

SHARE

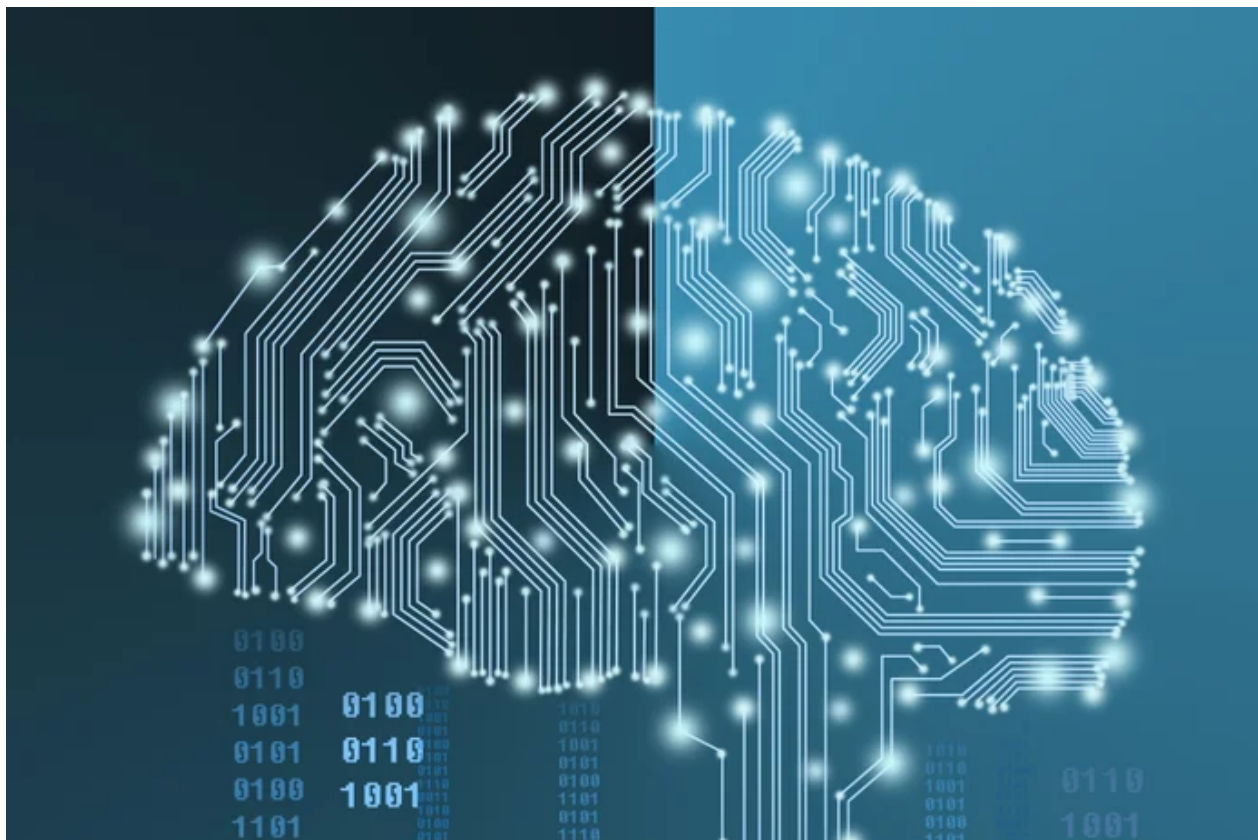
LATEST

POLICY & ETHICS | OPINION

Health Care AI Systems Are Biased

We need more diverse data to avoid perpetuating inequality in medicine

By Amit Kaushal, Russ Altman, Curt Langlotz on November 17, 2020



Credit: Getty Images

Thanks to advances in artificial intelligence (AI) and machine learning, computer systems can now diagnose skin cancer like a dermatologist would, pick out a stroke on a CT scan like a radiologist, and even detect potential cancers on a colonoscopy like a gastroenterologist. These new expert digital diagnosticians promise to put our caregivers

on technology's curve of bigger, better, faster, cheaper. But what if they make medicine more biased too?

At a time when the country is grappling with systemic bias in core societal institutions, we need technology to reduce health disparities, not exacerbate them. We've long known that AI algorithms that were trained with data that do not represent the whole population often perform worse for underrepresented groups. For example, algorithms trained with gender-imbalanced data do worse at reading chest x-rays for an underrepresented gender, and researchers are already concerned that skin-cancer detection algorithms, many of which are trained primarily on light-skinned individuals, do worse at detecting skin cancer affecting darker skin.

Given the consequences of an incorrect decision, high-stakes medical AI algorithms need to be trained with data sets drawn from diverse populations. Yet, this diverse training is not happening. In a recent study published in JAMA (the Journal of the American Medical Association), we reviewed over 70 publications that compared the diagnostic prowess of doctors against digital doppelgangers across several areas of clinical medicine. Most of the data used to train those AI algorithms came from just three states: California, New York and Massachusetts.

Whether by race, gender or geography, medical AI has a data diversity problem: researchers can't easily obtain large, diverse medical data sets—and that can lead to biased algorithms.

Why aren't better data available? One of our patients, a veteran, once remarked in frustration after trying to obtain his prior medical records: "Doc, why is it that we can see a specific car in a moving convoy on the other side of the world, but we can't see my CT scan from the hospital across the street?" Sharing data in medicine is hard enough for a single patient, never mind the hundreds or thousands of cases needed to reliably train machine learning algorithms. Whether in treating patients or building AI tools, data in medicine are locked in little silos everywhere.

Medical data sharing should be more commonplace. But the sanctity of medical data and the strength of relevant privacy laws provide strong incentives to protect data, and severe consequences for any error in data sharing. Data are sometimes sequestered for

economic reasons; one study found hospitals that shared data were more likely to lose patients to local competitors. And even when the will to share data exists, lack of interoperability between medical records systems remains a formidable technical barrier. The backlash from big tech's use of personal data over the past two decades has also cast a long shadow over medical data sharing. The public has become deeply skeptical of any attempt to aggregate personal data, even for a worthy purpose.

This is not the first time that medical data have lacked diversity. Since the early days of clinical trials, women and minority groups have been underrepresented as study participants; evidence mounted that these groups experienced fewer benefits and more side effects from approved medications. Addressing this imbalance ultimately required a joint effort from the NIH, FDA, researchers and industry, and an act of Congress in 1993; it remains a work in progress to this day. One of the companies racing toward a COVID vaccine recently announced a delay to recruit more diverse participants; it's that important.

It's not just medicine; AI has begun to play the role of trained expert in other high-stakes domains. AI tools help judges with sentencing decisions, redirect the focus of law enforcement, and suggest to bank officers whether to approve a loan application. Before algorithms become an integral part of high-stakes decisions that can enhance or derail the lives of everyday citizens, we must understand and mitigate embedded biases.

Bias in AI is a complex issue; simply providing diverse training data does not guarantee elimination of bias. Several other concerns have been raised—for example, lack of diversity among developers and funders of AI tools; framing of problems from the perspective of majority groups; implicitly biased assumptions about data; and use of outputs of AI tools to perpetuate biases, either inadvertently or explicitly. Because obtaining high-quality data is challenging, researchers are building algorithms that try to do more with less. From these innovations may emerge new ways to decrease AI's need for huge data sets. But for now, ensuring diversity of data used to train algorithms is central to our ability to understand and mitigate biases of AI.

To ensure that the algorithms of tomorrow are not just powerful but also fair, we must build the technical, regulatory, economic and privacy infrastructure to deliver the large and diverse data required to train these algorithms. We can no longer move forward

blindly, building and deploying tools with whatever data happen to be available, dazzled by a veneer of digital gloss and promises of progress, and then lament the “unforeseeable consequences.” The consequences are foreseeable. But they don’t have to be inevitable.

[Rights & Permissions](#)

ABOUT THE AUTHOR(S)

Amit Kaushal

Amit Kaushal, M.D., Ph.D., is an adjunct professor in the Department of Bioengineering at Stanford University.

Recent Articles

The Talent Gap Holding Back Digital Medicine

Russ Altman

Russ Altman, M.D., Ph.D., is a professor in the Department of Bioengineering at Stanford University, and a faculty affiliate at the Stanford Institute for Human Centered Artificial Intelligence.

Curt Langlotz

Curt Langlotz, M.D., Ph.D., is a professor in the Department of Radiology, Stanford University School of Medicine.