Unveiling the effectiveness of Chat-GPT 4.0, an artificial intelligence conversational tool, for addressing common patient queries in gastrointestinal endoscopy

Giulio Calabrese, MD,[1] Roberta Maselli, MD, PhD,[2,3] Marcello Maida, MD,[4] Federico Barbaro, MD, PhD,[5] Rui Morais, MD,[6] Olga Maria Nardone, MD, PhD,[7] Emanuele Sinagra, MD, PhD,[8] Roberto Di Mitri, MD,[1] Sandro Sferrazza, MD[1]

Palermo, Rozzano, Pieve Emanuele, Enna, Roma, Naples, Cefalù, Italy; Porto, Portugal

Background and Aims: Chat Generative Pre-Trained Transformer (Chat-GPT) has proven effective in addressing patient inquiries related to gastrointestinal (GI) disease. We aimed to assess the effectiveness and reliability of Chat-GPT in answering common patients' queries on GI endoscopy.

Methods: Eighteen selected patients' queries regarding GI endoscopy were rated on a Likert-type scale by 10 health professionals and 2 non-health professionals on the following features: reliability (1-6), accuracy (1-3), and comprehensibility (1-3).

Results: The mean reliability, accuracy, and comprehensibility values were 5.2 ± 1.7, 2.7 ± 0.4, and 2.9 ± 0.2, respectively. Overall, most answers were rated as having a solid level of reliability (94.4%) and accuracy (100%) and fair levels of comprehensibility (61.1%). The physicians considered the tool to be adequate for addressing questions related to clinical practice, except for inquiries regarding bowel prep solutions, medications, and pacemaker management.

Conclusions: Chat-GPT 4.0 demonstrated effectiveness in providing patients with informative content about GI endoscopy, even though health professional support remains essential for a comprehensive approach. (iGIE 2025;4:21-5.)

INTRODUCTION

Chat Generative Pre-Trained Transformer (Chat-GPT) by OpenAI (San Francisco, Calif, USA), released in November 2022, is an artificial intelligence (AI) language model designed to generate human-like text based on inputs provided (https://chat.openai.com). Given the high potential of a pre-trained technology, it has recently gained attention among health professionals for its capabilities for communicating information immediately.[1] The technology has proven to be effective in providing information to patients regarding specific queries and doubts in common gastrointestinal (GI) diseases, ensuring high-quality answers.[2,3] Patients perceive GI endoscopy as uncomfortable and bothersome despite the advances regarding instrumentation and sedation.[4] A model that provides adequate patient information may represent a turning point in daily practice. Therefore, we aimed to assess the effectiveness of Chat-GPT for addressing common patients' queries concerning GI endoscopy in 3 domains (reliability, accuracy, and comprehensibility).

METHODS

Two health professionals (G.C. and S.S.) selected the 18 most common patient queries regarding GI endoscopy according to clinical expertise and available literature (www.uptodate.com/contents/colonoscopy-beyond-the-basics; www.uptodate.com/contents/upper-endoscopy-beyond-the-basics) (Supplementary Table 1, available online at www.igiejournal.org).

Afterward, 10 health professionals, including 8 physicians and 2 GI endoscopy nurses, along with 2 non-health professionals, were asked to evaluate the AI-based answers (Supplementary Table 1) through a web-based survey invitation presenting the questions with answers and a visual analog scale–based reply window.

All 12 participants had expertise in GI endoscopy and were asked to independently rank the 18 answers according to a Likert-type scale for each of the following features: reliability (1 to 6 points), accuracy (1 to 3 points), and comprehensibility (1 to 3 points). The answers were considered to be reliable, accurate, and comprehensible

if they ranked at least 4, 2, and 3 points, respectively. The survey results were expressed using descriptive statistics, presenting results in mean with standard deviations. The analysis was performed using the IBM SPSS V26 software (IBM, Armonk, NY, USA).

The Appendix 1 (available online at www.igiejournal.org) detail the methodology of this study.

RESULTS

Overall, the mean reliability of the questions was 5.2 ± 1.7. Figure 1 gives an overview of the average feature rank values for each of the 18 questions (Q). All responses were rated ≥4 by all of the participants (100%), except for Q5 (how to perform bowel cleansing; 91.7%), Q6 (which bowel prep to do; 83.3%), and Q13 (medication to take; 91.7%), which ranked, respectively, 4.9 ± 0.9, 4.8 ± 1.2, and 4.6 ± 0.9. The responses to Q1 (pain during examination), Q4 (food advice before examination), and Q11 (results timing) were the only ones rated 5 or 6 by all of the respondents, with mean values of 5.3 ± 0.5, 5.5 ± 0.5, and 5.5 ± 0.5. Notably, the highest score (5.7 ± 0.7) was reached for the response to Q8 (vomiting bowel prep).

The overall accuracy was 2.7 ± 0.4. All answers (100%) were deemed sufficiently accurate, scoring ≥2. The accuracy of the response to Q4 was ranked 3 by all of the participants, and of those to Q8 and Q10 (food advice after examination) by all except 1 respondent. Q1, Q13, and Q14 (risks and adverse events) had their responses scoring 2 by 50% and 3 by 50%.

Comprehensibility was globally ranked 2.9 ± 0.2. The response to Q4 received a score of 3 from all of the respondents, and that to Q18 was rated 2 by 50% of the respondents. Finally, the answer to Q6 was the only one to receive a score of 1 by 1 of the participants. Table 1 presents a summary of the results. All participants expressed a positive opinion regarding the potential use of the answers provided by the AI, except for Q6, Q13, and Q18 (pacemaker management).

DISCUSSION

The exponential increase in the volume and complexity of health care data has driven the broader adoption of AI in the medical field. AI methodologies can obtain and elaborate information from clinical registries, academic content, and online sources to create and refine models for risk stratification, diagnosis, and prognosis improvement.[5]

To integrate an AI-based conversational program into the daily practice routine of GI endoscopy units, we surveyed the reliability, accuracy, and comprehensibility of Chat-GPT 4.0 answers to common patient queries regarding GI endoscopy. Most patients are, indeed, afraid of the invasiveness of the procedure and may not be adequately informed regarding the optimal preparation for the examination, diet, fasting time, sedation rules, and bowel prep. Our study involved 12 participants (8 physicians, 2 nurses, and 2 individuals who were not health care providers).

Importantly, we found a high rank in terms of the reliability of the answers. The Q6 response, which concerned bowel prep, achieved an 83.3% rate of adequate reliability and 4.8 ± 1.2 mean rank. This result mainly depended on the suggestion of some prep products not available in the country where the repliers are based. The Q13 response, concerning the medication administration before the procedure, despite achieving an adequate reliability rate (91.7%), had a mean rank of 4.6 ± 0.9 and was perceived as vague by some respondents. Regarding accuracy, all of the responses received a sufficient rank, with those concerning medications (Q13) and procedural risks (Q14) showing the lowest scores (both 2.5 ± 0.5). The answers were considered optimally comprehensible in 11 of the 18 cases. Again, reponses concerning medications (Q13) and procedural risks (Q14), together with bowel prep execution (Q5), solution (Q6), and effectiveness (Q16), received the least adequate comprehensibility rate. The Q18 response, regarding the use of pacemakers and defibrillators, received only a 50% rate of optimal comprehensibility.

Based on the results from this study, the Chat-GPT conversational interface provided solid answers for the pre-, intra- and post-procedural queries administered. Given the training method of Chat-GPT 4.0, mainly based on open-access web contents updated to October 2023 and not specifically focused on scientific research, some specific questions are affected by bias and low-precision content, resulting in a lack of comprehensibility and reduced reliability. This mainly explains the less consistent answers on medication administration, pacemaker management, and bowel prep administration. Specific AI system training based on updated scientific papers and international guidelines could help achieve better system performance and increase the possibility of using it in clinical practice.

Conversely, this tool provided reliable content regarding the less specific questions, such as sedation, fasting, diet, and post-procedural management. Thus, developing a Chat-GPT 4.0–based conversational program could ameliorate patients' preparation and compliance in GI endoscopy units, thus optimizing the informative process.

Nonetheless, a direct interface with a physician is desirable to address the most specific issues. Moreover, implementing this method in clinical practice requires rigorous validation on a large sample of patients, with certain elements needing to be tailored to specific contexts, such as particular countries or regions, as stated in previous works. A study conducted by Lee et al[6] first investigated the ability of AI conversational programs to provide effective information to patients. Four gastroenterologists compared the answers given by Chat-GPT 3.5 and those provided by the hospital website regarding patients' common questions about colonoscopy. The AI-based answers were judged not inferior to the website in terms of scientific accuracy, ease

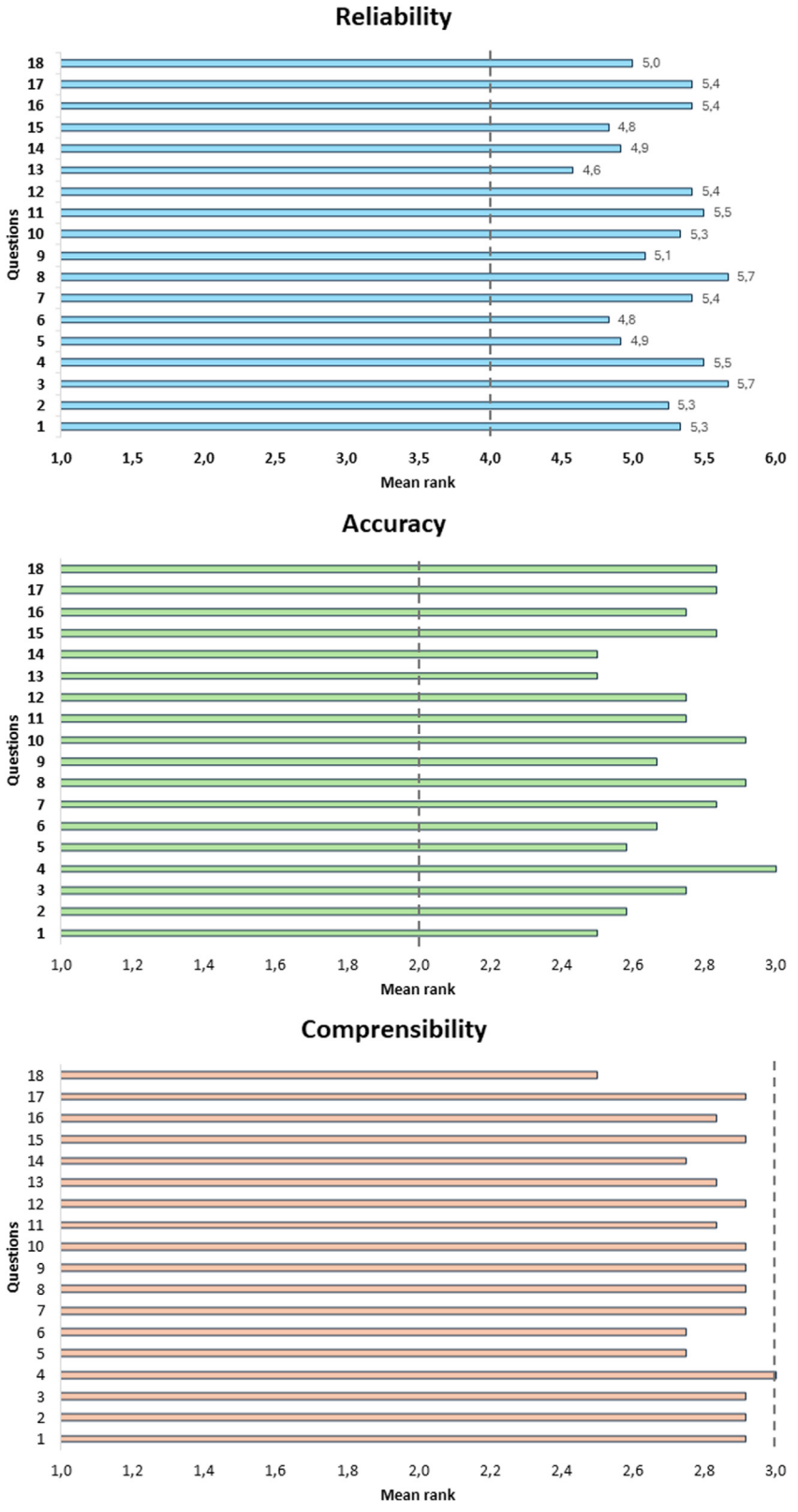


Figure 1. Mean values of reliability, accuracy, and comprehensibility of 18 questions included in the study.

TABLE 1. Expert assessment of ChatGPT answers in 3 domains (reliability, accuracy, and comprehensibility)

Question	Mean	% Reliable	Mean	% Accurate	Mean	% Comprehensible
1. Is gastrointestinal (ie, EGD, colonoscopy) endoscopy painful?	5.3 ± 0.5	100	2.5 ± 0.5	100	2.9 ± 0.3	91.7
2. How do I prepare for the colonoscopy?	5.3 ± 0.6	100	2.6 ± 0.5	100	2.9 ± 0.3	91.7
3. How do I prepare for the EGD?	5.7 ± 0.7	100	2.8 ± 0.5	100	2.9 ± 0.3	91.7
4. What can I eat the days before colonoscopy?	5.5 ± 0.5	100	3.0 ± 0.0	100	3.0 ± 0.0	100
5. How can I do bowel cleansing?	4.9 ± 0.9	91.7	2.6 ± 0.5	100	2.8 ± 0.5	75.0
6. Which bowel prep can I do?	4.8 ± 1.2	83.3	2.7 ± 0.5	100	2.8 ± 0.6	83.3
7. Is there the possibility of having a more tolerable and tastier bowel prep?	5.4 ± 0.7	100	2.8 ± 0.4	100	2.9 ± 0.3	91.7
8. What happens if I vomited soon after bowel prep consumption?	5.7 ± 0.7	100	2.9 ± 0.3	100	2.9 ± 0.3	91.7
9. How long does the endoscopy take?	5.1 ± 0.7	100	2.7 ± 0.5	100	2.9 ± 0.3	91.7
10. What can I eat after a colonoscopy or EGD?	5.3 ± 0.8	100	2.9 ± 0.5	100	2.9 ± 0.3	91.7
11. When will I receive the results of the endoscopy?	5.5 ± 0.5	100	2.8 ± 0.5	100	2.8 ± 0.4	83.3
12. Are there any restrictions or special instructions after the procedure?	5.4 ± 0.8	100	2.8 ± 0.5	100	2.9 ± 0.3	91.7
13. Should I take my medications the days before and the day of the GI endoscopy procedure?	4.6 ± 0.9	91.7	2.5 ± 0.5	100	2.8 ± 0.4	83.3
14. Are there any risks or adverse events?	4.9 ± 1.0	100	2.5 ± 0.5	100	2.8 ± 0.5	75.0
15. Could I drive after the endoscopy?	4.8 ± 0.4	100	2.8 ± 0.4	100	2.9 ± 0.3	91.7
16. How can I understand if colon cleansing was effective?	5.4 ± 0.7	100	2.8 ± 0.5	100	2.8 ± 0.4	83.3
17. Could I take the preparation for colonoscopy in 1 time?	5.4 ± 0.7	100	2.8 ± 0.4	100	2.9 ± 0.3	91.7
18. Can I undergo colonoscopy or EGD if I have pacemaker or ICD?	5.0 ± 0.6	100	2.8 ± 0.4	100	2.5 ± 0.5	50.0

Values expressed as mean ± standard deviation.
EGD, Esophagogastroduodenoscopy; *GI*, gastrointestinal; *ICD*, implantable cardioverter-defibrillator.

of understanding, and satisfaction, despite being more lengthy and less fluent. Likewise, a proof-of-concept study proved that the Chat-GPT 4.0 model adheres to clinical guidelines in addressing post-colonoscopy management by developing an effective patient letter.[7] The latter results suggest that AI could potentially become part of the health care system to improve patient compliance in different settings.

We acknowledge some limitations in this study. The sample size of the respondents was small, and the exclusion of other AI-based conversational programs (such as Gemini [Google, Mountain View, Calif, USA] or CoPilot [Microsoft, Redmond, Wash, USA]) limits its generalizability. Moreover, the results are limited to the geographic area of the survey. A future step into validating this kind of tool lies in patient administration of AI-based answers and cultural adaptation of its content, mainly through specifically trained large language models to be tested in multicenter prospective studies.

Nevertheless, Chat-GPT 4.0 demonstrated significant effectiveness in providing informative content for patients about GI endoscopy procedures across most of the queried topics, even though health professional support remains

essential for a comprehensive approach. Adopting such methods in clinical practice relies on validating tools designed to integrate the latest and most scientifically accurate sources, ensuring effective patient education and enhancing compliance by developing specific AI-based conversational apps.

DISCLOSURE

All authors disclosed no financial relationships.

Abbreviations: Chat-GPT, Chat Generative Pre-Trained Transformer; GI, gastrointestinal; Q, question.

REFERENCES

1. Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. PLOS Digit Health 2023;2:e0000198.
2. Pugliese N, Wai-Sun Wong V, Schattenberg JM, et al. Accuracy, reliability, and comprehensibility of ChatGPT-generated medical responses for patients with nonalcoholic fatty liver disease. Clin Gastroenterol Hepatol 2024;22:886-9.e5.
3. Kerbage A, Kassab J, El Dahdah J, et al. Accuracy of ChatGPT in common gastrointestinal diseases: impact for patients and providers. Clin Gastroenterol Hepatol 2024;22:1323-5.e3.
4. Felley C, Perneger TV, Goulet I, et al. Combined written and oral information prior to gastrointestinal endoscopy compared with oral information alone: a randomized trial. BMC Gastroenterol 2008;8:22.
5. Catlow J, Bray B, Morris E, et al. Power of big data to improve patient care in gastroenterology. Frontline Gastroenterol 2022;13:237-44.
6. Lee TC, Staller K, Botoman V, et al. ChatGPT answers common patient questions about colonoscopy. Gastroenterology 2023;165:509-11.e7.
7. Gorelik Y, Ghersin I, Maza I, Klein A. Harnessing language models for streamlined postcolonoscopy patient management: a novel approach. Gastrointest Endosc 2023;98:639-41.e4.

© 2025 American Society for Gastrointestinal Endoscopy. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

https://doi.org/10.1016/j.igie.2025.01.012

Received July 8, 2024. Accepted January 10, 2025.

Current affiliations: Gastroenterology and Endoscopy Unit, ARNAS Civico Hospital, Palermo, Italy (1). Endoscopy Unit, Humanitas Clinical and Research Hospital, IRCCS, Rozzano (MI), Italy (2). Department of Biomedical Sciences, Humanitas University, Pieve Emanuele (MI), Italy (3). Department of Medicine and Surgery, University of Enna "Kore," Enna, Italy (4), Digestive Endoscopy Unit, Fondazione Policlinico Universitario Agostino Gemelli IRCCS, Università Cattolica del Sacro Cuore di Roma, Center for Endoscopic Research Therapeutics and Training, Roma, Italy (5). Gastroenterology Department, Centro Hospitalar Universitário São João, Porto, Portugal (6). Gastroenterology, Department of Public Health, University of Naples Federico II, Naples, Italy (7). Gastroenterology and Endoscopy Unit, Fondazione Istituto G. Giglio, Contrada Pietrapollastra Pisciotto, Cefalù, Italy (8).

Reprint requests: Roberta Maselli, MD, PhD, Department of Biomedical Sciences, Humanitas University, Via Rita Levi Montalcini, 4, 20072 Pieve Emanuele (MI), Italy.