

New study finds that ChatGPT 4 excels at picking the right imaging tests

June 22 2023



Credit: CC0 Public Domain

A new study by investigators from Mass General Brigham has found that artificial intelligence (AI) language models like ChatGPT can accurately identify appropriate imaging services for two important clinical

presentations: breast cancer screening and breast pain. Their results suggest that large language models have the potential to assist decision-making for primary care doctors and referring providers in evaluating patients and ordering imaging tests for breast pain and breast cancer screenings. Their results are published in the *Journal of the American College of Radiology*.

"In this scenario, ChatGPT's abilities were impressive," said corresponding author Marc D. Succi, MD, associate chair of Innovation and Commercialization at Mass General Brigham Radiology and executive director of the MESH Incubator. "I see it acting like a bridge between the referring healthcare professional and the expert radiologist—stepping in as a trained consultant to recommend the right imaging test at the point of care, without delay. This could reduce administrative time on both referring and consulting physicians in making these evidence-backed decisions, optimize workflow, reduce burnout, and reduce patient confusion and wait times."

ChatGPT is a large language model (LLM) built on data from the internet to answer questions in a human-like way. Since ChatGPT was introduced in November 2022, researchers worldwide are diving into learning how these AI tools can be used in medical scenarios. Published as a [preprint](#) on February 7, 2023, this study is the first of its kind to test ChatGPT's clinical [decision-making](#) abilities, and the first to test GPT 4 as opposed to older iterations.

When a primary care doctor orders specialized testing, say for a patient who complains of breast pain, they may not know the best imaging test to choose. It might be an MRI, an ultrasound, a mammogram, or another imaging test. Radiologists generally follow the American College of Radiology's Appropriateness Criteria to make these decisions. These evidence-backed guidelines are well-known to specialists, but less known for non-specialists who many need to pick the best imaging test during a

patient's visit. This can cause confusion on the patient's side and can lead to patients getting tests they don't need or getting the wrong tests.

The researchers asked OpenAI's ChatGPT 3.5 and 4 to help them decide which imaging tests to use for 21 made-up patient scenarios involving the need for [breast cancer screening](#) or the reporting of breast pain using the appropriateness criteria.

They asked the AI in an open-ended way and by giving ChatGPT a list of options. They tested ChatGPT 3.5 as well as ChatGPT 4, a newer, more advanced version. ChatGPT 4 outperformed 3.5, especially when given the available imaging options. For example, when asked about breast cancer screenings, and given multiple choice imaging options, ChatGPT 3.5 answered an average of 88.9% of prompts correctly, and ChatGPT 4 got about 98.4% right.

"This study doesn't compare ChatGPT to existing radiologists because the existing gold standard is actually a set of guidelines from the American College of Radiology, which is the comparison we performed," Succi said. "This is purely an additive study, so we are not arguing that the AI is better than your doctor at choosing an imaging test but can be an excellent adjunct to optimize a doctor's time on non-interpretive tasks."

Integrating AI into medical decision making could happen at the point of care. When a primary care doctor enters data into an electronic health record, the program could alert them to the best imaging options—providing an answer to the patient for what to expect when they go for the test and suggesting to the doctor the right test to order.

Researchers added that a more advanced medical AI could be created using datasets from hospitals and other [research institutions](#) to make it more specific to health-focused applications.

"We may be able to fine-tune ChatGPT with different patient and therapeutic data and knowledge sets to tailor it to specific patient populations," Succi said. "At Mass General Brigham, we have specialized centers of excellence where we care for patients with some of the most complex and rare diseases. We can leverage our experience and lessons learned from caring for these patient cases to train a model to provide support for rare and complex diagnoses and then make that model available to centers around the world, especially centers that may treat these conditions less frequently."

But before any AI would be involved in medical decision-making, it would need to be extensively tested for bias, privacy concerns, and approved for use in medical setting. New regulations around medical AI could also play a big role in what makes it into patient care interactions.

More information: Arya Rao et al, Evaluating GPT as an Adjunct for Radiologic Decision Making: GPT-4 Versus GPT-3.5 in a Breast Imaging Pilot, *Journal of the American College of Radiology* (2023).
[DOI: 10.1016/j.jacr.2023.05.003](https://doi.org/10.1016/j.jacr.2023.05.003)

Provided by Mass General Brigham

Citation: New study finds that ChatGPT 4 excels at picking the right imaging tests (2023, June 22) retrieved 28 April 2024 from <https://medicalxpress.com/news/2023-06-chatgpt-excels-imaging.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.