


# The Daily: What 'Dr. ChatGPT' can—and can't—do, keeping drug prices down, and medical misinformation ... from doctors

Audio



On today's podcast episode, we discuss what "Dr. ChatGPT" is most likely to help with, how close it is to replacing your physician, and why it might not be ready for the patient exam room just yet. "In Other News," we talk about the US government's efforts to bring down prescription drug costs and the prevalence of health-related misinformation on social media. Tune in to the discussion with our analysts Rajiv Leventhal and Lisa Phillips.

*Subscribe to the "Behind the Numbers" podcast on [Apple Podcasts](#), [Spotify](#), [Pandora](#), [Stitcher](#), Podbean or wherever you listen to podcasts. [Follow us on Instagram](#)*



## Episode Transcript:

Marcus (00:00):

This episode is made possible by Awin. Two thirds of digital ad spend currently flows to the three big tech platforms, Google, Meta, and Amazon. But their auction-based ad models

favor their own bottom line and inflate costs at a time when every single marketing dollar counts. Awin's affiliate partnerships platform offers a real alternative to big tech and puts you back in control of your ad spend. Want to find out how? Visit [awin.com/emarketer](https://awin.com/emarketer) to learn more.

Lisa Phillips (00:25):

If anybody has ever crammed for an exam the night before, you know that you can memorize a lot of facts and figures and so on, which is what ChatGPT is good at. But when it comes to putting all those things together in a coherent way or a meaningful way when it comes to a medical diagnosis, there may be more things involved.

Marcus (00:48):

Hey gang, it's Tuesday, September 19th, Lisa, Rajiv, and listeners, welcome to the Behind the Numbers daily: an eMarketer Podcast made possible by Awin. I'm Marcus. Today I'm joined by two folks. We start with our senior analyst covering everything digital health-based out of New Jersey. It's Rajiv Leventhal.

Rajiv Leventhal (01:05):

Hey, Marcus. How are you?

Marcus (01:07):

Hey, fella. Good, sir. How are we doing?

Rajiv Leventhal (01:08):

Good.

Marcus (01:09):

Very nice. We're also joined by a principal analyst on that very team covering digital health based out of Connecticut. It's Lisa Phillips.

Lisa Phillips (01:15):

Hello, Marcus. How are you?

Marcus (01:16):

Hello there. Very well. Very well. How are you feeling today,

Lisa Phillips (01:19):

Chipper.

Marcus (01:20):

Good. Very good. Well, I've got a fact of the day, which will probably bring your mood right down. It starts off well, ends bad. From 1999 to today, global life expectancy has grown five years, so it's gone from 66 years to 71 years. Global life expectancy. Over that 20 plus year timeframe, life expectancy in the U.S. only increased by six months to reach 77 years. In the UK life expectancy jumped over three years during that time to 81, but Monaco has the highest life expectancy in the world at 86.

Rajiv Leventhal (01:58):

Of course, it does.

Marcus (02:00):

Yeah, exactly.

Lisa Phillips (02:01):

Very small sample size, I must say.

Marcus (02:03):

The smallest. Yeah, I think it's Monaco. Then Hong Kong is up there as well. At the other ends African countries, Chad, Nigeria, Lesotho are closer to 53 years life expectancy, which would be a 30-year gap. Or put another way, the life expectancy in the U.S. in 1911 was 53 years. So, 100 years ago that's what the life expectancy was in the U.S. These numbers, according to Our World in Data. I told you I was going to bring your mood right down, Lisa, I'm so sorry.

Lisa Phillips (02:34):

Thank you. Thank you, Marcus. I always count on you.

Marcus (02:37):

Off to a positive start. Anyway, today's real topic what Dr. ChatGPT can and can't do. In today's episode, first in the league we'll cover how close we are to ChatGPT being your physician. Then for another news, we'll discuss trying to get the cost of drugs down and doctors spreading misinformation on social media. We start with the lead, of course, is ChatGPT nearly ready to be your doctor? Ryan Heath of Axios writes that doctors are grappling with questions about what counts as an acceptable success rate for AI-supported diagnosis, and whether AI reliability under controlled research conditions will hold up in the real world. He cites a new study from Mass General Brigham testing ChatGPT's performance on textbook drawn case studies, which found the AI bot achieved 72% accuracy in overall clinical decision making. For example, things like making the final diagnosis and care decisions. Lisa, you write that AI agents, things like ChatGPT, are passing professional exams for doctors, nurses, dieticians, other healthcare people, but they're not ready for the patient exam room just yet. How come?

Lisa Phillips (03:56):

Well, if anybody has ever crammed for an exam the night before, you know that you can memorize a lot of facts and figures and so on, which is what ChatGPT is good at. But when it comes to putting all those things together in a coherent way or a meaningful way when it comes to a medical diagnosis, there may be more things involved I'll say. In that same study ChatGPT was strongest when making a final diagnosis, meaning it had a lot more information and its lowest accuracy at 61% or 60% was when it was making a correct initial diagnosis. So, this all comes down to as the Mass General Brigham researchers noted that it depends on what data these bots are being trained on for their large language modules and so on. And that's a very suspect thing too, which is one reason why a lot of doctors and everyone in the healthcare industry is a little worried about just letting a bot or ChatGPT make a diagnosis.

Rajiv Leventhal (04:49):

To your point, Lisa, the researchers I think said that, the physicians in that Mass General study, the physicians value their expertise is really at its highest when and adding the most value when it's in the early stages of patient care. And then when the AI agent gets more clinical information fed into it, that's when it can be more of an assistant. That's why I think it's important to distinguish between how it performed in the initial diagnosis versus the final diagnosis.

Marcus (05:15):

A few other examples of how it's being used to diagnose things. A study across over 170 hospitals in America in the Netherlands found that a machine learning model called LDA ICU could identify the illness severity of older adults in intensive care units, and determine who needed greater or earlier attention. And Marc Succi, executive director at Mass General Brigham's Innovation Incubator is saying, "AI could be used for ER triage as well." But I think he hit the nail on the head there, Rajiv used as an assistant, used to check over a doctor's work. Maybe used initially, similar to when you go into an urgent care clinic and you have a more junior doctor who sees you first or junior physician sees you first, and then that diagnosis gets signed off on by a more senior member of staff. It does seem like people are pretty trusting of AI making a diagnosis though already, 64% of U.S. adults, two in three said they would trust-

Lisa Phillips (05:15):

I challenge that though.

Marcus (06:13):

... a diagnosis made by AI over a human doctor from Innerbody research survey. How come?

Lisa Phillips (06:19):

Well, because that was very misleading to small study from Innerbody research. Because when I really dug into that data, it came down to when they were asked in a different way, "Are you concerned about the accuracy of the diagnosis?" Yeah, that was really a lot higher. 53% were concerned about the accuracy of the diagnosis. I'm not sure how the questions were asked, so that's why I think this study really shows that there's very high numbers of people like Gen Z. 82% of them said that they would trust a diagnosis made by AI over a human doctor. Later on they're asked, "Would you let it perform surgery?" "No."

Marcus (06:53):

Right.

Lisa Phillips (06:53):

I mean, the closer you get to your real condition and your real body, in general, yeah, it sounds fine. Yeah, AI knows better than a doctor, but.

Marcus (07:00):

I guess they want to trust it more as well, because it's easier to ask the question versus the doctor. So you want that easier, more convenient search for information too, because to the truth, but to your point, Lisa, when push comes to shove, less likely to accept what the generative AI is telling you versus a real doctor perhaps.

Lisa Phillips (07:16):

And that's where its expertise really is right now anyway, is in drafting emails for doctors to send to patients and so on. Because this other study, a bunch of licensed healthcare professionals were asked to rate some responses that ChatGPT questions people posted to a Reddit board where some doctors had volunteered to respond, and the doctor's responses were rated way less acceptable as than ChatGPT. But ChatGPT went on as like, "Oh yes, poor baby. It's really not good to have swelling in that area. You should do this." And whereas the doctor was like, "Yep, looks like a pilonidal cyst or something, go see a real doctor." And a lot of people, that's the kind of response they get from their doctor when they visit, whereas ChatGPT will just, "Oh, yes," empathize. It has all the time in the world to do that. It does it in a few seconds, whereas a doctor might take a minute just to say, "Yeah, that's what it is. Here's a prescription. See you later."

Marcus (08:07):

Rajiv, you had something?

Rajiv Leventhal (08:08):

Yeah, I was just going to say on the consumer piece, it really depends on the rub of these surveys is, what's the sample? How are the questions being framed? We covered another survey from Pure Research that surveyed 11,000 U.S. adults, and they had the opposite feedback from what you just said, Marcus, from this other study that two thirds said that they're uncomfortable if their healthcare provider relied on AI for diagnosing disease or recommending treatments, whereas one third or a little bit over a third said that they would be comfortable. It's interesting though, because even a third or 35%, whatever it was, even that many people saying that they would be comfortable is interesting to unpack, because that's still a lot of people that say that they'd be comfortable with AI in this type of a role. But it really does depend on the type of survey that you're referring to, because you're going to get different numbers.

(08:56):



And it also depends on where in the world you are, because there's this other study that just came out, *The Clinician of the Future* from Elsevier, and it shows that Chinese doctors are really all set to use AI, whereas doctors in the UK and the U.S. are far less ready to trust it. And given that there's so much bias, even in a clinical research, these days still, you don't know what the models have been trained on. Doctors are still puzzled about why a certain kind of dementia hits black people way harder and way more aggressively than white people. But the fact is, they've only studied mostly white people. They need to get more black people into these studies, and given how specific some of these conditions are that they're trying to study, that in itself is a tough lift.

Marcus (09:40):

Yeah, it does depend on where you are, how great the demand is for physicians, because if you are in a more desperate situation, maybe you're more likely to use these types of artificial intelligence chatbots, because you don't have enough doctors to go around. Also, depends what country you're in based on what the standard is in that country, how easy it is or maybe not easy it is to get these things up and running according to the government says the standard should be. So Lisa, it was surprising to see, even though folks to some of the research, Rajiv was just citing from Pure Research said that they wouldn't trust ChatGPT generative AI to give them a diagnosis. You did write an article saying that people did prefer them in some instances to real-life doctors when it came to bedside manner.

Lisa Phillips (10:25):

Yeah, what I was just mentioning is that yes, they ChatGPT shows a lot more empathy than doctors do. And I will say, I have been fired by doctors as a patient because I was questioning. I would say, "Well, you didn't check that off on the blood panel, but I did because I wanted to know." Like, what the? I mean, human doctors have a lot shorter patience span than ChatGPT does.

Marcus (10:49):

Yeah, I guess there's less ego involved as well. If you've trained for 12 years or more than that, 20, 30 years and someone comes in there and questions you, then human doctors may be more prone to say, "Hang on a second. I know what I'm talking about. You are the patient. Listen to me."

Lisa Phillips (11:03):

Exactly.

Marcus (11:03):

Yeah, it was staggering to see. So, folks preferred chatbots' responses nearly 80% of the time according to the study in JAMA Internal Medicine. The other responses being judged on the quality of information provided and the empathy or bedside manner provided as well. It seems like chatbots or Dr. ChatGPT could help with bedside manner to a certain extent. Taking notes, AI can assist doctors with low-risk early stage care, AWS HealthScribe, Amazon Web Services HealthScribe being used to take medical notes during patient office visits. We've talked about that before. Administrative tasks like folks trusting chatbots to answer logistical questions, schedule appointments, questions they have on insurance coverage, which can be incredibly tricky.

(11:44):

Software advice was noting this number one application for generative AI and health technology according to U.S. healthcare professionals was electronic medical record documentation. And also bringing the cost down. So 2021, America spent 18% of GDP on healthcare, nearly twice as much as the average advanced economy. The cost of using an AI agent was 20 cents an hour versus 20 to \$90 an hour for a human nurse dietician healthcare coder, according to Citi Global Insights, so cheaper in a lot of instances.

Rajiv Leventhal (12:15):

And it comes down to all the aspects that you mentioned, like documenting, note-taking, writing appeals to health insurers for prior authorizations if they deny coverage for a prescription. These are manual processes that we pay, or not we, but the healthcare system pays humans to do. It could be cheaper and more efficient with an AI agent, but these are not high-stakes clinical decisions that we're talking about.

Lisa Phillips (12:37):

It's not a bot sitting there telling you that you have cancer.

Marcus (12:40):

Right, right. I mean, these aren't high-stakes clinical decisions, but I'm wondering whether we need ChatGPT to get there before it's useful. I guess ChatGPT, you can argue doesn't have to

be as good as your doctor to be useful, because you have Dr. ChatGPT, versus doctor just Google something, versus Dr. WebMD, versus Dr. TikTok, versus Dr. Human. And so there are a lot of ways and places that people are getting medical information, and so if it can be better than some of those other ways, but maybe not yet as good as the physician, do you think it's still got some use there?

Rajiv Leventhal (13:12):

I think that's right. I mean, better than Google, but maybe not significantly better is a good way to frame it. But what does that mean? Does that mean that it should be relied on for treatment and diagnosis? I don't know. Would you rely on Google? We've talked about that on this podcast before. Perhaps, but you'd have to be careful about the information that you're seeing online. It's not all vetted and from authoritative sources. So, at this point, that's why I see it as an assistant, as an up and coming tool, but certainly not one that should run the show from a clinical standpoint.

Lisa Phillips (13:45):

I got to say, this Elsevier study of 2,700 some physicians around the world, this is only the second year they've done it. But last year they asked about AI tools and so on, and doctors said, "Yeah, we think we'll be using them, but maybe in about 10 years." This year when they did the study, many more doctors said they see this in a two to three year timeframe. And 55% of them said they think telehealth visits will be the norm, instead of in-person doctor visits and so on in the future, just a few years from now, like 2028. I think we may see more uses in the near future for ChatGPT in healthcare.

Marcus (14:19):

Yeah. Yeah, whether it's ChatGPT, again, doesn't need to maybe replace your physician, but play a meaningful role in some capacity. There's a few folks who believe that it can. Assistant Professor at Emory Eye Center, Nieraj Jain says that, "ChatGPT is definitely an improvement over just putting something into a Google search bar and seeing what you find." In June Emory University School of Medicine noted that ChatGPT compared quite well to human doctors who reviewed the same symptoms and much better than WebMD's symptom checker. Actually, WebMD is now partnering with startup HIA Technologies to provide interactive digital health assistance. And then also AI consultants James Benoit, who was a postdoctoral fellow in nursing at the University of Alberta, Canada, published a study in February reporting

that ChatGPT significantly outperformed online symptom checkers in evaluating a set of medical scenarios.

Rajiv Leventhal (15:12):

I was just going to say, this is timely, I think for our listeners. There was a story this week, it was actually on the TODAY Show. It was a young boy and he was in chronic pain, and his mom took him to 17 different specialists, couldn't figure out what was going on. The thought was that it had something to do with airway obstruction, but couldn't come up with a diagnosis that was relevant to what the pain was.

(15:32):

And the mother then went to ChatGPT and plugged in all of these symptoms that her son was having. Even all the MRI imaging notes, all these other details. And lo and behold, ChatGPT suggested that the kid was suffering from this neurological condition called tethered cord syndrome. And that turned out to be the correct diagnosis, and it enabled the family to find the right specialist and access the appropriate care. And I saw a physician commenting on it saying, "That the information to make that diagnosis was actually there in the MRI imaging, the specialist, the radiologist just missed it." That kind of story is interesting, because you think about humans versus technology, and of course, ChatGPT is going to hallucinate in many other situations, and we're not going to hear about those failures, but pretty interesting that it was able to succeed in this instance.

Marcus (16:17):

Yeah, using AI tools to analyze medical images, that's a really good use case. I want you to take on this quick, folks before we start to close out the lead. Marc Succi, so a guy I mentioned from Mass General Brigham's Innovation Incubator was saying, "That there are actually no benchmarks on how good ChatGPT needs to be." And was basically saying, "That ChatGPT is starting to exhibit capabilities of a newly graduated doctor." But since there are "no real benchmarks for success rates among doctors at different levels of seniority, judging whether AI is adding value to a doctor's work remain complicated." Asking, "Where do diagnostic success rates need to rise up to?" Because you can use ChatGPT to take the same test to pass an exam to become a doctor, but once you're a doctor, no one actually knows how often those doctors are getting the diagnosis right or not.

Lisa Phillips (17:07):

Exactly. I laughed when you first said it, because it's like, "Yeah, and who's judging the doctors?" I mean, 100% of doctors think they're above average. We know that isn't true.

Marcus (17:19):

Yeah, it seems similar to driverless cars. The margin of error is smaller for robots, just because they're newer. 40,000 people die from car accidents every year because of humans, 110 car accident deaths that equates to each day because of humans. But one driverless car accident will make the headlines immediately because of the newness and the technology and because the bar to acceptance is so much higher when we could really point to humans and say, "Well, we're doing pretty terrible job as is, and so actually the bar doesn't need to be as high to make a meaningful impact." That's it for the lead, time now for the halftime report. Folks, real quick, one thing that you think is most worth repeating from the first half? Lisa, I'll start with you.

Lisa Phillips (18:01):

I'll just say, I'd rather get an email from a ChatGPT bot than I would a diagnosis.

Marcus (18:06):

Rajiv?

Rajiv Leventhal (18:07):

Physicians aren't trusting this technology yet, but if you think about it, ChatGPT was what, introduced still less than a year ago. I think it's made pretty impressive headway, and it'll be interesting to see where things stand a year from now.

Marcus (18:20):

Yeah, a ton of issues to iron out. Bias, liability, lack of regulatory oversight, transparency, privacy, all the rest of the things. But John A is actually a computational epidemiologist from University of California, San Diego was saying, "100 million people have ChatGPT on their phone and are asking questions right now. People are going to use chatbots with or without us."

(18:41):

As we've got time for for the lead, time for the second half of the show. Today in other news, trying to get the cost down for the top 10 most expensive prescription drugs, and the

prevalence of health misinformation on social media from doctors.

(18:55):

Story one, Lisa, you write that the Centers for Medicare & Medicaid Services, CMS, just released its highly anticipated list for the first 10 prescription drugs covered by Medicare Part D subject to price negotiations. You point out that over 8 million Americans with Medicare Part D coverage use these drugs to treat a range of conditions, from cardiovascular disease to cancer. The Inflation Reduction Act is what we're talking about. That was passed in 2022 and told the CMS to choose drugs with the highest total Part D costs and negotiate prices directly with their makers, with the drug makers. But Lisa, the most interesting takeaway from your article on this is what?

Lisa Phillips (19:36):

That some of the drugs that were on the list really surprised the industry, the big pharma industry. And big pharma had already lined up a whole lot of lawsuits based on which drugs they thought were going to be on the list. Of course, Bristol Myers Squibb, their Eliquis blood thinner was on the list, because it costs 16.4 billion in just one year and spending on Medicare beneficiaries who needed it. But as soon as the list came out, a couple of the drug companies said, "Oh, we're not on it." They dropped their suits. And none of these prices will take effect until once they're negotiated over the next year. They won't take effect until January 1st, 2026. And then CMS is going to come out with another list next year, and it's going to have more. In five years we're going to have at least 50 to 60 drugs on this list, so there will be more fights. They're already going to court right now. But that said, it's good, but it may not make that big of a dent in overall consumer drug spending in the near future.

Marcus (20:30):

Yeah, and we'll certainly take a long time. Exactly. Story two, Rajiv, you recently wrote that some licensed physicians contributed to the spread of medical misinformation during the pandemic according to a study recently published in JAMA Open Network. Misinformation in this instance was defined as content that contradicted CDC guidelines for COVID-19 prevention or treatment. And you explained that researchers found over 50, 50 physicians in 28 different specialties across all regions of the U.S. disputing vaccine safety and effectiveness. Endorsing unproven medical treatments, arguing that masks don't work, things like that. Rajiv, the most interesting takeaway from your ask call was what?

Rajiv Leventhal (21:10):

Well, this is one reason why consumer trusts in the healthcare system and in authoritative physician experts is eroding. The findings of this study, I think back that up. In 52 doctors posting misinformation about COVID and other health issues, that might seem like a very small amount to some people, but keep in mind that this study didn't look at social media posts in 2020, which is when COVID started and misinformation was really spreading like wildfire. But these are a trusted authority figures, and now social media platforms need to step up and play a role in vetting posts and information from clinicians who patients do lean on. Because not doing so has serious repercussions. And we're seeing YouTube, we've written about YouTube doing exactly this, and hopefully other social media platforms will follow suit.

Marcus (21:56):

Yeah, at that point, while it's 50 doctors, it's not that many people. 50 doctors, say they've got 20,000 followers each. They probably have more, but-

Rajiv Leventhal (22:05):

I think that was in the study, combined they had over like a 100,000, so yeah.

Marcus (22:08):

Oh, wow. Well, I just did a quick back of the napkin, I mean, 100,000. That's insane. Let's pretend that the low end had 20,000 followers each times 50. That's a million people right there, which you are able to influence. I also thought it was so sobering and just heartbreaking that you highlighted researchers said about one third of the 1.1 million COVID-19 related deaths in the U.S. as of January 2023 were considered preventable. One third of those 1.1 million deaths preventable if public health recommendations had been followed.

Rajiv Leventhal (22:44):

And these were the same recommendations that were being refuted by people on social media, so yeah.

Marcus (22:49):

Yeah, it's absolutely heartbreaking. That's what we've got time for today's episode. Thank you so much to my guests. Thank you to, Lisa.

Lisa Phillips (22:54):

Thanks, Marcus.

Marcus (22:56):

Thank you to Rajiv.

Rajiv Leventhal (22:57):

Thanks, Marcus.

Marcus (22:58):

And thank you to Victoria who edits the show, James who copy edit it, and Stewart, who runs the team. Thanks to everyone for listening into the Behind the Numbers Daily: an eMarketer Podcast made possible by Awin. You can tune tomorrow if you want to hang out with Sara Lebow on the Reimagining Retail show as she speaks with senior analysts, Blake Roche and Corina Perkins about, how the way Gen Z is discovering and shopping for CPG brands is changing.