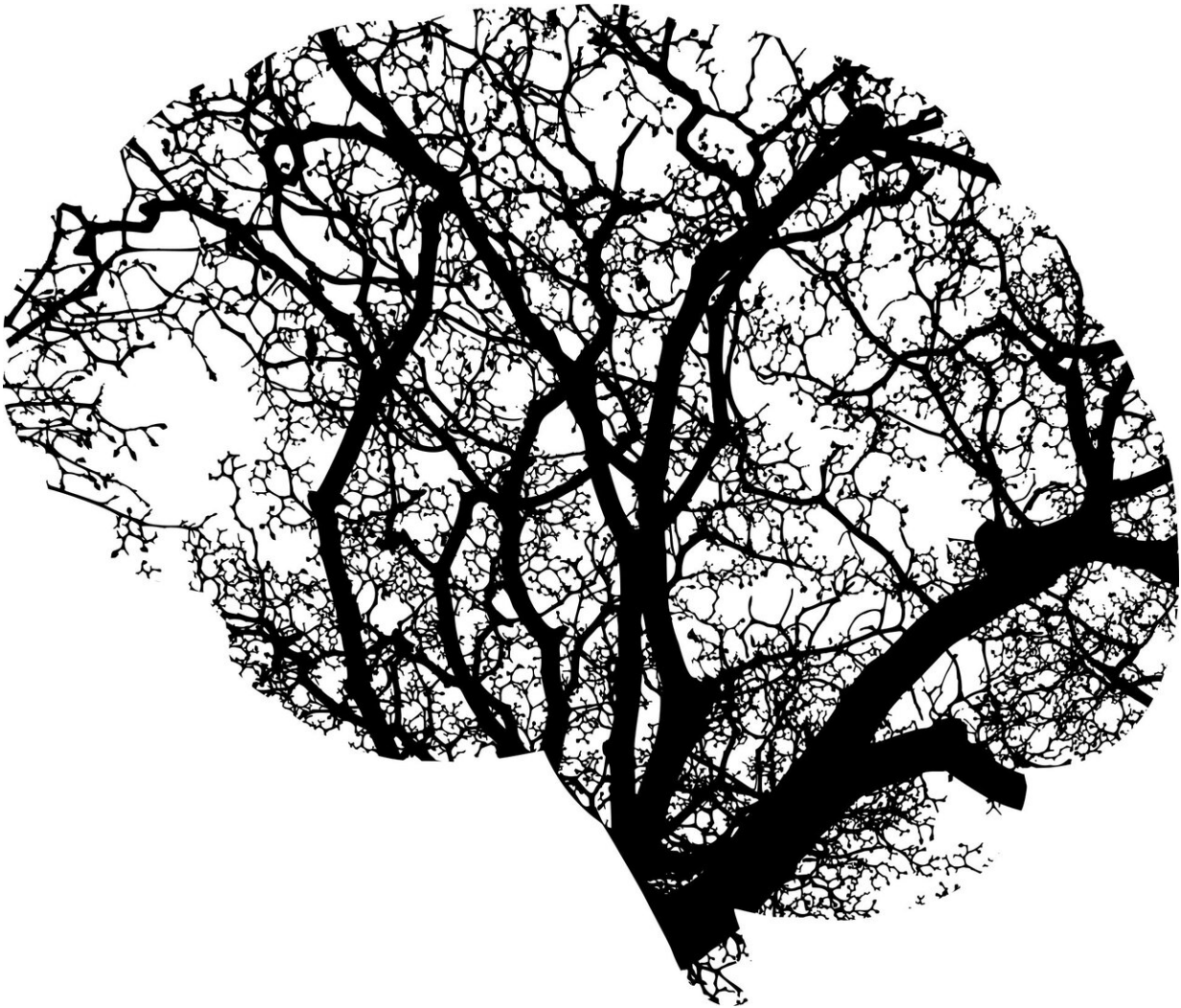


Artificial intelligence has come to medicine. Are patients being put at risk?

January 8 2020, by Liz Szabo



Credit: CC0 Public Domain

Health products powered by artificial intelligence, or AI, are streaming into our lives, from virtual doctor apps to wearable sensors and drugstore chatbots.

IBM boasted that its AI could "outthink cancer." Others say [computer systems](#) that read X-rays will make radiologists obsolete.

"There's nothing that I've seen in my 30-plus years studying medicine that could be as impactful and transformative" as AI, said Dr. Eric Topol, a cardiologist and executive vice president of Scripps Research in La Jolla, Calif. AI can help doctors interpret MRIs of the heart, CT scans of the head and photographs of the back of the eye, and could potentially take over many mundane medical chores, freeing doctors to spend more time talking to patients, Topol said.

Even the Food and Drug Administration—which has approved more than 40 AI products in the past five years—says "the potential of digital health is nothing short of revolutionary."

Yet many health industry experts fear AI-based products won't be able to match the hype. Many doctors and consumer advocates fear that the [tech industry](#), which lives by the mantra "fail fast and fix it later," is putting patients at risk—and that regulators aren't doing enough to keep consumers safe.

Early experiments in AI provide a reason for caution, said Mildred Cho, a professor of pediatrics at Stanford's Center for Biomedical Ethics.

Systems developed in one hospital often flop when deployed in a different facility, Cho said. Software used in the care of millions of Americans has been shown to discriminate against minorities. And AI systems sometimes learn to make predictions based on factors that have less to do with disease than the brand of MRI machine used, the time a

[blood test](#) is taken or whether a patient was visited by a chaplain. In one case, AI software incorrectly concluded that people with pneumonia were less likely to die if they had asthma—an error that could have led doctors to deprive asthma patients of the extra care they need.

"It's only a matter of time before something like this leads to a serious health problem," said Dr. Steven Nissen, chairman of cardiology at the Cleveland Clinic.

Medical AI, which pulled in \$1.6 billion in venture capital funding in the third quarter alone, is "nearly at the peak of inflated expectations," concluded a July report from the research company Gartner. "As the reality gets tested, there will likely be a rough slide into the trough of disillusionment."

That reality check could come in the form of disappointing results when AI products are ushered into the real world. Even Topol, the author of "Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again," acknowledges that many AI products are little more than hot air. "It's a mixed bag," he said.

Experts such as Dr. Bob Kocher, a partner at the venture capital firm Venrock, are blunter. "Most AI products have little evidence to support them," Kocher said. Some risks won't become apparent until an AI system has been used by large numbers of patients. "We're going to keep discovering a whole bunch of risks and unintended consequences of using AI on medical data," Kocher said.

None of the AI products sold in the U.S. have been tested in randomized clinical trials, the strongest source of medical evidence, Topol said. The first and only randomized trial of an AI system—which found that colonoscopy with computer-aided diagnosis found more small polyps than standard colonoscopy—was published online in October.

Few tech startups publish their research in peer-reviewed journals, which allow other scientists to scrutinize their work, according to a January article in the *European Journal of Clinical Investigation*. Such "stealth research"—described only in press releases or promotional events—often overstates a company's accomplishments.

And although [software developers](#) may boast about the accuracy of their AI devices, experts note that AI models are mostly tested on computers, not in hospitals or other medical facilities. Using unproven software "may make patients into unwitting guinea pigs," said Dr. Ron Li, medical informatics director for AI clinical integration at Stanford Health Care.

AI systems that learn to recognize patterns in data are often described as "black boxes" because even their developers don't know how they have reached their conclusions. Given that AI is so new—and many of its risks unknown—the field needs careful oversight, said Pilar Ossorio, a professor of law and bioethics at the University of Wisconsin-Madison.

Yet the majority of AI devices don't require FDA approval.

"None of the companies that I have invested in are covered by the FDA regulations," Kocher said.

Legislation passed by Congress in 2016—and championed by the tech industry—exempts many types of medical software from federal review, including certain fitness apps, electronic health records and tools that help doctors make medical decisions.

There's been little research on whether the 320,000 medical apps now in use actually improve health, according to a report on AI published Dec. 17 by the National Academy of Medicine.

"Almost none of the (AI) stuff marketed to patients really works," said Dr. Ezekiel Emanuel, professor of medical ethics and health policy in the Perelman School of Medicine at the University of Pennsylvania.

The FDA has long focused its attention on devices that pose the greatest threat to patients. And consumer advocates acknowledge that some devices—such as ones that help people count their daily steps—need less scrutiny than ones that diagnose or treat disease.

Some software developers don't bother to apply for FDA clearance or authorization, even when legally required, according to a 2018 study in *Annals of Internal Medicine*.

Industry analysts say that AI developers have little interest in conducting expensive and time-consuming trials. "It's not the main concern of these firms to submit themselves to rigorous evaluation that would be published in a peer-reviewed journal," said Joachim Roski, a principal at Booz Allen Hamilton, a technology consulting firm, and co-author of the National Academy's report. "That's not how the U.S. economy works."

But Oren Etzioni, chief executive officer at the Allen Institute for AI in Seattle, said AI developers have a financial incentive to make sure their medical products are safe.

"If failing fast means a whole bunch of people will die, I don't think we want to fail fast," Etzioni said. "Nobody is going to be happy, including investors, if people die or are severely hurt."

Relaxing Standards At The FDA

The FDA has come under fire in recent years for allowing the sale of dangerous medical devices, which have been linked by the International Consortium of Investigative Journalists to 80,000 deaths and 1.7 million

injuries over the past decade.

Many of these devices were cleared for use through a controversial process called the 510(k) pathway, which allows companies to market "moderate-risk" products with no clinical testing as long as they're deemed similar to existing devices.

In 2011, a committee of the National Academy of Medicine concluded the 510(k) process is so fundamentally flawed that the FDA should throw it out and start over.

Instead, the FDA is using the process to greenlight AI devices.

Of the 14 AI products authorized by the FDA in 2017 and 2018, 11 were cleared through the 510(k) process, according to a November article in *JAMA*. None of these appear to have had new clinical testing, the study said. The FDA cleared an AI device designed to help diagnose liver and lung cancer in 2018 based on its similarity to imaging software approved 20 years earlier. That software had itself been cleared because it was deemed "substantially equivalent" to products marketed before 1976.

AI products cleared by the FDA today are largely "locked," so that their calculations and results will not change after they enter the market, said Bakul Patel, director for digital health at the FDA's Center for Devices and Radiological Health. The FDA has not yet authorized "unlocked" AI devices, whose results could vary from month to month in ways that developers cannot predict.

To deal with the flood of AI products, the FDA is testing a radically different approach to digital device regulation, focusing on evaluating companies, not products.

The FDA's pilot "pre-certification" program, launched in 2017, is

designed to "reduce the time and cost of market entry for software developers," imposing the "least burdensome" system possible. FDA officials say they want to keep pace with AI software developers, who update their products much more frequently than makers of traditional devices, such as X-ray machines.

Scott Gottlieb said in 2017 while he was FDA commissioner that government regulators need to make sure its approach to innovative products "is efficient and that it fosters, not impedes, innovation."

Under the plan, the FDA would pre-certify companies that "demonstrate a culture of quality and organizational excellence," which would allow them to provide less upfront data about devices.

Pre-certified companies could then release devices with a "streamlined" review—or no FDA review at all. Once products are on the market, companies will be responsible for monitoring their own products' safety and reporting back to the FDA. Nine companies have been selected for the pilot: Apple, FitBit, Samsung, Johnson & Johnson, Pear Therapeutics, Phosphorus, Roche, Tidepool and Verily Life Sciences.

High-risk products, such as software used in pacemakers, will still get a comprehensive FDA evaluation. "We definitely don't want patients to be hurt," said Patel, who noted that devices cleared through pre-certification can be recalled if needed. "There are a lot of guardrails still in place."

But research shows that even low- and moderate-risk devices have been recalled due to serious risks to patients, said Diana Zuckerman, president of the National Center for Health Research. "People could be harmed because something wasn't required to be proven accurate or safe before it is widely used."

Johnson & Johnson, for example, has recalled hip implants and surgical mesh.

In a series of letters to the FDA, the American Medical Association and others have questioned the wisdom of allowing companies to monitor their own performance and product safety.

"The honor system is not a regulatory regime," said Dr. Jesse Ehrenfeld, who chairs the physician group's board of trustees.

In an October letter to the FDA, Sens. Elizabeth Warren (D-Mass.), Tina Smith (D-Minn.) and Patty Murray (D-Wash.) questioned the agency's ability to ensure company safety reports are "accurate, timely and based on all available information."

When Good Algorithms Go Bad

Some AI devices are more carefully tested than others.

An AI-powered screening tool for diabetic eye disease was studied in 900 patients at 10 primary care offices before being approved in 2018. The manufacturer, IDx Technologies, worked with the FDA for eight years to get the product right, said Dr. Michael Abramoff, the company's founder and executive chairman.

The test, sold as IDx-DR, screens patients for diabetic retinopathy, a leading cause of blindness, and refers high-risk patients to eye specialists, who make a definitive diagnosis.

IDx-DR is the first "autonomous" AI product—one that can make a screening decision without a doctor. The company is now installing it in primary care clinics and grocery stores, where it can be operated by employees with a high school diploma. Abramoff's company has taken

the unusual step of buying liability insurance to cover any patient injuries.

Yet some AI-based innovations intended to improve care have had the opposite effect.

A Canadian company, for example, developed AI software to predict a person's risk of Alzheimer's based on their speech. Predictions were more accurate for some patients than others. "Difficulty finding the right word may be due to unfamiliarity with English, rather than to cognitive impairment," said co-author Frank Rudzicz, an associate professor of computer science at the University of Toronto.

Doctors at New York's Mount Sinai Hospital hoped AI could help them use chest X-rays to predict which patients were at high risk of pneumonia. Although the system made accurate predictions from X-rays shot at Mount Sinai, the technology flopped when tested on images taken at other hospitals. Eventually, researchers realized the computer had merely learned to tell the difference between that hospital's portable chest X-rays—taken at a patient's bedside—with those taken in the radiology department. Doctors tend to use portable chest X-rays for patients too sick to leave their room, so it's not surprising that these patients had a greater risk of lung infection.

DeepMind, a company owned by Google, has created an AI-based mobile app that can predict which hospitalized patients will develop acute kidney failure up to 48 hours in advance. A blog post on the DeepMind website described the system, used at a London hospital, as a "game changer." But the AI system also produced two false alarms for every correct result, according to a July study in *Nature*. That may explain why patients' kidney function didn't improve, said Dr. Saurabh Jha, associate professor of radiology at the Hospital of the University of Pennsylvania. Any benefit from early detection of serious kidney

problems may have been diluted by a high rate of "overdiagnosis," in which the AI system flagged borderline kidney issues that didn't need treatment, Jha said. Google had no comment in response to Jha's conclusions.

False positives can harm patients by prompting doctors to order unnecessary tests or withhold recommended treatments, Jha said. For example, a doctor worried about a patient's kidneys might stop prescribing ibuprofen—a generally safe pain reliever that poses a small risk to kidney function—in favor of an opioid, which carries a serious risk of addiction.

As these studies show, software with impressive results in a computer lab can founder when tested in real time, Stanford's Cho said. That's because diseases are more complex—and the health care system far more dysfunctional—than many computer scientists anticipate.

Many AI developers cull electronic health records because they hold huge amounts of detailed data, Cho said. But those developers often aren't aware that they're building atop a deeply broken system. Electronic health records were developed for billing, not patient care, and are filled with mistakes or missing data.

A KHN investigation published in March found sometimes life-threatening errors in patients' medication lists, lab tests and allergies.

In view of the risks involved, doctors need to step in to protect their patients' interests, said Dr. Vikas Saini, a cardiologist and president of the nonprofit Lown Institute, which advocates for wider access to health care.

"While it is the job of entrepreneurs to think big and take risks," Saini said, "it is the job of doctors to protect their patients."

©2020 Kaiser Health News
Distributed by Tribune Content Agency, LLC.

Citation: Artificial intelligence has come to medicine. Are patients being put at risk? (2020, January 8) retrieved 25 April 2024 from <https://medicalxpress.com/news/2020-01-artificial-intelligence-medicine-patients.html>

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.