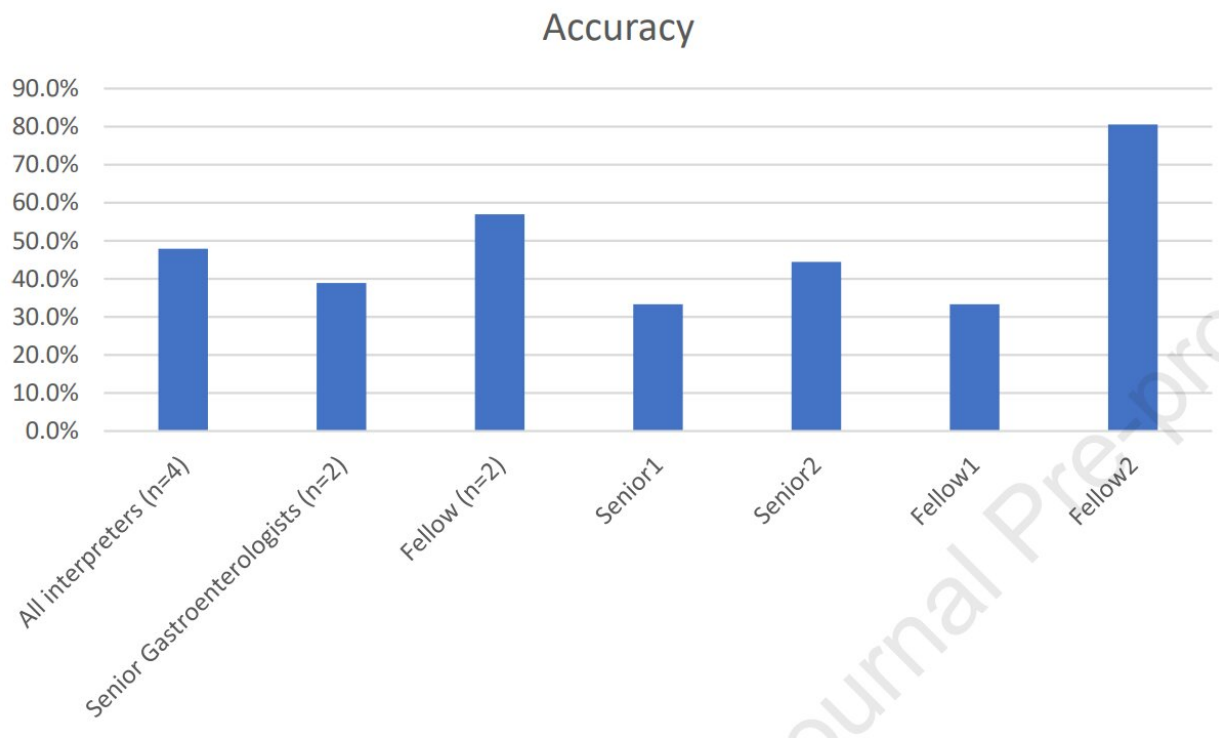# Q&A: ChatGPT answers common patient questions about colonoscopy

May 17 2023

Accuracy of detecting answers generated by AI by two senior and two fellow gastroenterologists. Credit: *Gastroenterology* (2023). DOI: 10.1053/j.gastro.2023.04.033

Braden Kuo, MD, a neurogastroenterologist and the director of the Center for Neurointestinal Health at MGH and an associate professor of Medicine at Harvard Medical School and Tsung-Chun Lee, MD, Ph.D.,

of Taipei Medical University Shuang Ho Hospital, in Taiwan are co-authors of a recent research letter published in *Gastroenterology*, "ChatGPT Answers Common Patient Questions About Colonoscopy."

Here, they answer questions about their latest study.

What was the question you set out to answer with this study?

ChatGPT, a new language processing tool driven by artificial intelligence (AI), provides conversational text responses to questions and can generate valuable information for enquiring individuals, but the quality of ChatGPT-generated answers to medical questions is currently unclear.

What methods or approach did you use?

We retrieved eight common questions and answers about colonoscopy from the publicly available webpages of three randomly-selected hospitals from the top-20 list of the US News & World Report Best Hospitals for Gastroenterology and Gastrointestinal Surgery.

We inputted these questions as prompts for ChatGPT for two times on the same day and recorded the ChatGPT-generated answers.

We then used a plagiarism detection software to compare the text similarity among all answers. Finally, to objectively interpret the quality of ChatGPT-generated answers, four gastroenterologists rated 36 random pairs of questions and answers for the following quality indicators on a 7-point scale:

- (1) ease of understanding
- (2) scientific adequacy

- (3) satisfaction with the answer

Raters were also asked to interpret whether the answers were AI-generated or not.

What did you find?

ChatGPT answers had extremely low text similarity compared with answers on hospital webpages, while the text similarity ranged from 28% to 77% between the two ChatGPT answers.

ChatGPT answers were rated similarly by gastroenterologists to non-AI answers in terms of ease of understanding, but with the average AI scores higher than non-AI scores. Scores were also similar related to scientific adequacy and satisfaction with the answers. The raters were only 48% accurate in telling which answers were provided by ChatGPT.

This study is the first of its kind to demonstrate that a contemporary large language model–derived conversational AI program is able to provide easy to understand, scientifically adequate, and generally satisfactory answers to common questions about colonoscopy, as determined by gastroenterologists.

Such programs may help to optimize clinical communication to patients, especially for high volume procedures like colonoscopy. Conversational AI empowered by large language models like ChatGPT has the potential to transform and benefit shared decision-making by patients and physicians.

What are the implications?

Future research should explore responses to a broader sample of patient questions and clinical conditions and include both patients and

physicians as raters.

More information: Tsung-Chun Lee et al, ChatGPT Answers Common Patient Questions About Colonoscopy, *Gastroenterology* (2023). DOI: 10.1053/j.gastro.2023.04.033

Provided by Massachusetts General Hospital