

AI came close to residents, medical students with clinical reasoning in studies

July 19 2023, by Amy Norton



ChatGPT may have some of the reasoning skills doctors need to

diagnose and treat health problems, a pair of studies suggests—though no one is predicting that chatbots will replace humans in lab coats.

In [one study](#), researchers found that—with the right prompting—ChatGPT was on par with [medical residents](#) in writing up a patient history. That's a summary of the course of a patient's current health issue, from the initial symptoms or injury to the ongoing problems.

Doctors use it in making diagnoses and coming up with a treatment plan.

Recording a good history is more complicated than simply transcribing an interview with a patient. It requires an ability to synthesize information, extract the pertinent points and put it all together into a narrative, explained [Dr. Ashwin Nayak](#), the lead researcher on the study.

"It takes [medical students](#) and residents years to learn," said Nayak, a clinical assistant professor of medicine at Stanford University, in California.

Yet, his team found that ChatGPT was able to do it about as well as a group of medical residents (doctors in training). The catch was, the prompt had to be good enough: The chatbot's performance was decidedly subpar when the prompt was short on detail.

ChatGPT is driven by [artificial intelligence](#) (AI) technology that allows it to have human-like conversations—instantly generating responses to just about any prompt a person can cook up. Those responses are based on the chatbot's "pre-training" with a massive amount of data, including information gathered from the internet.

The technology was launched last November, and within two months it had a record-setting 100 million monthly users, according to a report

from the investment bank UBS.

ChatGPT has also made headlines by reportedly scoring high on SAT college entrance exams, and even passing the U.S. medical licensing exam.

Experts warn, however, that the chatbot should not be anyone's go-to for [medical information](#).

Studies have pointed to both the technology's promise and its limitations. For one, the accuracy of its information depends in large part on the prompt the user gives. In general, the more specific the question, the more reliable the response.

A recent study focused on [breast cancer](#), for example, found that ChatGPT often gave appropriate answers to the questions researchers posed. But if the question was broad and complex—"How do I prevent breast cancer?"—the chatbot was unreliable, giving different answers each time the question was repeated.

There's also the well-documented issue of "hallucinations." That is, the chatbot has a tendency to make stuff up at times, especially when the prompt is about a complicated subject.

That was borne out in Nayak's study, which was published online July 17 as a research letter in *JAMA Internal Medicine*.

The researchers pitted ChatGPT against four senior medical residents in writing up histories based on "interviews" with hypothetical patients. Thirty attending physicians (residents' supervisors) graded the results on level of detail, succinctness and organization.

The researchers used three different prompts to set the chatbot on the

task, and results varied widely. With the least-detailed prompt—"Read the following patient interview and write a [history]. Do not use abbreviations or acronyms"—the chatbot fared poorly. Only 10% of its reports were considered acceptable.

It took a much more detailed prompt to nudge the technology to a 43% acceptance rate—on par with the residents. In addition, the [chatbot](#) was more prone to hallucinations—such as making up a patient's age or gender—when the prompt "quality" was lower.

"The concerning thing is, in the real world people aren't going to engineer the 'best' prompt," said [Dr. Cary Gross](#), a professor at Yale School of Medicine who co-wrote [a commentary](#) published with the findings.

Gross said that AI has "tremendous" potential as a tool to aid medical professionals in arriving at diagnoses and other critical tasks. But the kinks still need to be ironed out.

"This is not ready for prime time," Gross said.

In the [second study](#), another Stanford team found that the latest model of ChatGPT (as of April 2023) outperformed medical students in final exam questions that require "clinical reasoning"—the ability to synthesize information on a hypothetical patient's symptoms and history, and come up with a likely diagnosis.

Again, Gross said, the implications of that are not yet clear, but no one is suggesting that chatbots make better doctors than humans do.

A broad question, he said, is how AI should be incorporated into medical education and training.

While the studies were doctor-centric, both Nayak and Gross said they offer similar take-aways for the [general public](#): In a nutshell, prompts matter, and hallucinations are real.

"You might find [accurate information](#), you might find unintentionally fabricated information," Gross said. "I would not advise anyone to base medical decisions on this."

One of the main appeals of chatbots is the conversational nature. But that's also a potential pitfall, Nayak said.

"They sound like someone who has a sophisticated knowledge of the subject," he noted.

But if you have questions about a serious medical issue, Nayak said, bring them to your human health care provider.

More information: Ashwin Nayak et al, Comparison of History of Present Illness Summaries Generated by a Chatbot and Senior Internal Medicine Residents, *JAMA Internal Medicine* (2023). [DOI: 10.1001/jamainternmed.2023.2561](#)

Eric Ward et al, Evolving Methods to Assess Chatbot Performance in Health Sciences Research, *JAMA Internal Medicine* (2023). [DOI: 10.1001/jamainternmed.2023.2567](#)

Copyright © 2023 [HealthDay](#). All rights reserved.

Citation: AI came close to residents, medical students with clinical reasoning in studies (2023, July 19) retrieved 14 December 2023 from <https://medicalxpress.com/news/2023-07-ai-residents-medical-students-clinical.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.