

## A.I. Could Worsen Health Disparities

By Dhruv Khullar

THE NEW YORK TIMES, Jan. 31, 2019 (abridged)

Artificial intelligence is beginning to meet (and sometimes exceed) assessments by doctors in various clinical situations. [...]

There are many questions about whether A.I. actually works in medicine, and where it works: can it pick up pneumonia, detect cancer, predict death? But those questions focus on the technical, not the ethical. And in a health system riddled with inequity, we have to ask: Could the use of A.I. in medicine worsen health disparities? There are at least three reasons to believe it might.

The first is a training problem. A.I. must learn to diagnose disease on large data sets, and if that data doesn't include enough patients from a particular background, it won't be as reliable for them. Evidence from other fields suggests this isn't just a theoretical concern. A recent study found that some facial recognition programs incorrectly classify less than 1 percent of light-skinned men but more than one-third of dark-skinned women. What happens when we rely on such algorithms to diagnose melanoma on light versus dark skin? [...]

Second, because A.I. is trained on real-world data, it risks incorporating, entrenching and perpetuating the economic and social biases that contribute to health disparities in the first place. Again, evidence from other fields is instructive. A.I. programs used to help judges predict which criminals are most likely to reoffend have shown troubling racial biases, as have those designed to help child protective services decide which calls require further investigation. [...]

Finally, even ostensibly fair, neutral A.I. has the potential to worsen disparities if its implementation has disproportionate effects for certain groups. Consider a program that helps doctors decide whether a patient should go home or to a rehab facility after knee surgery. It's a decision imbued with uncertainty but has real consequences: Evidence suggests discharge to an institution is associated with higher costs and higher risk of readmission. If an algorithm incorporates residence in a low-income neighborhood as a marker for poor social support, it may recommend minority patients go to nursing facilities instead of receive home-based physical therapy. Worse yet, a program designed to maximize efficiency or lower medical costs might discourage operating on those patients altogether.

To some extent, all these problems already exist in medicine. American health care has always struggled with income- and race-based inequities rooted in various forms of bias. The risk with A.I. is that these biases become automated and invisible — that we begin to accept the wisdom of machines over the wisdom of our own clinical and moral intuition. Many A.I. programs are black boxes: We don't know exactly what's going on inside and why they produce the output they do. But we may increasingly be expected to honor their recommendations. [...]

Still, A.I. holds tremendous potential to improve medicine. It may well make care more efficient, more accurate and — if properly deployed — more equitable. But realizing this promise requires being aware of the potential for bias and guarding against it. It means regularly monitoring both the output of algorithms and the downstream consequences. In some cases, this will necessitate counter-bias algorithms that hunt for and correct subtle, systematic discrimination.

But most fundamentally, it means recognizing that humans, not machines, are still responsible for caring for patients. It is our duty to ensure that we're using AI as another tool at our disposal — not the other way around.

*Dr. Khullar is an assistant professor of health care policy and research.*