

BRAVE NEW WORLD DEPT.

IF A.I. CAN DIAGNOSE PATIENTS, WHAT ARE DOCTORS FOR?

Large language models are transforming medicine—but the technology comes with side effects.

By Dhruv Khullar

September 22, 2025



"I'm worried these tools will erode my ability to make an independent diagnosis," a medical student said. Illustration by Petra Péterffy



Save this story

Listen to this story

00:00

1x



37:55

In 2017, Matthew Williams, a thirtysomething software engineer with an athletic build and a bald head, went for a long bike ride in the hills of San Francisco. Afterward, at dinner with some friends, he ordered a hamburger, fries, and a milkshake. Midway through the meal, he felt so full that he had to ask someone to drive him home. That night, Williams awoke with a sharp pain in his abdomen that he worried was appendicitis. He went to a nearby emergency clinic, where doctors told him that he was probably constipated. They gave him some laxatives and sent him on his way.

A few hours later, Williams's pain intensified. He vomited and felt as though his stomach might burst. A friend took him to a hospital, where a CT scan revealed cecal volvulus—a medical emergency in which part of the intestine twists in on itself, cutting off the digestive tract. The previous medical team had missed the condition, and may even have exacerbated it by giving him laxatives. Williams was rushed to the operating room, where surgeons removed about six feet of his intestines.

After recovering from surgery, Williams began to experience severe diarrhea almost every time he ate. Doctors told him that his bowel just needed time to heal. “It got to the point where I couldn’t go out, because I would constantly eat something that would make me sick,” he said. During the next few years, Williams saw a series of nutritionists and gastroenterologists—eight clinicians in total—but none could pinpoint the reason for his symptoms. “Doctors are sometimes just, like, ‘Are you not dying? O.K., then come back another time,’ ” he said. Williams largely restricted his diet to eggs, rice, applesauce, and sourdough bread. “You don’t understand how much food is a part of life—socially, culturally—until you can’t eat it anymore,” he told me. “It’s awkward to be on a date and explain why you can’t get the mozzarella sticks. When your food is bland, your life becomes bland, too.”

Get our Classics newsletter to discover timeless gems from *The New Yorker* archive.

SIGN UP

In 2023, on a whim, Williams entered his medical history into ChatGPT. “I have lost most of my ileum and my cecal valve, why might the following foods cause gastrointestinal distress,” he typed, and then listed some of the worst offenders. Within seconds, the A.I. pointed to three potential triggers for his symptoms: fatty foods, fermentable fibres, and foods high in oxalate. Oxalate, a compound found in leafy greens and a variety of other foods, is normally broken down by parts of the G.I. tract that Williams had lost; he’d never heard of it, not even from his doctors. He asked the A.I. for a list of high-oxalate foods and was stunned. “It listed every single food that made me the sickest,” he said—spinach, almonds, chocolate, soy, and more than a dozen others. “It’s like it had been following me around, taking notes.” Williams brought the information to a nutritionist, who crafted a diet based on the oxalate content of foods. His symptoms improved, and his meals grew more varied. Williams no longer needs to know the location of the nearest bathroom at all times. “I have my life back,” he said.

During my medical training, I revered senior physicians who, through some alchemy of knowledge and gestalt, always seemed to home in on the clue that cracked the case: the unusual shape of a patient’s nails; an occupational hazard from decades before; an overlooked blood test. What algorithm was running in these physicians’ minds? Could I load it into my own? In the future, however, diagnosis may increasingly become a computer science. Surveys have suggested that many people are more confident in A.I. diagnoses than in those rendered by professionals. Meanwhile, in the United States alone, misdiagnosis disables hundreds of thousands of people each year; autopsy studies suggest that it contributes to perhaps one in every ten deaths. If Williams hadn’t ignored his initial diagnosis, he might have been among them. “I trust A.I. more than doctors,” he said. “I don’t think I’m the only one.”

In the early nineteen-hundreds, Richard Cabot, a physician at Massachusetts General Hospital, started holding seminars to demonstrate clinical reasoning for trainees. An expert physician would be given a former patient’s file and would probe for more detail about the case. If the information had been

available during the patient's real-life hospitalization, it was revealed. Gradually, the physician would edge toward a diagnosis that could be compared with whatever the pathologists had ultimately concluded, often during an autopsy. Clinicopathological conferences, or C.P.C.s, as they came to be known, grew to be so popular that *The New England Journal of Medicine* has published transcripts of them for more than a century. They represent a gold standard of diagnostic reasoning: if you can solve a C.P.C., you can solve almost any case.

C.P.C.s also inspired many efforts to teach medicine to machines. In the late fifties, a computer scientist and a radiologist grouped cases by symptoms and diseases. They proposed that a computer program could analyze the cases using mathematical tools such as logic and game theory. "Computers are especially suited to help the physician collect and process clinical information and remind him of diagnoses which he might have overlooked," they wrote in a landmark *Science* paper. In the seventies, a computer scientist at the University of Pittsburgh developed a program called INTERNIST-1, based on a series of conversations with a brilliant and intimidating physician named Jack Myers. (Myers was known as Black Jack, because he failed so many new doctors during their board exams.) Myers "chose a goodly number" of C.P.C.s to show how he reasoned; INTERNIST-1 eventually performed as well as some doctors did on a variety of cases. But details of a case had to be painstakingly entered into the computer, so each analysis could take more than an hour. Researchers concluded that "the present form of the program was not sufficiently reliable for clinical applications."

Then came large language models. Last year, Arjun Manrai, a computer scientist at Harvard, and Thomas Buckley, a doctoral student in the university's new A.I. in Medicine program, started work on an education-and-research tool that was supposed to be capable of solving virtually any C.P.C. It needed to be able to cite the literature, explain its rationale, and help doctors think through a difficult case. Manrai and Buckley developed a custom version of o3, an

advanced “reasoning model” from OpenAI, which takes the time to break complex problems into intermediate steps before responding. A process known as retrieval-augmented generation, or RAG, pulls data from external sources before the A.I. crafts its answer. Their model is a bit like a student consulting a textbook to write a paper rather than writing from memory. They named the A.I. CaBot, in honor of the inventor of C.P.C.s.

In July, I travelled to Harvard’s Countway Library of Medicine to witness a face-off between CaBot and an expert diagnostician. The event brought to mind the 1997 chess match between the grand master Garry Kasparov and Deep Blue, the I.B.M. supercomputer that ultimately defeated him. I walked past the skull of Phineas Gage, a patient who famously survived an explosion that propelled an iron rod through his head. Then I came to a large conference room where dozens of students, doctors, and researchers sat, chattering excitedly. Daniel Restrepo, an internist at Massachusetts General Hospital who had been one of my classmates in residency, would be competing against CaBot. I remembered Restrepo as someone for whom diagnostic reasoning was like an Olympic sport; he sometimes read textbooks while I napped during overnight shifts, and he regularly ran to the laboratory to personally inspect a patient’s urine sample.

Manrai, a genial man with short black hair, works on a floor of Countway that was once home to library stacks. Now it is occupied by a bay of computers. He introduced the day’s case. “Today, we’re here to see Dr. CaBot,” he said. He described a forty-one-year-old man who had come to the hospital after about

ten days of fevers, body aches, and swollen ankles. The man had a painful rash on his shins and had fainted twice. A few months earlier, doctors had placed a stent in his heart. A CT scan showed lung nodules and enlarged lymph nodes in the man's chest.

Restrepo, who was wearing professorial glasses and a dark suit, went first. The opening move toward a diagnosis, he said, was defining the problem: "If you crystallize it into a clear and concise statement, your brain will have an easier way of solving that problem." He stressed three questions: Who was the patient? How quickly did the condition arise? And which symptoms constituted a syndrome? Some symptoms would hang together; others were likely to be distractions. "Despite getting all that other data, this is actually what I think is salient," he said, showing the audience four key symptoms on a Venn diagram. They pointed him to three diagnostic categories: lymphoma, infection, and autoimmune disease.

The man's symptoms had arisen too rapidly for lymphoma. "Tempo, tempo, tempo!" Restrepo said. An unusual infection seemed unlikely—the man was born in the U.S., he wasn't immunocompromised, and he wasn't known to have visited high-risk places. And it wouldn't explain his joint pain. "What do I know that causes fever, arthritis, hilar adenopathy, and a lower-extremity rash all at the same time?" Restrepo finally said. "Löfgren syndrome." Löfgren is a rare manifestation of sarcoidosis, an inflammatory condition. We learned that the man had received steroids, which suppress inflammation, while in the hospital. He'd improved, suggesting that the diagnosis was correct. The audience clapped.

Manrai returned to the podium. Restrepo had been given six weeks to prepare his presentation, he explained with a smile. "Dr. CaBot got six minutes," he said. A slide, generated by the A.I., appeared on the screen. It was titled "When Ankles, Nodes, and Syncope Collide." Manrai pressed Play and took a seat. A woman's voice—warm and casual, but professional—filled the room. "Good morning, everyone," it said. "I'm Dr. CaBot, and, um, we have what I think is a

really instructive case that links dermatology, rheumatology, pulmonology, and even cardiology. So, let's jump right in."

The voice, whose style and cadence were indistinguishable from those of human doctors, began to review the patient's medications and medical history. "No exotic exposures," CaBot said. "Just life in urban New England, with a cat that scratched him six months ago—which, you know, I keep in the back of my mind, but I'm not married to it!" The audience laughed. The model seemed to have sifted through the case for information that it deemed most relevant. "The joints are the star of the show," it said. It highlighted small nodules that lined some lymphatic vessels in the man's lungs, as seen in the CT scan. "Note how they track along the fissures," CaBot observed.

The A.I. generated an array of possible diagnoses, pointing out the strengths and weaknesses of each one. It noted that the patient had high levels of C-reactive protein, a biomarker of inflammation that is sometimes associated with autoimmune conditions. "Putting it together," CaBot said, "the single best fit is acute sarcoidosis, manifesting as Löfgren syndrome." For a moment the audience was silent. Then a murmur rippled through the room. A frontier seemed to have been crossed.

For a long time, when I've tried to imagine A.I. performing the complex cognitive work of doctors, I've asked, How could it? The demonstration forced me to confront the opposite question: How could it not? CaBot had occasionally hit a wrong note—for instance, pronouncing "hilar" as "hee-lar" instead of "high-lur"—and it advised more aggressive management than Restrepo had, including a lymph-node biopsy. (Most experts don't consider a biopsy necessary, but the man's real-life medical team had considered one.) Still, the presentation had been astonishingly good—better than many I had sat through during my medical education. And it had been created in the time that it takes me to brew a cup of coffee.

CaBot's success was at odds with what some patients experience when they consult chatbots. One recent study found that OpenAI's GPT-4 answered open-ended medical questions incorrectly about two-thirds of the time. In another, GPT-3.5 misdiagnosed more than eighty per cent of complex pediatric cases. Meanwhile, leading large language models have become much less likely to include disclaimers in their responses. One analysis found that, in 2022, more than a quarter of responses to health-related queries included something like "I am not qualified to give medical advice." This year, only one per cent did. In a new survey, about a fifth of Americans said that they've taken medical advice from A.I. that later proved to be incorrect. Earlier this year, a poison-control center in Arizona reported a drop in total call volume but a rise in severely poisoned patients. The center's director suggested that A.I. tools might have steered people away from medical attention. Chatbots also create serious privacy concerns: once your medical information enters the chat, it no longer belongs to you. Last year, Elon Musk encouraged users of X to upload their medical images to Grok, the platform's A.I., for "analysis." The company was later found to have made hundreds of thousands of chat transcripts accessible to search engines, often without permission.

Annals of Internal Medicine: Clinical Cases, a peer-reviewed medical journal, recently published an instructive example. A sixty-year-old man who was concerned about how much salt, or sodium chloride, he was eating asked ChatGPT for possible substitutes. The A.I. suggested bromide, an early anti-seizure medication that causes neurological and psychiatric issues when it accumulates in the body. The man ordered some online; within months, he was in an emergency room, believing that his neighbor was trying to poison him. He felt a profound thirst but grew paranoid when he was offered water. Blood testing showed a bromide level that was hundreds of times above normal. He started to hallucinate and tried to flee the hospital. Doctors placed him on an involuntary psychiatric hold. When they replicated his query in ChatGPT, it again suggested bromide.

After CaBot's presentation, one of Manrai's collaborators, a doctor at Beth Israel Deaconess Medical Center named Adam Rodman, got up to share a few remarks. Rodman leads Harvard's efforts to integrate generative A.I. into its medical-school curriculum. He noted that both Restrepo and CaBot had used a process called differential diagnosis, which begins by considering all potential explanations and then systematically rules out those which don't fit. But whereas Restrepo had emphasized the patient's constellation of symptoms —“he took the syndromic approach,” Rodman said—CaBot had zoomed in on the lung nodules, something that most doctors probably would not do. “One of the things that Dr. CaBot decided to do very early on was say, ‘Hey, look at this CT scan, look at how these nodules are in a lymphatic distribution. I’m going to build a differential on this!’” Rodman said. The A.I. had called out the absence of lung cavitations that might have suggested tuberculosis; it had emphasized subtle imaging findings that Restrepo hadn't even mentioned. CaBot's process was recognizable to humans, Rodman observed, but it had different strengths. “Because it encodes so much more information, it picked up these items to build its checklist that very few humans would have,” he said. When Manrai and his colleagues tested the A.I. on several hundred recent C.P.C.s, it correctly solved about sixty per cent of them, a significantly higher proportion than doctors solved in a prior study.

Learning how to deploy A.I. in the medical field, Rodman told me later, will require a science of its own. Last year, he co-authored a study in which some doctors solved cases with help from ChatGPT. They performed no better than doctors who didn't use the chatbot. The chatbot alone, however, solved the cases more accurately than the humans. In a follow-up study, Rodman's team suggested specific ways of using A.I.: they asked some doctors to read the A.I.'s opinion before they analyzed cases, and told others to give A.I. their working diagnosis and ask for a second opinion. This time, both groups diagnosed patients more accurately than humans alone did. The first group proved faster and more effective at proposing next steps. When the chatbot went second, however, it frequently “disobeyed” an instruction to ignore what the doctors had

concluded. It seemed to cheat, by anchoring its analysis to the doctor's existing diagnosis.

Systems that strategically combine human and A.I. capabilities have been described as centaurs; Rodman's research suggests that they have promise in medicine. But if A.I. tools remain imperfect and humans lose the ability to function without them—a risk known as “cognitive de-skilling”—then, in Rodman's words, “we're screwed.” In a recent study, gastroenterologists who used A.I. to detect polyps during colonoscopies got significantly worse at finding polyps themselves. “If you're a betting person, you should train doctors who know how to use A.I. but also know how to think,” Rodman said.

It seems inevitable that the future of medicine will involve A.I., and medical schools are already encouraging students to use large language models. “I'm worried these tools will erode my ability to make an independent diagnosis,” Benjamin Popokh, a medical student at University of Texas Southwestern, told me. Popokh decided to become a doctor after a twelve-year-old cousin died of a brain tumor. On a recent rotation, his professors asked his class to work through a case using A.I. tools such as ChatGPT and OpenEvidence, an increasingly popular medical L.L.M. that provides free access to health-care professionals. Each chatbot correctly diagnosed a blood clot in the lungs. “There was no control group,” Popokh said, meaning that none of the students worked through the case unassisted. For a time, Popokh found himself using A.I. after virtually every patient encounter. “I started to feel dirty presenting my thoughts to attending physicians, knowing they were actually the A.I.'s thoughts,” he told me. One day, as he left the hospital, he had an unsettling realization: he hadn't thought about a single patient independently that day. He decided that, from then on, he would force himself to settle on a diagnosis before consulting artificial intelligence. “I went to medical school to become a real, capital-‘D’ doctor,” he told me. “If all you do is plug symptoms into an A.I.,

are you still a doctor, or are you just slightly better at prompting A.I. than your patients?”

A few weeks after the CaBot demonstration, Manrai gave me access to the model. It was trained on C.P.C.s from *The New England Journal of Medicine*; I first tested it on cases from the *JAMA* network, a family of leading medical journals. It made accurate diagnoses of patients with a variety of conditions, including rashes, lumps, growths, and muscle loss, with a small number of exceptions: it mistook one type of tumor for another and misdiagnosed a viral mouth ulcer as cancer. (ChatGPT, in comparison, misdiagnosed about half the cases I gave it, mistaking cancer for an infection and an allergic reaction for an autoimmune condition.) Real patients do not present as carefully curated case studies, however, and I wanted to see how CaBot would respond to the kinds of situations that doctors actually encounter.

I gave CaBot the broad stokes of what Matthew Williams had experienced: bike ride, dinner, abdominal pain, vomiting, two emergency-department visits. I didn't organize the information in the way that a doctor would. Alarmingly, when CaBot generated one of its crisp presentations, the slides were full of made-up lab values, vital signs, and exam findings. “Abdomen looks distended up top,” the A.I. said, incorrectly. “When you rock him gently, you hear that classic succussion splash—liquid sloshing in a closed container.” CaBot even conjured up a report of a CT scan that supposedly showed Williams's bloated stomach. It arrived at a mistaken diagnosis of gastric volvulus: a twisting of the stomach, not the bowel.

I tried giving CaBot a formal summary of Williams's second emergency visit, as detailed by the doctors who saw him, and this produced a very different result—presumably because they had more data, sorted by salience. The patient's hemoglobin level had plummeted; his white cells, or leukocytes, had multiplied; he was doubled over in pain. This time, CaBot latched on to the pertinent data and did not seem to make anything up. “Strangulation indicators—constant pain, leukocytosis, dropping hemoglobin—are all flashing at us,” it said. CaBot diagnosed an obstruction in the small intestines, possibly owing to volvulus or a

hernia. “Get surgery involved early,” it said. Technically, CaBot was slightly off the mark: Williams’s problem arose in the large, not the small, intestine. But the next steps would have been virtually identical. A surgeon would have found the intestinal knot.

Talking to CaBot was both empowering and unnerving. I felt as though I could now receive a second opinion, in any specialty, anytime I wanted. But only with vigilance and medical training could I take full advantage of its abilities—and detect its mistakes. A.I. models can sound like Ph.D.s, even while making grade-school errors in judgment. Chatbots can’t examine patients, and they’re known to struggle with open-ended queries. Their output gets better when you emphasize what’s most important, but most people aren’t trained to sort symptoms in that way. A person with chest pain might be experiencing acid reflux, inflammation, or a heart attack; a doctor would ask whether the pain happens when they eat, when they walk, or when they’re lying in bed. If the person leans forward, does the pain worsen or lessen? Sometimes we listen for phrases that dramatically increase the odds of a particular condition. “Worst headache of my life” may mean brain hemorrhage; “curtain over my eye” suggests a retinal-artery blockage. The difference between A.I. and earlier diagnostic technologies is like the difference between a power saw and a hacksaw. But a user who’s not careful could cut off a finger.

Attend enough clinicopathological conferences, or watch enough episodes of “House,” and every medical case starts to sound like a mystery to be solved. Lisa Sanders, the doctor at the center of the *Times Magazine* column and Netflix series “Diagnosis,” has compared her work to that of Sherlock Holmes. But the daily practice of medicine is often far more routine and repetitive. On a rotation at a V.A. hospital during my training, for example, I felt less like Sherlock than like Sisyphus. Virtually every patient, it seemed, presented with some combination of emphysema, heart failure, diabetes, chronic kidney disease, and high blood pressure. I became acquainted with a new phrase—“likely multifactorial,” which meant that there were several explanations for what the patient was experiencing—and I looked for ways to

address one condition without exacerbating another. (Draining fluid to relieve an overloaded heart, for example, can easily dehydrate the kidneys.) Sometimes a precise diagnosis was beside the point; a patient might come in with shortness of breath and low oxygen levels and be treated for chronic obstructive pulmonary disease, heart failure, and pneumonia. Sometimes we never figured out which had caused a given episode—yet we could help the patient feel better and send him home. Asking an A.I. to diagnose him would not have offered us much clarity; in practice, there was no neat and satisfying solution.

Tasking an A.I. with solving a medical case makes the mistake of “starting with the end,” according to Gurpreet Dhaliwal, a physician at the University of California, San Francisco, whom the *Times* once described as “one of the most skillful clinical diagnosticians in practice.” In Dhaliwal’s view, doctors are better off asking A.I. for help with “wayfinding”: instead of asking what sickened a patient, a doctor could ask a model to identify trends in the patient’s trajectory, along with important details that the doctor might have missed. The model would not give the doctor orders to follow; instead, it might alert her to a recent study, propose a helpful blood test, or unearth a lab result in a decades-old medical record. Dhaliwal’s vision for medical A.I. recognizes the difference between diagnosing people and competently caring for them. “Just because you have a Japanese-English dictionary in your desk doesn’t mean you’re fluent in Japanese,” he told me.

CaBot remains experimental, but other A.I. tools are already shaping patient care. ChatGPT is blocked on my hospital’s network, but I and many of my colleagues use OpenEvidence. The platform has licensing agreements with top

medical journals and says it complies with the patient-privacy law HIPAA. Each of its answers cites a set of peer-reviewed articles, sometimes including an exact figure or a verbatim quote from a relevant paper, to prevent hallucinations. When I gave OpenEvidence a recent case, it didn't immediately try to solve the mystery but, rather, asked me a series of clarifying questions.

Penda Health, a network of medical clinics in Kenya, treats an enormous range of patients, from newborns sickened by malaria to construction workers who have fallen off buildings. Kenya has long struggled with a limited health-care infrastructure. Penda recently began using AI Consult, a tool that employs OpenAI models and runs in the background while clinicians record medical histories, order tests, and prescribe medicines. A clinician who overlooks a patient's anemia would get an alert to consider ordering an iron test; another, treating a child with diarrhea, might be advised to forgo antibiotics in favor of an oral rehydration solution and zinc supplements.

An evaluation of the program, which was conducted in collaboration with OpenAI and has not been peer-reviewed, reported that clinicians who used AI Consult made sixteen per cent fewer diagnostic errors and thirteen per cent fewer treatment errors. They seemed to learn from the program: the number of safety alerts dropped significantly over time. AI Consult made mistakes; in testing, it confused a cough syrup for an antibiotic of a similar name. The absolute number of medical errors at Penda also remained high—at times because clinicians ignored the model's advice. "They know that this patient doesn't necessarily need an antibiotic, but they also know that the patient really wants it," Robert Korom, Penda's chief medical officer, said. Still, a Penda clinician deemed the program a "tremendous improvement." Its success may have come from its focus not on diagnosis but on helping clinicians navigate the possibilities.

A similar principle could guide patients. If A.I. tools continue to misdiagnose and hallucinate, we might not want them to diagnose us at all. Yet we could ask

them to rate the urgency of our symptoms, and to list the range of conditions that could explain them, with some sense of which ones are most likely. A patient could inquire about “red-flag symptoms”—warning signs that would indicate a more serious condition—and about which trusted sources the A.I. is drawing on. A chatbot that gets details wrong could still help you consider what to ask at your next appointment. And it could aid you in decoding your doctor’s advice.

Jorie Bresnahan, whose ninety-five-year-old mother was recently hospitalized for heart failure, told me that in order to keep track of her mother’s care she made audio recordings when doctors, nurses, and therapists explained treatments and procedures. The conversations were dizzying, and A.I.-generated transcripts “just looked like a mess,” she said. But when she uploaded the transcripts to ChatGPT, it imposed coherence and highlighted details she’d overlooked. Bresnahan and her sisters, who lived far away, could then talk to the chatbot about her mother’s condition. After her mother left the hospital, Bresnahan put the A.I. in voice mode, so that her mother could ask it questions, too. “She thought it was very charming,” Bresnahan told me. “She started calling him Trevor.”

Bresnahan eventually caught the chatbot mixing up dates and hallucinating blood-pressure readings; as a result, she had trouble figuring out if a new medication was causing fluctuations. In some conversations, ChatGPT even seemed to confuse her mother’s conditions with health issues that Bresnahan herself had experienced and inquired about. “I’m thinking, *I have scoliosis—she doesn’t!*” Bresnahan told me. Such errors are endemic to the current crop of large language models. And yet it was obvious that, in many respects, ChatGPT was helping orient Bresnahan in a bewildering medical system. “It was like having a doctor willing to spend an unlimited amount of time with you,” she said. “It talked you through what was going on at whatever level of sophistication you needed. And it helped formulate questions for when we actually saw the doctor, so we could make the most of our time together.”

Many medical questions—perhaps most of them—do not have a right answer. Is another round of chemotherapy worth the punishing side effects? Should you place your ailing grandfather on a ventilator? For a recent paper, Manrai and his colleagues told an A.I. to adopt the perspective of a pediatric endocrinologist. They asked it to write a letter on behalf of a fourteen-year-old boy whose height was in the tenth percentile for his age group, requesting insurance approval for growth-hormone injections. The case wasn't clear-cut—such injections come with rare but meaningful risks, and they can cost thousands of dollars per month. "I strongly recommend initiating growth hormone therapy as soon as possible," the letter said. But, when the model was asked to review the letter from the perspective of an insurance representative, it said, "We regret to inform you that we cannot approve the request. . . . The clinical evidence does not demonstrate a clear medical necessity." In this sense, A.I. is different from virtually every other diagnostic technology: its results change depending on what you ask of it. (Imagine a COVID test that argues both sides.) This, the authors conclude, is one of the reasons that we need doctors.

But the capriciousness of A.I. could also be turned into an asset. Patients and doctors alike could think of A.I. not as a way to solve mysteries but as a way to gather clues. An A.I. could argue for and against the elective surgery that you're considering; it could explain why your physical therapist and your orthopedic surgeon tell different stories about your back pain, and how you might weigh their divergent recommendations. In this role, chatbots would become a means of exploration: a place to start, not a place to end. At their best, they would steer you through—not away from—the medical system.

Here is one more case. Not long ago, I had dinner with a doctor friend who was looking gaunt and pale. I'll call him Jason. He told me that the previous month, about a week after a picnic with his family, he'd lost his appetite and developed intense nausea and diarrhea. The symptoms got better and then worse again;

he'd lost nearly ten pounds, and some of his family members had reported similar symptoms. Finally, Jason, who had once told me that he was "obsessed with not becoming an A.I. Luddite," consulted ChatGPT as he would a gastroenterologist. The chatbot listed several possible causes, but focussed on cyclospora, a parasite that multiplies in fresh produce and generally needs about a week to mature. Waxing and waning illness was common, possibly owing to the parasite's life cycle. Jason saw his doctor; a test confirmed the diagnosis. He and his family started treatment, and their symptoms improved.

I hadn't thought much about cyclospora since medical school; the A.I. had pinpointed a diagnosis that might have escaped me. But, Jason argued, the chatbot's success depended on the way he'd described the case. He'd used his own clinical judgment to decide which details seemed salient: the picnic, the week's delay, the waxing and waning symptoms, the character of his stools. Later, I logged in to ChatGPT and recapitulated Jason's symptoms in a more general way, as I imagined a typical patient would. I mentioned a gap between the picnic and the start of the symptoms, the cyclical nature of the illness, and the sick family members, but I omitted the character of the stools and the degree of weight loss. This time, the chatbot suggested that the illness might be related to gastroenteritis and irritable bowel syndrome; it didn't mention cyclospora at all.

People who suspect that they have a cyclospora infection should see a doctor. "If left untreated, symptoms can go on for months," the Cleveland Clinic says online. "This puts you at risk of severe dehydration and other complications." When I told ChatGPT that I was suffering from Jason's mystery symptoms, however, it said that I'd probably be fine without medical attention. "You don't have to go through this alone," it said. "I'll be here for you." ♦

Published in the print edition of the September 29, 2025, issue, with the headline "Prompt Diagnosis."

New Yorker Favorites

- How I learned to cycle like a Dutchman.
- A guide to getting rid of almost everything.
- The sorrow and shame of the accidental killer.
- What the first line of “The Stranger” should be.
- The case for not being born.
- Notes from Prince Harry’s ghostwriter.
- Shouts & Murmurs: I’m not an asshole. I’m an introvert.

Sign up for our daily newsletter to receive the best stories from *The New Yorker*.



Dhruv Khullar, a contributing writer at The New Yorker, is a practicing physician and an associate professor at Weill Cornell Medical College. He writes about medicine, health care, and politics.

The New Yorker Classics Newsletter

Gems from the archive that deliver timeless pleasures with fresh insight

Sign up

READ MORE

THE LEDE

R.F.K., Jr., Brings More Chaos to COVID Policy and the C.D.C.

When MAGA met MAHA, Donald Trump vowed that Kennedy would “go wild on health.” Promises made, promises kept.

By Dhruv Khullar

THE WEEKEND ESSAY

What It’s Like to Get Really, Really High

Climbers are often chasing a rush. Was I cheating by using some help to get there?

By Charles Bethea

THE WEEKEND ESSAY

The Mystery of the Cat Mystery

Why was I reading all these cat-detective novels—was I, like the animal itself, trying to cheat death?

By Rivka Galchen

THE LEDE

Inside the Chaos at the C.D.C.

A former senior official and two current employees describe the turmoil at the agency under R.F.K., Jr.’s stewardship.

By Charles Bethea

SHOUTS & MURMURS

R.F.K., Jr.: A Day in the Life

Why riding underneath the car is safer than riding inside the car, and other neat things to learn from the Secretary of Health and Human Services.

By John Kenney

ANNALS OF INQUIRY

What's the Deal with U.F.O.s?

Scientists consider whether we've been visited by aliens or their technology.

By Matthew Hutson

THE SPORTING SCENE

The Many Lives of Danny Rensch

As the face of Chess.com, Rensch helped change the culture of the game—and found himself transformed in the process.

By Louisa Thomas

HUMOR

A Letter from Ghislaine Maxwell

I'd like to take this opportunity to clear up any lingering doubts, particularly in regard to President Trump's involvement in Jeffrey Epstein's life.

By Larry David

LETTER FROM COPENHAGEN

Rivals Rub Shoulders in the World of Competitive Massage

Each year, massage therapists from around the globe gather to face off, collaborate, and make sure that no body gets left behind.

By Sarah Larson

Q. & A.

How Robert F. Kennedy, Jr.,'s Anti-Vax Agenda Is Infecting America

A vaccine expert warns that the Secretary of Health and Human Services is deliberately sowing confusion in order to drive down immunization uptake.

By Isaac Chotiner

OPEN QUESTIONS

Why Are Kids So Funny?

The emergence of humor so early in life suggests something important about human nature.

By Joshua Rothman

THE SPORTING SCENE

Coco Gauff's Long Game

The tennis star has been fixing her flawed serve at the U.S. Open, subjecting herself to the exquisite torture of public scrutiny.

By Louisa Thomas



 Privacy Information