the ENGR184 course using the blueprint curriculum in Ref [1,2] and captured in a collection [3]."

METHODS

Standpoint Theory

This paper uses standpoint theory as a framework to examine the use of AI tools in the healthcare system and its implications on marginalized communities.

Standpoint theory states that knowledge is socially situated rather than being an isolated truth, and that marginalized people's experiences offer an epistemic advantage due to their oppressed positions and thus lead to true objectivity. This framework highlights the importance of examining AI not only as a technically functioning and optimized instrument, but also as a product of social context that accounts for marginalized experiences.

In this case, the broad, hegamonic understanding is that AI models can achieve proper analytical capabilities and become fully functional by the standard AI model training. As previously mentioned, AI models are trained using large datasets, where algorithms recognize patterns and relationships between the presented data. Using the framework of standpoint theory, we question what additional requirements are essential for the true preparation of AI technologies. By examining the demographics across datasets, AI training reveals gaps in its preparation as it overlooks marginalized patient experiences.

RESULTS AND INTERPRETATION

MIMIC III Dataset: Biased Representation

In "Artificial Intelligence in Medicine and Healthcare - AI Datasets", The Philadelphia College of Osteopathic Medicine Library lists several datasets from medical organizations that are used for training and developing AI in healthcare, one of which is from the Beth Israel Deaconess Medical Center between 2001 and 2012 [9]. This dataset provided data for machine learning approaches in areas such as predicting patient outcomes, clinical implications of blood pressure monitoring techniques, and semantic analysis of unstructured patient notes [10]. Demographic descriptions, however, reveal a bias towards white/caucasian patients, with around 70% patient demographics belonging to White/Caucasians, 8–10% to Black/African Americans, 3–5% to Hispanic/Latino, 2–4% to Asians, and 10-15% to Other/Unknown/Multiracial patients [11].

This dataset provides an example of unequal representation, where one demographic group - White/Caucasian patients - dominates the dataset, which leads to a heavily biased algorithm in AI technologies, as AI machines are mostly exposed to one group.

This disparity presents an issue because people of the same ethnicity generally share a higher percentage of DNA compared to those from different ethnicities. Consequently, some genetic disorders are more likely to occur among the same ethnicities due to the presence of disease-causing variants among common ancestry [12]. For instance, sickle cell disease is more common among African and Mediterranean heritages, while Tay-Sachs disease is more common among eastern and central European heritages [12]. (It is important to note, however, that these disorders can occur in any ethnic group.

Therefore, in a heavily biased dataset, AI algorithms wouldn't be given enough chances to learn patterns among specific ethnic demographics on account of the misrepresentation.

Diverse Dermatology Images (DDI) Dataset; Diagnostic Imaging Bias

The Diverse Dermatology Images (DDI) dataset consists of 656 images created to evaluate diverse skin tones and determine whether AI systems in dermatology applications perform equally well across different skin tones. This study was prompted by the fact that most AI models haven't been rigorously assessed on images of diverse skin tones or uncommon diseases, and so researchers were concerned the system would perform inadequately were it to be tested across darker skin tones [14]. As expected, the results showed that the standard, state-of-the-art dermatology AI models perform significantly worse on the DDI dataset, with the accuracy dropping by approximately 27–36% compared to the performance reported during the original model testing [14].

These results provide yet another example that demonstrates how the composition of the datasets directly affects the accuracy and reliability of AI technologies. Because many dermatology datasets contain a larger proportion of images from lighter-skinned patients, AI models are exposed to fewer examples of how diseases appear on darker skin tones during training. As a result, the algorithm learns patterns that are more representative of the groups most frequently included in the dataset. When the system encounters patients whose characteristics are underrepresented in the training data, diagnostic accuracy Is likely to decrease.

Risk Prediction Algorithms and Social and Political Biases

Using standpoint theory as a framework helps clarify another point regarding the described defects. Standpoint theory argues that knowledge is shaped by the social, thus institutional, contexts in which it is produced. Since artificial intelligence in medicine reflects the populations that have historically been most represented in medical research and documentation, which is in favor of societal power structures, the biased algorithms serve as yet another method against marginalized communities.

In addition to diagnostic accuracy concerns, biased algorithms could also benefit institutions that prioritize efficiency and cost management within healthcare systems. An example of that is a healthcare risk prediction algorithm analyzed by Obermeyer et al., which was widely used by hospitals in the United States to identify patients who should receive additional medical care management [15]. The algorithm predicted which patients were most likely to develop serious health complications, but instead of directly measuring illness, the model used healthcare spending as a proxy for patient health needs. Because healthcare spending historically differs across racial groups due to unequal access to care, the algorithm tended to underestimate the health needs of Black patients compared to White patients who struggled with similar conditions [15]. The algorithm functioned effectively for institutional cost rather than health risk prediction, which benefits healthcare systems attempting to manage spending and maximize profits, while failing to reflect the true medical needs of all patient populations.

Several social and political factors may also explain why these disparities can remain unnoticed or insufficiently addressed. Many healthcare algorithms are developed by private companies and integrated into hospital systems as proprietary technologies, meaning the internal logic of the models is often difficult for clinicians or researchers to fully evaluate [15]. Additionally, algorithms are frequently perceived as objective because they rely on statistical analysis and large datasets, which may discourage questioning the assumptions embedded within the data used to train them. Historical inequalities in healthcare access further contribute to the problem. Marginalized communities often receive less medical care due to barriers such as insurance limitations, geographic access to healthcare facilities, and socioeconomic constraints [16]. When AI systems are trained on data produced within these unequal systems, they may reproduce those same patterns without being immediately recognized as biased. From the perspective of standpoint theory, this demonstrates how dominant institutional perspectives embedded in healthcare datasets can obscure the experiences of marginalized populations, whose perspectives may reveal limitations that remain invisible within dominant knowledge systems.

CONCLUSION

The increasing use of artificial intelligence in healthcare has introduced significant opportunities for improving efficiency, diagnostic accuracy, and clinical decision-making. However, the effectiveness of these systems is closely tied to the data used to train them. As shown through the examples discussed in this study, including the demographic imbalance present in the MIMIC III dataset and the diagnostic performance differences observed in dermatology AI systems tested on the Diverse Dermatology Images dataset, unequal representation within training data can directly influence the accuracy and reliability of AI technologies. When certain demographic groups are overrepresented while others are underrepresented, algorithms may learn patterns that are more reflective of dominant patient populations while overlooking the medical characteristics of marginalized groups. Through the perspective of standpoint theory, this issue demonstrates how datasets produced within existing healthcare systems may reflect dominant institutional perspectives while failing to capture the full range of patient experiences.

This is not to imply that Artificial Intelligence should be eliminated or limited in medicine, for it has shown immense promise and potential benefit to the field. This is not to imply that Artificial Intelligence should be eliminated or limited in medicine, for it has shown immense promise and potential benefit to the field. For example, AI-generated reports were compared to surgeon-written reports involving 14,606 patients across 158 cases. The AI-generated reports held a 87.3% accuracy, outperforming the surgeon-written ones, which had 72.8% accuracy [22].

These results illustrate how AI technologies can meaningfully assist medical practitioners when used appropriately. However, ensuring that these benefits apply to all patient populations requires careful consideration of how AI systems are developed and trained. By incorporating more diverse datasets and considering the perspectives of marginalized communities, researchers and healthcare institutions can work toward developing AI systems that produce more accurate and equitable outcomes.

Going forward, research in this field should focus on improving the representation of datasets used to train AI models. Because AI systems learn patterns from the data they are exposed to, expanding datasets to include patients from a wider range of hospitals, geographic regions, and demographic groups would allow algorithms to learn patterns that are more representative of the broader population.

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. 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. Carbajo, S. (2026). Queered Science & Technology Center: Volume 4. Queered Science & Technology Center, 4. Retrieved from https://escholarship.org/uc/item/0zb930m0

4. American Medical Association. "Physician Enthusiasm Grows for Health Care AI." *American Medical Association*, 2024. Retrieved from https://www.ama-assn.org/press-center/ama-press-releases/ama-physician-enthusiasm-grows-health-care-ai

5. Office of the National Coordinator for Health Information Technology. "Hospital Trends in Use, Evaluation, and Governance of Predictive AI (2023–2024)." *HealthIT.gov*, 2024. Retrieved from https://healthit.gov/data/data-briefs/hospital-trends-use-evaluation-and-governance-predictive-ai-2023-2024/

6. Qubit Capital. “AI Healthcare Investment Trends.” *Qubit Capital*, 2024. Retrieved from https://qubit.capital/blog/ai-healthcare-investment-trends
7. Fortune Business Insights. “Artificial Intelligence in Healthcare Market Size, Share & Industry Analysis.” *Fortune Business Insights*, 2024. Retrieved from https://www.fortunebusinessinsights.com/industry-reports/artificial-intelligence-in-healthcare-market-100534
8. Advisory Board. “How Physicians Are Using AI in Clinical Practice.” *Advisory Board Daily Briefing*, 2025. Retrieved from https://www.advisory.com/daily-briefing/2025/02/17/ai-use
9. Philadelphia College of Osteopathic Medicine Library. “Artificial Intelligence in Medicine and Healthcare – AI Datasets.” *PCOM Library Guides*, 2024. Retrieved from https://libguides.pcom.edu/c.php?g=1386314&p=10554413
10. Johnson, Alistair E. W., et al. “MIMIC-III, a Freely Accessible Critical Care Database.” *Scientific Data*, 2016, v. 3, article 160035. Retrieved from https://www.nature.com/articles/sdata201635
11. MIT Laboratory for Computational Physiology. “MIMIC-III Clinical Database Documentation.” *MIMIC Database Documentation*, 2023. Retrieved from https://mimic.mit.edu/docs/iii/
12. MedlinePlus Genetics. “Genetic Disorders and Ethnic Groups.” *U.S. National Library of Medicine*, 2023. Retrieved from https://medlineplus.gov/genetics/understanding/inheritance/ethnicgroup/
13. U.S. News & World Report. “Beth Israel Deaconess Medical Center.” *U.S. News Health*, 2024. Retrieved from https://health.usnews.com/best-hospitals/area/ma/beth-israel-deaconess-medical-center-6140013
14. Daneshjou, Roxana, et al. “Evaluating Dermatology AI Using the Diverse Dermatology Images (DDI) Dataset.” *arXiv preprint*, arXiv:2203.08807, 2022. Retrieved from https://arxiv.org/abs/2203.08807
15. Obermeyer, Ziad, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. “Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations.” *Science*, 2019, v. 366, n. 6464, pp. 447–453. doi:10.1126/science.aax2342
16. Artiga, Samantha, Kendal Orgera, and Anthony Damico. “Beyond Health Care: The Role of Social Determinants in Promoting Health and Health Equity.” *Kaiser Family Foundation*, 2020. Retrieved from https://www.kff.org/racial-equity-and-health-policy/report/beyond-health-care-the-role-of-social-determinants-in-promoting-health-and-health-equity/
17. Tyson, Alec, Giancarlo Pasquini, Alison Spencer, and Cary Funk. “60% of Americans Would Be Uncomfortable With Provider Relying on AI in Their Own Health Care.” *Pew Research Center*, 2023. Retrieved from https://www.pewresearch.org/science/2023/02/22/60-of-americans-would-be-uncomfortable-with-provider-relying-on-ai-in-their-own-health-care/
18. Pew Research Center. “How the U.S. Public and AI Experts View Artificial Intelligence.” *Pew Research Center*, 2025. Retrieved from https://www.pewresearch.org/internet/2025/04/03/how-the-us-public-and-ai-experts-view-artificial-intelligence/