

Editors: Raymond D. Aller, MD, & Dennis Winsten

How labs can make the most of ChatGPT and other LLMs

February 2024—The key to using ChatGPT and other large language models effectively in pathology is understanding not only what they are designed to do but, just as importantly, what they are not designed to do, says Eric Glassy, MD, medical director at Affiliated Pathologists Medical Group, Rancho Dominguez, Calif., and past chair of the CAP Information Technology Leadership Committee.

The models should not be thought of as databases that retrieve facts but as predictors that generate predictions that are often right and sometimes wrong, says Dr. Glassy, who conducted a presentation on ChatGPT and other large language models at CAP23. LLMs are designed to identify answers with the highest probability of satisfying users, he explains. “That’s an answer you would like to have but not necessarily the true answer.”

Incorrect predictions, or hallucinations, generated by LLMs can pose numerous risks to the practice of pathology, Dr. Glassy says. These risks can be linked to models failing to recognize the difference between public and private information, demonstrating racial or ethnic bias, and providing medical information that could potentially be harmful.

While it’s important to verify answers generated by LLMs, pathologist users of the technology can take steps to minimize hallucinations and steer models toward providing more accurate and appropriate answers using offerings such as the following.



Dr. Glassy

GPT-4 versus GPT-3.5. GPT-4, the latest version of OpenAI’s large language models, provides more accurate and coherent answers than its predecessor, GPT-3.5, but it is also more expensive. Yet GPT-4 may be worth the subscription cost because incorrect answers could lead to improper medical treatment. For example, when GPT-3.5 was asked how to treat a pregnant woman who had contracted Lyme disease, it suggested tetracycline, which is effective at treating the disease but can cause a range of developmental abnormalities in a fetus, Dr. Glassy says. GPT-4, on the other hand, correctly identified amoxicillin as the treatment that would effectively and safely treat the disease in a pregnant woman.

GPT-4 Turbo, the latest version of the software, is available for \$20 per month. It allows users to expand prompts to approximately 300 pages of text, generate images from a text prompt using DALL-E technology, accept images as inputs, and perform text-to-speech conversion, among other tasks. Those who do not want to pay the subscription fee may want to check out Microsoft’s Bing Chat in creative mode, adds Dr. Glassy. The latter uses GPT-4 and is available at no cost. The Google Bard and Microsoft Copilot artificial intelligence chatbots are also available at no charge.

Prompts. How a query is written can make a big difference in how LLMs respond. The results are noticeable enough that Boston Children’s Hospital hired an artificial intelligence prompt engineer to help physicians and other hospital employees query LLMs more effectively, Dr. Glassy says.

Some prompts improve the accuracy of LLM responses by targeting algorithms. Telling ChatGPT (the application

powered by GPT AI models) to provide a step-by-step answer, for example, or asking it to request three questions before providing an answer can guide it toward a more sequence-based approach to processing a query, which tends to reduce hallucinations, Dr. Glassy says. Even asking an LLM to “slow down and take a deep breath” before responding has been shown to result in more deliberate and accurate answers, he adds.

Asking ChatGPT to provide a confidence score for an answer, with zero being not confident and 100 being very confident, can have a similar effect, Dr. Glassy says. However, users should verify all responses, even those with high confidence scores, through other sources.

Other prompt techniques involve narrowing the focus of the question to elicit more specific information. Asking for heart disease symptoms, for example, would generate a wealth of information, but requesting the top five symptoms of heart disease according to the latest medical guidelines would generate a more targeted response, he says.

Instructing ChatGPT to provide an answer from a particular perspective can also guide it toward providing specific, tailored responses. “You can say that you are a pathologist ‘who is an expert in soft-tissue tumors and molecular pathology so walk me through the differential diagnosis between liposarcoma and synovial sarcoma,’” Dr. Glassy says.

Plug-ins. Plug-ins, the specialized programs accessible over the Internet that work in conjunction with ChatGPT, add functionality that increases the model’s effectiveness, Dr. Glassy says. The Wolfram plug-in, from Wolfram Research, for example, performs complex mathematical computations and can leverage Wolfram’s subject-specific databases related to science, technology, and other fields to help ChatGPT arrive at more in-depth answers. Other plug-ins, such as Show Me Diagrams, can make it easier to create a wide variety of diagrams for visualizing complex information through ChatGPT.

A subscription to ChatGPT Plus provides access to hundreds of plug-ins via OpenAI’s plug-in store.

GPTs and other customized chatbots. GPTs, a feature of the ChatGPT Plus service, are sets of customized instructions created by ChatGPT users that function as small applications for performing specific tasks. GPTs were developed because users were maintaining long sets of carefully crafted prompts that they would manually input into ChatGPT every time they used it, according to OpenAI.

Dr. Glassy tested the process of creating a GPT by uploading a CAP synoptic report to ChatGPT and instructing the model to use the report as a template. The process took about 15 minutes and allowed him to put information into standard synoptic report format quickly and easily.

Users can publish their GPTs on OpenAI’s site or they can choose to keep them private. OpenAI launched its GPT Store, “which is similar to the app store for Apple and Google devices,” last month, says Dr. Glassy. “There are now hundreds of free GPTs [available to ChatGPT Plus subscribers], some of which are applicable to medicine and pathology.”

Pathologists can also access multiple proprietary and open-source technologies via the Web to build chatbots that exclusively access information from their institutions rather than a plethora of information from the Internet. Pathologists with a curated collection of 100 hematology papers, for example, could create a chatbot to answer questions based only on information from that collection of papers, Dr. Glassy says.

Other applications. Dr. Glassy, who serves as a trustee of the American Board of Pathology, has used ChatGPT to create better distractors, or wrong answers, for the ABPath CertLink questions he writes. “It’s also good at catching grammatical mistakes in the CertLink critiques,” he says.

Some pathologists too have leveraged ChatGPT’s sophisticated translation capabilities to translate lab reports for non-English-speaking patients, he notes.

Recently, newer methods for training algorithms have shown promise in reducing the number of hallucinations, Dr.

Glassy says. For example, reinforced learning from human feedback is a technique that allows users to rate the answers they receive from an LLM to optimize that model. And retrieval-augmented generation prompts LLMs to check information against authoritative sources outside the training set before issuing an answer to a query.

Despite continual improvements, Dr. Glassy expects that hallucinations will never be eliminated completely. You can't assume an answer is right, he says. "It may sound correct, but it could still be dead wrong. These different mitigating techniques can limit errors, but you still need to confirm."

—Renee Caruthers

XiFin offers tool that focuses on payer rate transparency

XiFin has released its Payor Rate Transparency Monitor, an interactive visualization tool that compares contracted rates published by UnitedHealthcare, Aetna, and Cigna by mapping aggregate billing code reimbursement data.

The monitor helps pathology laboratories and others to compare contracted rates for most common laboratory services and reimbursement rates to better negotiate with payers. It works in part by drilling down into billing codes and modifiers developed by the Centers for Medicare and Medicaid Services.

Users of the monitor can compare UnitedHealthcare, Aetna, and Cigna across common billing codes and related billing modifiers and spotlight detailed visualizations through bubble charts and comprehensive analyses that showcase the lowest, highest, and weighted average rates for specific services of each payer. The monitor initially focuses on 23 common codes, such as lipid panel and urinalysis, but each month XiFin will highlight a new set of codes.

The monitor allows XiFin's clients "to make informed, timely decisions that maximize reimbursement," said the company's chief operating officer, Kyle Fetter, in a press statement. "But beyond the benefit to individual laboratories, these data shed light on the reimbursement landscape as a whole and have the potential to clarify pricing data, such as industry averages for services, to inform policies. Among other things, the monitor highlights absurdities, such as a basic metabolic panel with contracted rates that vary from less than a penny to several hundred dollars. Additionally, it underscores the unsustainably low average rates offered by some of the largest payers in the country for certain services."

[XiFin](#), 866-934-6364

CAP online course addresses cyberattack awareness

The CAP is offering an online cyberattack preparedness activity to members and nonmembers that focuses on taking preventive measures to protect laboratory information technology systems, equipment, and data.

The activity, "2023 ICBE-D: Cyberattack Awareness: Tips for Planning and Preparedness," uses details of an actual cyberattack on a laboratory to share best practices for pathologists in private and academic health care institutions.

The objectives of the educational offering are to identify the impact of prolonged system outages on pathology and laboratory medicine services and share guidance on adapting processes so labs can function during such outages, as well as to assess laboratories' readiness for unplanned, prolonged system outages.

The authors of the activity are William O. Humphrey, MD, CAP Informatics Committee member and neuropathology fellow, Mayo Clinic, Rochester, Minn., and Tara Corona, MS, instructional designer, College of American Pathologists.

To enroll or for more information, go to <http://tinyurl.com/bfvbpc6p>.

TriMetis and iSpecimen undertake partnership

iSpecimen, an online global marketplace that connects scientists who require biospecimens for medical research with a network of health care specimen providers, has entered a strategic partnership with TriMetis Life Sciences, a provider of digital pathology, laboratory, and artificial intelligence workflow and automation solutions.

Under the pilot program agreement, iSpecimen, its suppliers, and clients can use technology solutions from TriMetis.

The partnership will initially focus on a subset of solid tumor types and allow iSpecimen users to employ TriMetis' computer-assisted pathology quality control AI to help standardize and enhance tissue sample evaluation. It will also offer iSpecimen users access to TriMetis' ARCH ecosystem. This includes ARCH marketplace, for buying and selling biospecimens and images, and ARCH LabFlow, for automating digital image workflow and processes.

[TriMetis Life Sciences](#), 901-410-1441

Dr. Aller practices clinical informatics in Southern California. He can be reached at rayaller@gmail.com. Dennis Winsten is founder of Dennis Winsten & Associates, Healthcare Systems Consultants. He can be reached at dennis.winsten@gmail.com.