

Researchers Customize AI Tools for Digital Pathology

A refined model of ChatGPT could provide accurate responses to questions about digital pathology and compile detailed results.

July 25, 2024 By Dana-Farber Cancer Institute

Scientists from Dana-Farber Cancer Institute and Weill Cornell Medicine have developed and tested new artificial intelligence (AI) tools tailored to digital pathology—a relatively new field that uses high-resolution digital images that are created from tissue samples to diagnose diseases and inform treatment decisions.

Their paper, [published in The Lancet Digital Health](#) on July 9, 2024, demonstrates that ChatGPT, an AI language model developed to understand and generate text, can be tailored using an artificial intelligence technique called retrieval-augmented generation, to provide accurate responses to questions about digital pathology and compile detailed results. The authors also found that ChatGPT can help pathologists, without extensive coding experience, use complex software that analyzes tissue samples, effectively helping bridge the gap between pathology skills and digital pathology skills.

ChatGPT is a large language model (LLM), meaning it generates text on a wide range of topics using extensive amounts of data. “LLMs are good for general tasks, but they aren’t the best tools for getting useful information for specialized fields,” said the study’s lead author, Mohamed Omar, MD, assistant professor of research in pathology and laboratory medicine and a member of the Division of Computational and Systems Pathology at Weill Cornell Medicine, and lead scientist at Dana-Farber Cancer Institute.

To create AI tools that could increase the efficiency and precision of digital pathology, corresponding author Renato Umeton, PhD, director of Artificial Intelligence Operations and Data Science Services, Informatics & Analytics Department at Dana-Farber, spearheaded the effort to customize and augment ChatGPT capabilities for this specific purpose.

Boosting AI Accuracy for Pathology

“There are two major problems with general LLMs. First, they often provide lengthy generic responses that, while correct, don’t contain detailed enough information,” Omar said. “Second,

these models can hallucinate and make things up out of nowhere, including literature citations. This is especially bad in specialized fields like digital pathology and cancer biology, for example.”

To address the glitches, Umeton started with a safe, private, and secure ChatGPT variant operationalized at Dana-Farber ([GPT4DFCI](#)). They augmented GPT4DFCI with access to a comprehensive and curated database of the latest developments in digital pathology, consisting of 650 publications from 2022 onward, which added up to over 10,000 pages of literature. “We could ask this new system to catch us up on many specific topics or techniques in digital pathology and get results in seconds, with a level of detail, depth, and summarization that does not exist in current scientific literature tools or search engines. This effectively augmented researchers’ capabilities,” Umeton noted.

They used a technique called retrieval-augmented generation (RAG), which enabled GPT4DFCI to access relevant documents or information from this specialized database and generate accurate responses to user prompts about digital pathology, but nothing outside of that realm.

Omar and his colleagues asked GPT4DFCI questions specific to digital pathology and compared the responses to those provided by ChatGPT 4. By requiring GPT4DFCI to provide links to the specific publications it used to generate responses, they determined that the answers were accurate and grounded. The refined model provided more precise and relevant answers than ChatGPT 4 and did not hallucinate—not even once. “My hope is that this will be a catalyst for more domain-specific tools in other fields of medicine or medical research,” Omar said.

AI Provides a Helping Hand with Coding

The second AI program the team developed helps pathologists use PathML, a specialized software library that requires familiarity with the programming language Python for analyzing vast and complex pathology image datasets. “Pathologists or scientists without prior coding experience could find PathML challenging to use for image analysis tasks,” Omar said.

The researchers integrated PathML with ChatGPT, allowing users to interact with PathML documentation through the chat function. Users can simply type in their questions about using PathML for analysis of histopathology images—of multiplex images, of tissue-microarray, or for biomarker quantitative assessments, for instance—and the tool will provide step-by-step, accurate instructions on coding their analyses.

“Generative AI has proved useful in providing structured guidance around what material to consult and how to organize the learning journey for new topics,” Umeton said. “Our research shows that, when combined with the proper information retrieval techniques, ChatGPT and safeguarded AI tools, like GPT4DFCI, can be extremely effective in supporting basic researchers. These tools are helpful even across very complex topics that need extremely precise answers, like digital pathology.”

This work was supported by the National Cancer Institute grants P50CA211024, P01CA265768,

and U54CA273956.

[This news release](#) was originally published July 10, 2024, by Dana-Farber Cancer Institute. It was adapted from a press release issued by Weill Cornell Medicine.

© 2026 Smart + Strong All Rights Reserved.

<http://beta.docker.cancerhealth.com/article/researchers-customize-ai-tools-digital-pathology>