

Can AI Chatbots Correctly Answer Questions About Cancer?

Artificial intelligence-backed chatbots can provide accurate info about cancer—but also include errors and omissions, two studies show.

October 17, 2023 By [National Cancer Institute](#)

Artificial intelligence (AI) technologies have become part of everyday life for many people, helping with mundane tasks such as shopping online and using social media. But can AI-driven chatbots give people accurate information about cancer and its treatment?

According to two new studies, the answer may be: Not yet. Although AI chatbots can gather cancer information from reputable sources, their responses can include errors, omissions, and language written for health care professionals rather than for patients, researchers found.

“We’re still in the very early days of AI,” said Danielle Bitterman, MD, of the Artificial Intelligence in Medicine Program at Mass General Brigham in Boston, who led one of the studies. “AI chatbots can synthesize medical information, but they are not yet able to consistently generate reliable responses to clinical questions from patients.”

She and her colleagues [asked ChatGPT version 3.5 to describe basic treatment approaches](#) for various forms of cancer (e.g., “What is the treatment for stage I breast cancer?”).

Almost all of the chatbot responses included at least one treatment approach that was consistent with expert clinical guidelines. However, about a third of these responses had at least one recommendation that did not match the clinical guidelines.

The second study prompted four chatbots, including the same version of ChatGPT, with questions about common cancers (e.g., “What is prostate cancer?”). The [chatbots generally produced accurate information about the cancers](#), but many of the responses were too technical for the average patient, the researchers found.

Both groups published their findings in JAMA Oncology on August 24.

“These studies illustrate why we are not ready to rely on these tools or steer patients and the public to them for cancer information,” said Wen-Ying Sylvia Chou, PhD, a health communication researcher in NCI’s Division of Cancer Control and Population Sciences who was not involved in the studies.

“But AI technology is here to stay,” Dr. Chou continued. “This research opens the door to a thoughtful discussion of the benefits and harms related to the use of AI in cancer communication and cancer care.”

Testing Off-the-Shelf Chatbots

Artificial intelligence refers to a computer’s ability to perform functions that are usually thought of as intelligent human behavior, such as learning, reasoning and solving problems.

Chatbots are a type of AI known as large language models. They can interpret questions and can generate text responses that read as though they were written by a person. The models are trained on large amounts of information, such as text from the internet.

The chatbots evaluated in these studies were “off the shelf” models that presumably had not been trained on selected medical information. But newer models that have had specific health care training are being released, wrote Atul Butte, MD, PhD, who directs the Bakar Computational Health Sciences Institute at the University of California, San Francisco, [in an accompanying editorial](#).

The two studies highlight both the promise and the current limitations of some chatbots. For instance, chatbots can seamlessly combine correct and incorrect treatment recommendations, making it difficult even for experts to detect the errors, Dr. Bitterman noted.

“Large language models are trained to predict the next word in a sentence,” she said. “Their primary goal is to give a response that’s fluent and that makes linguistic sense, so we were not surprised by incorrect responses.”

Asking a Chatbot Questions About Common Cancers

Dr. Bitterman’s team focused on three common cancers—breast, lung and prostate. They developed 4 slightly different questions, or prompts, about treatment strategies for 26 distinct diagnoses for these cancers, such as early stage (or “localized”) breast cancer or advanced lung cancer.

The researchers compared the responses against the National Comprehensive Cancer Network (NCCN) guidelines for those cancers. The study used 2021 NCCN guidelines as a benchmark because the version of ChatGPT that they used, GPT-3.5-turbo-0301, was developed based on data available up to September 2021.

Three oncologists scored the 104 responses. The chatbot provided at least 1 recommendation for 102 of 104 (98%) prompts. But 35 of 102 (34.3%) of these responses also recommended 1 or more treatments that did not agree with recommendations, the researchers found.

In addition, the oncologists often disagreed about whether a ChatGPT response was correct. The researchers believe the disagreement reflects the complexity of NCCN guidelines and the fact that ChatGPT responses can be unclear or hard to interpret.

Notably, 13 responses (13%) included treatment strategies that were not part of any recommended treatment. These answers incorporated treatment strategies that did not appear at all in the guidelines or were nonsensical, which AI researchers refer to as “hallucinations.”

Putting Four Chatbots to the Test

The second study evaluated 100 responses to prompts about the 5 most common cancers (skin, lung, breast, colorectal, and prostate) from the chatbots ChatGPT-3.5, Perplexity, Chatsonic, and Bing AI.

The prompts, such as “colorectal screening,” “breast cancer signs,” and “melanoma,” were based on the top Google search queries for these cancers during 2021–2022.

Three of the chatbots had responded with cancer information from reputable sources, including American Cancer Society, Mayo Clinic, NCI, and the Centers for Disease Control and Prevention, the researchers found.

But chatbots may not be able to explain complex medical concepts using text alone, the researchers said. They noted that concepts such as swollen lymph nodes, for example, might be difficult to explain without diagrams or visual aids.

Some responses were written at a college reading level, the researchers cautioned. They plan to test prompts that might help get more accessible responses in future studies.

“Many of the chatbot responses were too complex for the average patient to understand,” said study leader Abdo Kabarriti, MD, of the State University of New York Downstate Health Sciences University.

“Chatbots can be very smart,” Dr. Kabarriti continued. “But when it comes to answering a patient’s questions about cancer, they are not a replacement for doctors.”

If chatbots become more accurate, a more appropriate role for the technology, the researchers suggested, would be as an additional resource for people with cancer and their family members.

Many patients hear so much information at their medical appointments that it can be difficult to remember everything, said Dr. Kabarriti. Chatbots could potentially be useful by “re-emphasizing” some of the general ideas from these discussions, he added.

Expanding Medical Applications for Chatbots

The new studies “are among the first to explore how large language models can be used in cancer research and care,” Dr. Butte said, adding that he expects many more to follow.

Chatbots have already shown promise in health communication. In one study, researchers found that AI-based tools [wrote empathetic responses to patients’ questions](#) that had been posted in an online forum. The findings, they said, suggest that AI might be able to help busy physicians draft email responses to patients.

And after [ChatGPT-3.5 recently passed parts of a medical licensing exam](#), some researchers suggested that large language models may be able to assist with medical education and potentially with clinical decision-making.

But chatbots might be better able to answer questions on a licensing exam than to respond to medical questions asked by individual patients, Dr. Bitterman noted. Her concern is that patients might take advice from a chatbot at face value.

“The promise of AI technology is really exciting, but we cannot compromise on patient safety,” she said. “We have to take the time to optimize these large language models and evaluate their performance and safety now.”

Developing New AI Systems for Seeking Cancer Information

AI technology is changing so rapidly that some of the chatbots the researchers evaluated may already be out of date.

Just as AI evolves, cancer treatment recommendations can also change. The organizations and companies developing AI technology need to figure out how to ensure that chatbots draw on the latest medical knowledge, several experts said.

Learning more about how AI is currently being used to search for cancer information would help inform decisions about future AI systems, noted Dr. Chou.

And as these systems are built, she continued, developers need to be aware that decisions such as which data are used to “train” the language models can have consequences. If those data exclude certain patient groups, for instance, that bias will be reflected in the responses.

“As a community, we need to develop benchmarks and standards for how we are going to evaluate these models as we start moving toward more clinical applications,” Dr. Bitterman said.

A Chatbot Responds to a Question About Cancer Treatment

When Dr. Bitterman and her colleagues asked ChatGPT-3.5 “How do you treat stage III breast cancer?” the chatbot responded:

“As an AI language model, I do not have the ability to treat breast cancer. However, in general treatment of Stage III breast cancer may include: 1. Surgery: This involves removing the tumor and surrounding tissue, and possibly removing lymph nodes from under the arm. 2. Radiation therapy....”

AI’s Potential to Expand Access to Information about Best Practices in Cancer Care

Chatbots may be imperfect, but they have the potential to help expand access to knowledge about cancer far beyond the leading centers, according to Dr. Butte.

He envisions a day when chatbots are trained on clinical data from top cancer centers. Digital health tools powered by AI could then share recommendations on cancer diagnostics and treatment with any doctor who has internet access. AI could even help explain the care to patients, Dr. Butte suggested.

“I believe that AI and machine learning models could help expand the top level of cancer care to reach many more people,” he said.

[This post](#) was originally published October 3, 2023, by the National Cancer Institute. It is republished with permission.