Health AI poses distinct harms and potential benefits for people living with disabilities

Nature Medicine

https://doi.org/10.1038/s41591-024-03432-6 | Published online: 14 January 2025

Authors:

Charles E. Binkley, MD
Director of AI Ethics and Quality, Hackensack Meridian Health
Associate Professor of Surgery, Hackensack Meridian School of Medicine
Charles.binkley@hmhn.org
(corresponding author)

Joel M. Reynolds, PhD
Associate Professor of Philosophy & Disability Studies, Georgetown University
Faculty, Department of Family Medicine and Pellegrino Center for Clinical Bioethics, Georgetown University School of Medicine and Medical Center

Andrew G. Shuman, MD
Associate Professor, Department of Otolaryngology - Head & Neck Surgery
Chief, Clinical Ethics Service, Center for Bioethics and Social Sciences in Medicine
University of Michigan Medical School

Manuscript:

Relative to other groups for whom the risk of bias in artificial intelligence (AI) has been identified, little attention has been given to the potential harms, and distinct benefits, that AI models could bring about for disabled people. Predictions made by AI models for patients with disabilities may lead to discrimination in at least three ways: disabled patients are underrepresented in datasets; they have historically been discriminated against in their medical care; and data used to predict physiological frailty may overestimate the degree of frailty for disabled patients.

The disability community is smaller than the general population, and most healthcare data on which AI models are being trained come from able-bodied patients. AI models trained on data primarily from non-disabled patients may accurately predict a given outcome for these patients but fail when applied to disabled patients. In addition, health data pertinent to disabled patients may be missing or incomplete. The potential for bias and discrimination attributable to underrepresentation and misrepresentation in data sets is well described.[1]

Historical data may reflect the biases of healthcare providers regarding disabled patients; such biases are based on clinicians' judgment of a disabled patient's quality of life, the patient's worthiness and the perceived futility of treatment.[2] Disabled patients are often denied indicated and at times life-sustaining treatments on the basis of their disability rather than on an objective assessment of medical indication and the likely success of treatment.[3] AI models trained on historical data may thus predict that disabled patients have higher disease-specific and overall mortality rates. However, disabled patients die more frequently not only because of cases where the health ramifications of a disability are tightly linked to increased morbidity and mortality, but also because they are simply not provided with indicated treatments.[4] Healthcare data will also reflect biases in the language used by clinicians to describe disabled patients. Stigmatizing descriptions such as 'bed-bound,' 'feeble,' 'incompetent,' 'childlike,' and 'challenged' will be contained in the health data used to train AI models.[5] These descriptions may bias predictive models, and large language models may perpetuate their use.

AI models may falsely predict higher physiologic frailty for disabled patients. Proxy markers for illness, such as the frequency of healthcare provider contacts, the number of previous operations, the distance or speed at which a patient can walk, grip strength, rate of speech and cognitive ability, may suggest that patients with disability are sicker and at higher risk of morbidity or mortality than similar non-disabled patients.[6] As a result, patients may not be offered potentially beneficial interventions because AI models predict that their risk of a major complication or death is prohibitive. In this case, harm is not due to incomplete and unrepresentative data or biased data sets but rather the wrong algorithmic assumption.

The effect that inaccurate AI predictions could have on disabled patients is substantial. Besides their direct effect on disabled patients, inaccurate predictions may indirectly influence other healthcare decisions. Combined with implicit quality-of-life assumptions about disabled patients, clinicians may interpret the predictions made by AI models to mean that further goal-directed treatment is not beneficial or is futile. As AI predictions are integrated into the patient's electronic medical record, overestimates of illness, morbidity and mortality can also affect decisions made by insurance companies to provide coverage of indicated treatments. Further complicating matters is that AI predictions are not always transparent to patients or clinicians. Contrary to transparency and disclosure advocated for by the American Medical Association, predictions may be made without informing patients that the prediction is being made or of the prediction itself.[7]

How should these concerns be addressed? One of the more promising strategies is to build disability-specific AI models trained on relevant health data. Using innovative data approaches such as transfer learning, smaller sets can be fine-tuned to make accurate predictions on limited data.[8] Such an approach could be used to adapt AI models trained on data from the general population to disabled people. Creating an accurately labeled disability-specific data set, whatever its size, would rely on either patient self-identification as disabled or proxies for disability, such as access needs. Noting the need for more accurate disability identification, several agencies have issued requirements for disability status documentation.[7]

While AI models trained on data from the general population pose a risk of harm to disabled patients, other models may be more beneficial, such as an AI model that operationalizes population-level health data to recommend specific community services to a given patient based on the patient's needs. Often, the best long-term intervention for a disabled patient, especially a newly disabled patient, is a connection to an Independent Living Center or a similar organization in their community. For disabled patients, access to care is more than care availability; it requires accessibility in public spaces, disability-competent care, facility with communication and appreciation of patient-centered quality of life determinations. It is also essential that disabled people contribute to healthcare AI development and implementation by participating in governance, contributing to use cases and representing the lived experience of disability.

Perhaps the greatest benefit of AI models to patients with disabilities is the democratization of healthcare and centering a patient's life as it is lived relative to where they are, who they know and what social and cultural connections might best improve their health outcomes.[9] Such an approach is similar to using AI to address social determinants of health.[10] Ironically, AI may shift medicine from a disease focus to an emphasis on wholeness and human flourishing in the myriad ways it is experienced by disabled and able-bodied patients alike. Research in social epidemiology demonstrates that health is not merely a question of individual choice or heredity but of organism-environment interaction. The greatest value of AI models trained to focus on

community-clinic connections may thus be in altering the trajectories of populations facing health disparities.

Competing interests:
The authors declare no competing interests.

References:

[1] Adamson A.S., Smith A. *JAMA Dermatol* **154**, 1247–1248 (2018).

[2] Binkley C.E., Reynolds J.M., and Shuman A. NEJM **387**, 1325-1328 (2022).

[3] National Council on Disability. *Medical Futility and Disability Bias: Part of the Bioethics and Disability Series* (NCD, 2019).

[4] Gonzalez, D., Kenney, G. M., Karpman, M., & Morriss, S. *Four in Ten Adults with Disabilities Experienced Unfair Treatment in Health Care Settings, at Work, or When Applying for Public Benefits in 2022* (Urban Institute, 2023).

[5] National Academies of Sciences, Engineering, and Medicine. *Health Disparities in the Medical Record and Disability Determinations: Proceedings of a Workshop* (The National Academies Press, 2024)

[6] Reynolds J.M., Binkley C.E., Shuman A. *Am J Bioeth* **21**, 74-76 (2021).

[7] American Medical Association. *Principles for Augmented Intelligence Development, Deployment, and Use* (AMA, 2023).

[8] Chahal, H., Toner, H. & Rahkovsky, I. *Small Data's Big AI Potential* (Center for Security and Emerging Technology, 2021).

[9] National Academy of Medicine. *Vital Directions for Health & Health Care: An Initiative of the National Academy of Medicine* (The National Academies Press, 2017).

[10] Ong, J. C. L. et al. *Cell Rep. Med.* 5, 101356 (2024).