

How an archaeological approach can help leverage biased data in AI to improve medicine

September 14 2023, by Alex Ouyang



In a new paper, computer science and bioethics professors from MIT, Johns Hopkins University, and the Alan Turing Institute call for an alternative approach to understanding biased data used in medical machine learning—one that views biased clinical data as akin to archaeological artifacts that connect back to societal values, practices, and patterns of inequity. Credit: Marzyeh Ghassemi via Midjourney

The classic computer science adage "garbage in, garbage out" lacks nuance when it comes to understanding biased medical data, argue computer science and bioethics professors from MIT, Johns Hopkins University, and the Alan Turing Institute in a [new opinion piece](#) published in a recent edition of the *New England Journal of Medicine* (*NEJM*).

The rising popularity of artificial intelligence has brought increased scrutiny to the matter of biased AI models resulting in algorithmic discrimination, which the White House Office of Science and Technology identified as a key issue in their recent Blueprint for an AI Bill of Rights.

When encountering biased data, particularly for AI models used in medical settings, the typical response is to either collect more data from underrepresented groups or generate synthetic data making up for missing parts to ensure that the model performs equally well across an array of patient populations. But the authors argue that this technical approach should be augmented with a sociotechnical perspective that takes both historical and current social factors into account. By doing so, researchers can be more effective in addressing bias in [public health](#).

"The three of us had been discussing the ways in which we often treat issues with data from a [machine learning](#) perspective as irritations that need to be managed with a technical solution," recalls co-author Marzyeh Ghassemi, an assistant professor in [electrical engineering](#) and computer science and an affiliate of the Abdul Latif Jameel Clinic for Machine Learning in Health (Jameel Clinic), the Computer Science and Artificial Intelligence Laboratory (CSAIL), and Institute of Medical Engineering and Science (IMES).

"We had used analogies of data as an artifact that gives a partial view of past practices, or a cracked mirror holding up a reflection. In both cases the information is perhaps not entirely accurate or favorable: Maybe we think that we behave in certain ways as a society—but when you actually look at the data, it tells a different story. We might not like what that story is, but once you unearth an understanding of the past you can move forward and take steps to address poor practices."

Data as artifact

In their paper, titled "Considering Biased Data as Informative Artifacts in AI-Assisted Health Care," Ghassemi, Kadija Ferryman, and Maxine Mackintosh make the case for viewing biased [clinical data](#) as "artifacts" in the same way anthropologists or archaeologists would view physical objects: pieces of civilization-revealing practices, [belief systems](#), and cultural values—in the case of the paper, specifically those that have led to existing inequities in the health care system.

For example, [a 2019 study](#) showed that an algorithm widely considered to be an industry standard used health-care expenditures as an indicator of need, leading to the erroneous conclusion that sicker Black patients require the same level of care as healthier white patients. What researchers found was algorithmic discrimination failing to account for unequal access to care.

In this instance, rather than viewing biased datasets or lack of data as problems that only require disposal or fixing, Ghassemi and her colleagues recommend the "artifacts" approach as a way to raise awareness around social and historical elements influencing how data are collected and alternative approaches to clinical AI development.

"If the goal of your model is deployment in a clinical setting, you should engage a bioethicist or a clinician with appropriate training reasonably

early on in problem formulation," says Ghassemi. "As computer scientists, we often don't have a complete picture of the different social and historical factors that have gone into creating data that we'll be using. We need expertise in discerning when models generalized from existing data may not work well for specific subgroups."

When more data can actually harm performance

The authors acknowledge that one of the more challenging aspects of implementing an artifact-based approach is being able to assess whether data have been racially corrected: i.e., using white, male bodies as the conventional standard that other bodies are measured against. The opinion piece cites [an example](#) from the Chronic Kidney Disease Collaboration in 2021, which developed a new equation to measure kidney function because the old equation had previously been "corrected" under the blanket assumption that Black people have higher muscle mass. Ghassemi says that researchers should be prepared to investigate race-based correction as part of the research process.

In another [recent paper](#) accepted to this year's [International Conference on Machine Learning](#) co-authored by Ghassemi's Ph.D. student Vinith Suriyakumar and University of California at San Diego Assistant Professor Berk Ustun, the researchers found that assuming the inclusion of personalized attributes like self-reported race improve the performance of ML models can actually lead to worse risk scores, models, and metrics for minority and minoritized populations.

"There's no single right solution for whether or not to include self-reported race in a clinical risk score. Self-reported race is a social construct that is both a proxy for other information, and deeply proxied itself in other [medical data](#). The solution needs to fit the evidence," explains Ghassemi.

How to move forward

This is not to say that biased datasets should be enshrined, or biased algorithms don't require fixing—quality training data is still key to developing safe, high-performance clinical AI models, and the NEJM piece highlights the role of the National Institutes of Health (NIH) in driving ethical practices.

"Generating high-quality, ethically sourced datasets is crucial for enabling the use of next-generation AI technologies that transform how we do research," NIH acting director Lawrence Tabak stated in a press release when the NIH announced its \$130 million Bridge2AI Program last year.

Ghassemi agrees, pointing out that the NIH has "prioritized data collection in ethical ways that cover information we have not previously emphasized the value of in human health—such as environmental factors and social determinants. I'm very excited about their prioritization of, and strong investments towards, achieving meaningful health outcomes."

Elaine Nsoesie, an associate professor at the Boston University of Public Health, believes there are many potential benefits to treating biased datasets as artifacts rather than garbage, starting with the focus on context. "Biases present in a dataset collected for lung cancer patients in a hospital in Uganda might be different from a dataset collected in the U.S. for the same patient population," she explains. "In considering local context, we can train algorithms to better serve specific populations."

Nsoesie says that understanding the historical and contemporary factors shaping a dataset can make it easier to identify discriminatory practices that might be coded in algorithms or systems in ways that are not immediately obvious. She also notes that an artifact-based approach could lead to the development of new policies and structures ensuring

that the root causes of bias in a particular dataset are eliminated.

"People often tell me that they are very afraid of AI, especially in health. They'll say, 'I'm really scared of an AI misdiagnosing me,' or 'I'm concerned it will treat me poorly,'" Ghassemi says. "I tell them, you shouldn't be scared of some hypothetical AI in health tomorrow, you should be scared of what health is right now. If we take a narrow technical view of the data we extract from systems, we could naively replicate poor practices. That's not the only option—realizing there is a problem is our first step towards a larger opportunity."

More information: Kadija Ferryman et al, Considering Biased Data as Informative Artifacts in AI-Assisted Health Care, *New England Journal of Medicine* (2023). [DOI: 10.1056/NEJMra2214964](https://doi.org/10.1056/NEJMra2214964)

This story is republished courtesy of MIT News (web.mit.edu/newsoffice/), a popular site that covers news about MIT research, innovation and teaching.

Provided by Massachusetts Institute of Technology

Citation: How an archaeological approach can help leverage biased data in AI to improve medicine (2023, September 14) retrieved 27 April 2024 from <https://medicalxpress.com/news/2023-09-archaeological-approach-leverage-biased-ai.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--