EDITORIAL

AI (Artificial Intelligence) as an IA (Intelligent Assistant): ChatGPT for Surveillance Colonoscopy Questions

This study by Mukherjee et al[1] evaluated the performance of Chat Generative Pre-Trained Transformer (ChatGPT), a large language model, on answers for common queries about the intervals of surveillance colonoscopy.

The study has proposed 12 questions on colonoscopy. The questions ranged from simple to complex content based on colonoscopy findings which consider size, number, pathology, and completeness of polyp resection. The answers were assessed by 4 gastroenterology fellows. They aimed to determine (1) the accuracy of answers by ChatGPT regarding intervals of colonoscopy as referenced to the standard guidelines and (2) the patient education usage.

The first 3 questions (Q1 to Q3) are categorized as simple and they achieved the highest answer score – "addresses the query and is factually entirely correct." On the contrary, Q4 to Q12 have more complicated wording on the interval of surveillance colonoscopy after the positive findings. The percentage of entirely accurate answers dropped to 0%–25% in such questions, which detailed polyp findings in the index colonoscopy. The same trend is also observed with the use of ChatGTP for patient self-education with 75% to 100% in the first 3 simple questions and 0% in the rest.

This is the first study focused on queries regarding interval of surveillance colonoscopy based on various findings. It differs from previous studies on colonoscopy-related queries aimed at general patient education such as colonoscopy preparation in preprocedure, intraprocedure, and postprocedure phases and related complications.[2,3] Thus, the current study by Mukherjee et al explores an unmet need for many patients who are eager to understand their colonoscopy findings and how the results may influence their future colonoscopy intervals. Whether the gastroenterology community is ready or not, patients will increasingly utilize large language models such as ChatGPT for health-related questions, and so it is important that we become aware of the information they are being provided.

The question at hand is what and how reliable the data sources of ChatGPT are. According to this study, trustworthy data sources, such as professional guidelines, provide different recommendations compared to ChatGPT. For example, ChatGPT answers the incorrect age, 50 vs 45 years, of starting screening colonoscopy and interval of surveillance colonoscopy. Interestingly, regeneration of responses in the same day yielded different results.[4] The time period of data training potentially plays a role in these findings.

Clinicians need to be cautious, if implementing ChatGPT use in clinical practice, given the rapidly expanding medical literature and potential legal ramifications of leveraging outdated sources for patient education.

The study did not specify the version of ChatGPT used. ChatGPT 4.0 as a paid version has been shown to yield a better outcome than version 3.5 with more updated training database. Tariq et al.[3] reported that 3 large language models were compared for 47 general questions on colonoscopy. ChatGPT 4.0 yielded the most completely correct answers with 91.4% vs 14.9% and 6.4% by Bard and ChatGPT 3.5, respectively. Thus, a subsequent study would be interesting to investigate the comparative efficacy among ChatGPT 3.5, 4.0, and Bard on these specific queries on intervals of surveillance colonoscopy.

The patient usability was addressed by the study with good rating by physicians on the general questions, similarly to prior studies which compared the questions based on the hospital webpage for patient education.[2] On the contrary, questions on interval of surveillance colonoscopy requiring polyp information yielded very low satisfaction. Further studies including patient raters are warranted to verify the findings.

In conclusion, this study emphasizes how "AI may serve as an IA," the artificial intelligence goal to become an intelligent assistant. ChatGPT is becoming an IA for providers and patients alike, which can be used as an adjunctive tool to help reduce the work burden for physicians. However, the quality control of the assistant is an area for improvement. Regulation of generative AI is required to monitor the concerns for safety, usage, data update, and accuracy for training.

PASSISD LAOVEERAVAT
Division of Digestive Disease and Nutrition
Department of Internal Medicine
University of Kentucky
Lexington, Kentucky

DOUGLAS A. SIMONETTO
Division of Gastroenterology and Hepatology
Mayo Clinic
Rochester, Minnesota

References

1. Mukherjee S, Durkin C, PeBenito AM, et al. Assessing ChatGPT's ability to reply to queries regarding colon cancer screening based on multi-society guidelines. Gastro Hep Adv 2023;2(8):1040–1043.
2. Lee TC, Staller K, Botoman V, et al. ChatGPT answers common patient questions about colonoscopy. Gastroenterology 2023;165(2):509–511.e7.
3. Tariq R, Malik S, Khanna S. Evolving landscape of large language models: an evaluation of ChatGPT and bard in answering patient queries on colonoscopy. Gastroenterology 2023.

4. Yeo YH, Samaan JS, Ng WH, et al. Assessing the performance of ChatGPT in answering questions regarding cirrhosis and hepatocellular carcinoma. Clin Mol Hepatol 2023;29(3):721–732.

Correspondence:
Address correspondence to: Passisd Laoveeravat, MD, Division of Digestive Disease and Nutrition, Department of Internal Medicine, University of Kentucky, 800 Rose St, Lexington, Kentucky 40536. e-mail: passisd.laoveeravat@uky.edu.