

Artificial Intelligence Makes Bad Medicine Even Worse

Christie Aschwanden

Google researchers made [headlines](#) early this month for a [study](#) that claimed their artificial intelligence system could outperform human experts at finding breast cancers on mammograms. It sounded like a big win, and yet another example of how [AI will soon transform health care](#): More cancers found! Fewer false positives! A better, cheaper way to provide high-quality medical care!

Hold on to your exclamation points. Machine-enabled health care may bring us many benefits in the years to come, but those will be contingent on the ways in which it's used. If doctors ask the wrong questions to begin with—if they put AI to work pursuing faulty premises—then the technology will be a bust. It could even serve to amplify our earlier mistakes.

In a sense, that's what happened with the recent Google paper. It's trying to replicate, and then exceed, human performance on what is at its core a deeply flawed medical intervention. In case you haven't been following the decades-long [controversy over cancer screening](#), it boils down to this: When you subject symptom-free people to mammograms and the like, you'll end up finding a lot of things that look like cancer but will never threaten anyone's life. As the science of cancer biology has advanced and screening has become widespread, researchers have learned that not every tumor is destined to become deadly. In fact, many people harbor indolent forms of cancer that do not actually pose a risk to their health. Unfortunately, standard screening tests have proven most adept at finding precisely the latter—the slower-growing ones that would better be ignored.

This might not be so bad, in theory. When a screening test uncovers harmless cancer, you can just ignore it, right? The problem is, it's almost impossible to know at the time of screening whether any particular lesion will end up dangerous or no big deal. In practice, most doctors are inclined to treat any cancer that's discovered as a potential threat, and the question of whether or not mammograms actually save lives is a matter of intense [debate](#). Some studies suggest [they do](#), others find that [they don't](#), but even if we take the rosier interpretations of the literature at face value, the number of lives saved by this massive, widespread intervention is small. [Some](#) researchers have even [calculated](#) that mammography is, in balance, bad for patients' health; i.e. that its aggregate harms, in terms of the excess treatment it inspires and the tumors brought on by its radiation, outweigh any benefits.

In other words, AI systems like the one from Google promise to combine humans and machines in order to facilitate cancer diagnosis, but they also have the potential to worsen pre-existing problems such as overtesting, overdiagnosis, and overtreatment. It's not even clear whether the improvements in false-positive and false-negative rates reported this month would apply in real-world settings. The Google study found that AI performed better than radiologists who were not specifically trained in examining mammograms. Would it come out on top against a team of more specialized experts? It's hard to say without a trial. Furthermore, most of the images assessed in the study were created with imaging devices made by a single company. It remains to be seen whether these results would generalize to images from other machines.

The problem goes beyond just breast-cancer screening. Part of the appeal of AI is that it can scan through reams of familiar data, and pick out variables that we never realized were important. In principle, that power could help us to diagnose any early-stage disease, in the same way the subtle squiggles of a seismograph can give us early warnings of an earthquake. (AI [helps](#) there, too, by the way.) But sometimes those hidden variables really *aren't* important. For instance, your data set might be drawing from a cancer screening clinic that is only open for lung cancer tests on Fridays. As a result, an AI algorithm could decide that scans taken on Fridays are more likely to be lung cancer. That trivial relationship would then get baked into the formula for making further diagnoses.

Even when they're accurate, early diagnoses of disease may not always be a boon. Other recent medical AI projects have focused on early detection of [Alzheimer's](#) and [autism](#), two conditions where faster detection probably won't change a patient's outcome much anyway. These are gee-whiz opportunities to showcase how an algorithm can learn to identify characteristics we teach it to find, but they don't represent advancements that will make a difference in patients' lives.

SUBSCRIBE

[Subscribe](#) to WIRED and stay smart with more of your favorite [Ideas](#) writers.

Some uses of algorithms and machine learning may also introduce new and perplexing problems for clinicians. Consider the Apple watch's feature to detect [atrial fibrillation](#), a type of heart arrhythmia that's a risk factor for stroke. Atrial fibrillation is treated with blood thinners, which have side effects that can turn a minor fall into a life-threatening injury. If you're truly in danger of having a stroke, that's a risk worth taking. What about people whose atrial fibrillation was picked up by their smartwatch, though? Traditionally, the condition is diagnosed when someone comes into the doctor complaining of symptoms; now Apple monitors healthy people without symptoms and finds new cases that may have never shown up in a clinic. It's not clear whether this group of patients would see the same net benefit from treatment.

"We don't actually know that these two populations of people are the same," says Venkatesh Murthy, a cardiologist at Frankel Cardiovascular Center in Ann Arbor, Michigan. The more fruitful approach would be to use AI to identify the people who get the most benefit from the available treatments.

If AI is going to prove truly revolutionary, it will need to do more than just reinstate the status quo in medicine; and before any such approach is adopted, it's important to address a pair of fundamental questions: What problem is the technology trying to address, and how will it improve patient outcomes? It may take some time to find the necessary answers.

That's why the famous Mark Zuckerberg motto, "Move fast and break things" might be fine for Facebook, but it's not great for medicine, AI-assisted or not. According to Vinay Prasad, author of [Ending Medical Reversal and](#) a hematologist-oncologist at the Oregon Health & Science University School of Medicine, the Silicon Valley mindset can be dangerous for clinicians. It's that kind of attitude—when lives are at stake we need to implement promising new ideas as quickly as possible—that got us into this cancer-screening mess in the first place. Mammography was adopted before all the evidence was in, Prasad says, and once a medical practice has become standard, it's very difficult to phase it out. "In a culture that's used to immediacy and inflated claims, it's difficult to have humility and patience."

More Great WIRED Stories

- Live your best life—[on and off your phone](#)—this year
- How to secure your Wi-Fi router and [protect your home network](#)
- Mid-century modern motels, [in all their neon glory](#)
- How Hong Kong's protests [turned into a *Mad Max* tableau](#)
- Why the “queen of shitty robots” [renounced her crown](#)
- 👁️ Will AI as a field ["hit the wall" soon](#)? Plus, the [latest news on artificial intelligence](#)
- 🎧 Things not sounding right? Check out our favorite [wireless headphones](#), [soundbars](#), and [Bluetooth speakers](#)