

ChatGPT Answers Common Patient Questions About Colonoscopy



See editorial on page 336.

ChatGPT (OpenAI) is a 175 billion-parameter large language model (LLM) artificial intelligence (AI) that was released in November 2022. ChatGPT is developed based on the generative pretrained transformer (GPT) 3.5 natural language processing technology and provides a conversational text response to a given prompt.¹

One potential application of ChatGPT is answering patients' medical questions. With more than 70 million procedures annually in the United States,² screening colonoscopies are frequently the subject of questions in gastroenterology. In this study, we examine the quality of ChatGPT-generated answers to common questions (CQs) about colonoscopy.

We retrieved 8 CQs and answers about colonoscopy from the publicly available webpages of 3 randomly selected hospitals from the top-20 list of the *US News & World Report* Best Hospitals for Gastroenterology and GI Surgery (Supplementary Methods). We input these questions as prompts for ChatGPT (January 30, 2023, version) twice on the same day and recorded the ChatGPT-generated answers as AI1 and AI2, respectively. We compared the text similarity among all answers using a plagiarism detection software (Supplementary Table 1). To objectively interpret the quality of ChatGPT-generated answers, 4 gastroenterologists (2 senior gastroenterologists and 2 fellows) rated 36 pairs of CQs and answers, randomly displayed, for the following quality indicators on a 7-point Likert scale: (1) ease of understanding, (2) scientific adequacy, and (3) satisfaction with the answer (Table 1). Raters were also requested to interpret whether the answers were AI generated or not.

We found that ChatGPT answers had extremely low text similarity (0%–16%) compared to answers on hospital webpages, and the text similarity ranged from 28% to 77% between the 2 ChatGPT answers, except for the CQ7 (Supplementary Table 1).

Gastroenterologists rated ChatGPT answers similarly to non-AI answers in ease of understanding (AI, 5.0–6.4 vs non-AI, 4.8–5.8; all $P > .00089$ comparing means after Bonferroni adjustment for 56 multiple comparisons), with the AI mean scores higher than non-AI scores. Scientific adequacy scores were also similar (AI, 5.4–6.5 vs non-AI, 5.1–6.3, nonsignificant), with the AI mean score higher than non-AI 63% of the time. AI and non-AI answers received similar ratings regarding satisfaction with the answers (AI, 4.9–6.3 vs non-AI, 4.8–5.8; nonsignificant) (Table 1).

The raters demonstrated only 48% accuracy in identifying AI-generated answers, with 41% sensitivity and 54% specificity. Three raters had an accuracy of less than 50%,

and 1 (a fellow) had 81% accuracy (Supplementary Figure 1 and Supplementary Table 2).

This study is the first of its kind, to our knowledge, to demonstrate that a contemporary LLM-derived conversational AI program is able to provide easy-to-understand, scientifically adequate, and generally satisfactory answers to CQs about colonoscopy as determined by gastroenterologists.

One surprising finding was the low sensitivity in identifying AI-generated answers (sensitivity of 6%, 25%, and 44%, respectively). Heuristic feedbacks from the outperforming fellow revealed that “ChatGPT answers tended to be lengthy, used many colons (‘:’) in the long list of possibilities it gave, and tended to be more of a list rather than a narrative paragraph in response.” Contrastingly, answers from hospital webpages were “more like verbal responses to a patient as opposed to something more encyclopedic.”

This study suggests a potential role of conversational AI programs in optimizing the communication between patients and health care providers, especially for high-volume procedures like colonoscopy. Despite similar ratings, there was little overlap or plagiarism between the AI and non-AI answers as well as between the 2 AI answers (Supplementary Table 1 and Supplementary Table 3), which suggested the inherent plagiarism-avoiding design in LLMs and the capabilities of LLMs to create unique answers to the same question.

Accumulated publications about ChatGPT in PubMed grew 10-fold from 20 on February 3, 2023, to 246 on April 14, 2023 (Supplementary Figure 2), with topics including board examinations,³ authorship, editorial policies,⁴ medical education,⁵ clinical decision support,⁶ a LLM assessment framework,⁷ etc.

Although early in the adoption curve,⁸ LLMs (ChatGPT, BioGPT, BARD, and others) may represent a transformative innovation in how medical information (MI) is created by physicians and consumed by patients. Especially in the current era of shared decision making and the consumerization of health care, patients have been actively consuming MI through multiple channels and accessing providers through electronic patient portals at an exponential magnitude, which has the potential to benefit patients but, simultaneously, represents a heavy burden for providers

Abbreviations used in this paper: AI, artificial intelligence; CQs, common questions; GPT, generative pretrained transformer; LLM, large language model; MI, medical information.

Most current article

© 2023 The Author(s). Published by Elsevier Inc. on behalf of the AGA Institute. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

0016-5085

<https://doi.org/10.1053/j.gastro.2023.04.033>

Table 1. Quality Indicators (Ease of Understanding, Scientific Adequacy, Satisfaction) for Answers From AI and From Non-AI Sources

Common questions (CQ)	Source of answers	"The answers are easy to understand."		"The answers are scientifically adequate."		"I am satisfied with the answers."	
		Mean	P	Mean	P	Mean	P
CQ1 What is a colonoscopy?	AI	6.4	.54	5.9	.94	5.6	.78
	Non-AI	5.7		5.7		5.7	
CQ2 Why is a colonoscopy performed?	AI	5.9	.20	5.8	.65	5.6	.39
	Non-AI	4.8		5.5		4.9	
CQ3 How to prepare for a colonoscopy?	AI	5.9	.59	6.1	.72	5.8	1
	Non-AI	5.8		5.9		5.6	
CQ4 What to expect during the colonoscopy procedure?	AI	5.9	.78	5.6	.60	5.3	.66
	Non-AI	5.6		6.0		5.5	
CQ5 What to expect after the colonoscopy procedure?	AI	6.3	.05	5.9	.16	6.1	.12
	Non-AI	5.3		5.1		5.0	
CQ6 What to do after a negative colonoscopy result?	AI	6.4	.44	5.9	.28	6.1	.77
	Non-AI	5.8		6.3		5.8	
CQ7 What to do after a positive colonoscopy result?	AI	5.0	.87	5.4	.11	4.9	.83
	Non-AI	4.6		6.0		4.8	
CQ8 What to expect about complications?	AI	6.1	.40	6.5	.01	6.3	.02
	Non-AI	5.6		5.4		4.8	

Interpreted by four physicians with 7-points Likert Scale. (7 = Strongly agree, 4 = Neutral, 1 = Strongly disagree)

Interpretation 1: "The answers are easy to understand."

Interpretation 2: "The answers are scientifically adequate."

Interpretation 3: "I am satisfied with the answers."

Statistical analysis by Mann Whitney U test. After adjustment for multiple comparison, new Bonferroni corrected alpha value = $0.05 / 56 = 0.00089$ *: $P < .00089$ as significant.

and staff. We envision that AI-generated MI, with appropriate provider oversight, accreditation, and periodic surveillance, could improve the efficiency of care and free providers for more cognitively intensive patient communications.

Nevertheless, potential pitfalls have to be addressed. Currently, ChatGPT-generated MI is not constructed on the basis of clinical evidence but is created through an LLM trained on diverse Internet texts with reinforcement learning by human feedback.¹ LLM outputs may be sensitive and vulnerable to prompt engineering, that is, manipulation by subtle changes in inputting prompts, and the consistency of performance might be in "a state of constant change."⁹ Thus, there remains a large gap, technology- and format-wise, regarding the use LLMs in responsible clinical care.¹⁰ Implicit bias is another concern, because the clinical utility might differ for patients with or without resources. Furthermore, readability analyses using validated reading-level metrics (Flesch-Kincaid Grade Level, Gunning Fog Index) revealed that the AI-generated answers were written with significantly higher grade reading levels than the hospital webpages ($P < .001$), far exceeding the eighth grade thresholds recommended (Supplementary Table 4).

This study has several limitations. First, we did not include patient raters, the group to which colonoscopy preparation answers will be ultimately provided. For this

study, we aimed to initially critique AI-generated MI through the lens of medical professionals. Future research should explore responses to a broader sample of questions and clinical conditions, as well as the inclusion of patient raters. Second, numbers of both the hospital webpages and raters were small, which limited broad generalizability. Finally, webpages of randomly selected top-tier hospitals may not be comprehensive.

This study shows that a conversational AI program can generate credible MI in response to common patient questions. With dedicated domain training, there is meaningful potential to optimize clinical communication to patients.

Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Gastroenterology* at www.gastrojournal.org, and at <https://doi.org/10.1053/j.gastro.2023.04.033>.

TSUNG-CHUN LEE

Division of Gastroenterology and Hepatology

Department of Internal Medicine

Taipei Medical University Shuang Ho Hospital

New Taipei City, Taiwan, and

Department of Internal Medicine, School of Medicine
College of Medicine, Taipei Medical University
Taipei, Taiwan, *and*
TMU Research Center for Digestive Medicine
Taipei Medical University
Taipei, Taiwan

KYLE STALLER

VLAICU BOTOMAN

MYTHILI P. PATHIPATI

SANSKRITI VARMA

BRADEN KUO

Center for Neurointestinal Health
Massachusetts General Hospital
Boston, Massachusetts, *and*
Division of Gastroenterology
Massachusetts General Hospital
Boston, Massachusetts, *and*
Department of Medicine
Harvard Medical School
Boston, Massachusetts

References

1. OpenAI. <https://openai.com/blog/chatgpt>. Accessed February 8, 2023.
2. Ladabaum U, et al. *Gastroenterology* 2019;157:137–148.
3. Gilson A, et al. *JMIR Med Educ* 2023;9:e45312.
4. Stokel-Walker C, et al. *Nature* 2023;614:214–216.
5. Mbakwe AB, et al. *PLOS Digit Health* 2023;2:e0000205.

6. Gaumgartner C. *Clin Transl Med* 2023;13:e1206.
7. Howard A, et al. *Lancet Infect Dis* 2023;23:405–406.
8. Rogers EM. *Diffusion of innovations*. 5th ed. New York: Simon and Schuster, 2003.
9. Lee P, et al. *N Engl J Med* 2023;388:1233–1239.
10. Sackett DL, et al. *Br J Med* 1996;312:71–72.

Received March 6, 2023. Accepted April 30, 2023.

Correspondence

Address correspondence to: Braden Kuo, MD, Harvard Medical School, GI Unit, Massachusetts General Hospital, GRJ-719, 55 Fruit Street, Boston, Massachusetts 02114. e-mail: bkuo@partners.org.

CRedit Authorship Contributions

Tsung-Chun Lee, MD (Conceptualization: Equal; Data curation: Equal; Formal analysis: Equal; Investigation: Equal; Methodology: Equal; Project administration: Equal; Writing – original draft: Equal; Writing – review & editing: Equal).

Kyle Staller, MD (Data curation: Equal; Formal analysis: Equal; Methodology: Equal; Writing – review & editing: Equal).

Vlaicu Botoman, MD (Data curation: Equal; Writing – review & editing: Equal).
Mythili P. Pathipati, MD (Data curation: Equal; Writing – review & editing: Equal).

Sanskriti Varma, MD (Data curation: Equal; Writing – review & editing: Equal).

Braden Kuo, MD (Conceptualization: Lead; Data curation: Lead; Formal analysis: Lead; Investigation: Lead; Methodology: Lead; Project administration: Lead; Resources: Lead; Supervision: Lead; Writing – original draft: Lead; Writing – review & editing: Lead).

Conflicts of interest

The authors disclose no conflicts.

Funding

Kyle Staller is supported by the National Institutes of Health (K23 DK120945).

Data Availability

Data and study materials can be made available to other researchers upon request.

Supplementary Methods

Data Collection and Objective Evaluation of Text Similarity

From the top-20 list of the *US News & World Report* Best Hospitals for Gastroenterology and GI Surgery,¹ we randomly selected 1 university-affiliated teaching medical center located on the East Coast, 1 in the Midwest, and 1 on the West Coast. We retrieved 8 common questions (CQs) and answers about colonoscopy from publicly available webpages of these 3 randomly selected hospitals. These 8 CQs about colonoscopy were frequently asked questions by patients. We used these questions as prompts for ChatGPT (January 30, 2023, version) and recorded its answers. To compare the consistency of artificial intelligence (AI) answers, we entered the same prompt 2 times on the same day and recorded the answers as AI1 and AI2, respectively. We compared the text similarity of answers between AI1 and AI2 as well as between AI1 and hospital webpages (non-AI) using plagiarism detection software,² and the results were shown in [Supplementary Table 1](#).

Objective Evaluation of Quality Indicators

To objectively interpret the quality of ChatGPT-generated answers, 4 gastroenterologists (2 senior gastroenterologists, 2 fellows) rated the answers. All 4 raters were blind to the sources of answers. For a total of 36 pairs of CQs and answers displayed in a random fashion, raters interpreted the following quality indicators on a 7-point Likert scale: (1) ease of understanding, (2) scientific adequacy, and (3) satisfaction with the answer ([Table 1](#)). Raters were also requested to interpret whether the answers were AI generated or not. Raters' performances in identifying AI-generated answers are shown in [Supplementary Figure 1](#) and [Supplementary Table 2](#). Example answers from AI and non-AI sources to CQ1, "What is a colonoscopy?" and to CQ2, "Why is a colonoscopy performed?" are exhibited in [Supplementary Table 3](#).

Objective Evaluation of Reading Level

Medical information provided to patients is recommended to be readable such that an eighth grader could understand it.³ We measured the reading levels of all answers to CQs by 2 objective indexes of reading level of texts: Flesch-Kincaid Grade Level⁴ and Gunning Fog Index⁵ ([Supplementary Table 4](#)). Both indexes are

well-recognized objective measures, in which index number x represents the corresponding x th grade reading level.³ Medical information given to patients ideally should have an index of 8 or less. Measurements were performed by an online readability tool (<https://readable.com>, accessed on April 12, 2023).

ChatGPT Publications in PubMed

We searched the PubMed database with the keyword of "ChatGPT" and obtained the list of publications that involved ChatGPT.⁶ The chronology of ChatGPT publications in PubMed is shown in [Supplementary Figure 2](#).

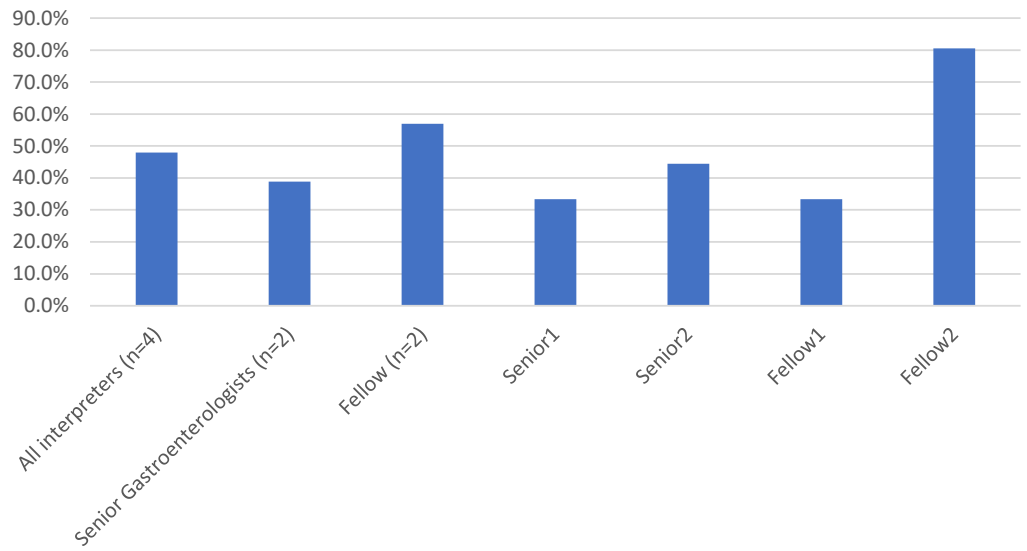
Statistical Analysis

Data are shown as mean or mean (standard deviation). Comparison of quality indicators on answers from AI vs non-AI sources was performed using the Mann-Whitney U test. After adjustment for multiple comparisons, the new Bonferroni-corrected α value was calculated as the following: 0.05 divided by 56 comparisons, that is, 0.00089. Therefore, $P < .00089$ was regarded as significant in the comparison of the 3 quality indicators among answers from AI and non-AI sources. Each rater's performance in detecting AI-generated answers was collectively calculated and expressed as the sensitivity, specificity, positive predictive value, negative predictive value, and accuracy. In reading level analyses, we compared the reading levels of the answers from the individual sources with an eighth grade reading level in 2-sided Student t tests and also compared the reading levels of answers from AI and non-AI sources using Mann-Whitney U tests. Statistical analyses were performed using SAS version 14 (SAS Institute).

Supplementary References

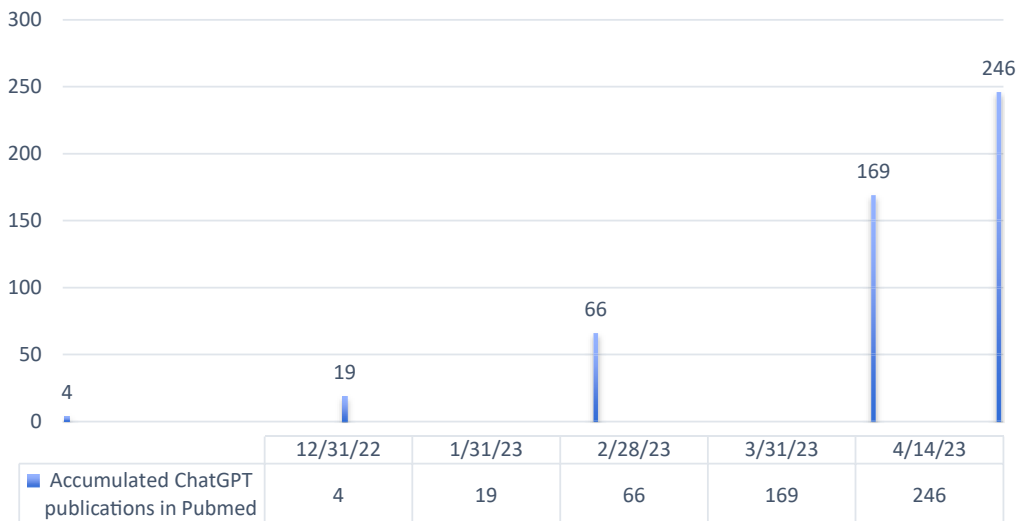
1. US News and World Report. <https://health.usnews.com/best-hospitals/rankings/gastroenterology-and-gi-surgery>. Accessed February 8, 2023.
2. Copyleaks. <https://copyleaks.com>. Accessed February 15, 2023.
3. Murphy B, et al. *Surgeon* 2022;20:E366–E370.
4. Kincaid JP, et al. <http://stars.library.ucf.edu/istlibrary/56>. Accessed April 12, 2023.
5. Avra TD, et al. *J Vasc Surg* 2022;76:1728–1732.
6. National Library of Medicine. <https://pubmed.ncbi.nlm.nih.gov>. Accessed April 12, 2023.

Accuracy



Supplementary Figure 1. Accuracy of detecting answers generated by AI by 2 senior and 2 fellow gastroenterologists.

Accumulated ChatGPT publications in Pubmed



Supplementary Figure 2. Numbers of accumulated publications related to ChatGPT in the PubMed database (accessed on April 14, 2023).

Supplementary Table 1. Comparison of Answers to 8 Common Questions About Colonoscopy

CQs	AI1 vs AI2		AI1 vs Hospital1		AI1 vs Hospital2		AI3 vs Hospital3	
	Words	Match, %	Words	Match, %	Words	Match, %	Words	Match, %
CQ1: What is a colonoscopy?	59 vs 136	28	59 vs 112	0	59 vs 89	0	59 vs 247	0
CQ2: Why is a colonoscopy performed?	133 vs 215	36	133 vs 129	0	133 vs 161	16	133 vs 167	0
CQ3: How to prepare for a colonoscopy?	223 vs 219	37	223 vs 168	0	223 vs 272	3	223 vs 299	0
CQ4: What to expect during the colonoscopy procedure?	201 vs 257	39	201 vs 132	0	201 vs 205	0	201 vs 344	0
CQ5: What to expect after the colonoscopy procedure?	178 vs 238	60	178 vs 41	0	178 vs 156	3	178 vs 202	0
CQ6: What to do after a negative colonoscopy result?	198 vs 78	44	198 vs NA	NA	198 vs 161	0	198 vs NA	NA
CQ7: What to do after a positive colonoscopy result?	189 vs 144	0	189 vs 72	0	189 vs 259	0	189 vs NA	NA
CQ8: What to expect about complications?	219 vs 165	77	219 vs NA	NA	219 vs 73	0	219 vs 82	0

NOTE. Answers designated as AI1 and AI2 were obtained with the 8 CQs as prompts to ChatGPT on the same day. Answers designated Hospital1–3 were retrieved from webpages of 3 top-tier gastroenterology hospitals in the United States. Text similarity of answers was compared with CopyLeaks² and is presented as the match percentage. AI answers shared 28%–77% similarity, except the CQ7. AI answers had extremely low text similarity (0%–16%) to those from the hospital webpages. NA, not applicable.

Supplementary Table 2. Performance of 4 Interpreters in Detecting Answers Generated by AI

Detecting answers generated by AI	Sensitivity, %	Specificity, %	PPV, %	NPV, %	Accuracy, %
All interpreters (n = 4)	40.63	53.75	41.27	53.09	47.92
Senior gastroenterologists (n = 2)	15.63	57.50	22.73	46.00	38.89
Fellows (n = 2)	65.63	50.00	51.22	64.52	56.94
Senior1	6.25	55.00	10.00	42.31	33.33
Senior2	25.00	60.00	33.33	50.00	44.44
Fellow1	43.75	25.00	31.82	35.71	33.33
Fellow2	87.50	75.00	73.68	88.24	80.56

Supplementary Table 3. Example Answers From AI and Non-AI (Hospital Webpages)

CQs	AI1 answer	AI2 answer	Hospital1 answer	Hospital2 answer	Hospital3 answer
CQ1: What is a colonoscopy?	A colonoscopy is a medical procedure in which a doctor uses a long, flexible, lighted tube called a colonoscope to examine the inside of the rectum and colon (large intestine) for any abnormalities such as polyps, tumors, or signs of colorectal cancer. The procedure is performed while the patient is sedated and typically lasts 30 minutes to an hour.	A colonoscopy is a diagnostic procedure that allows a doctor to examine the inside of the large intestine (colon) and rectum. The procedure is performed using a thin, flexible tube called a colonoscope, which is equipped with a light and a camera. The doctor inserts the colonoscope through the anus and advances it into the colon to visualize the inside of the intestine. The procedure allows the doctor to identify any abnormal growths, such as polyps or cancers, as well as to assess the overall health of the colon. A colonoscopy is an important screening tool for colon cancer and can help detect the disease in its early stages, when it is most treatable. The procedure typically takes 30 minutes to an hour to complete and is performed while the patient is sedated to minimize discomfort.	A colonoscopy is a safe, private and usually painless procedure that takes about 30 minutes to complete. A colonoscopy may be done in a hospital outpatient department, clinic or in a doctor’s office. During the procedure, a general surgeon or gastroenterologist (a doctor who specializes in identifying and treating gastrointestinal diseases) uses a colonoscope to see the entire colon. A colonoscope is a long, flexible tube. If a polyp is found, the doctor may remove it during your colonoscopy. During your colonoscopy, the doctor may use the colonoscope to take a biopsy – a small tissue sample – if anything looks abnormal. The tissue sample will be sent to the lab for further testing.	A colonoscopy (koe-lun-OS-kuh-pee) is an exam used to look for changes — such as swollen, irritated tissues, polyps or cancer — in the large intestine (colon) and rectum. During a colonoscopy, a long, flexible tube (colonoscope) is inserted into the rectum. A tiny video camera at the tip of the tube allows the doctor to view the inside of the entire colon. If necessary, polyps or other types of abnormal tissue can be removed through the scope during a colonoscopy. Tissue samples (biopsies) can be taken during a colonoscopy as well.	Colonoscopy is a procedure that lets your healthcare provider check the inside of your entire large intestine or colon. The procedure is done using a long, flexible tube (colonoscope). The tube has a light and tiny camera on one end. It is put in your rectum and moved into your colon. In addition to letting your provider see the inside of your colon, the tube can be used to: Clean the lining of your colon with a water jet (irrigate) Remove any liquid stool with a suction device Inject air in your bowel to make it easier to see inside Work inside your bowel with surgical tools During a colonoscopy, your provider may remove tissue or abnormal growths (polyps) for further exam. He or she may also be able to treat problems that are found. The large intestine or colon is the last part of your digestive system. It absorbs water to change waste from liquid to solid stool. The large intestine is about 5 feet long in adults. It has 4 sections: Ascending colon. This extends upward on the right side of your belly. Transverse colon. This extends from the ascending colon across your body to the left side.

Supplementary Table 3. Continued

CQs	AI1 answer	AI2 answer	Hospital1 answer	Hospital2 answer	Hospital3 answer
CQ2: Why is a colonoscopy performed?	<p>A colonoscopy is performed for several reasons, including:</p> <p>Screening for colorectal cancer: A colonoscopy is an effective way to screen for colorectal cancer and is often recommended for people over the age of 50 or for those with a family history of the disease.</p> <p>Investigation of digestive symptoms: If a person is experiencing symptoms such as abdominal pain, rectal bleeding, or changes in bowel habits, a colonoscopy may be performed to identify the underlying cause.</p> <p>Follow-up after abnormal results: If a person has had an abnormal result on a stool test or X-ray, a colonoscopy may be performed for further evaluation.</p> <p>Monitoring of pre-existing conditions: For individuals with conditions such as inflammatory bowel disease (IBD), a</p>	<p>A colonoscopy is performed to screen for and diagnose conditions affecting the large intestine (colon) and rectum. Some common reasons for having a colonoscopy include:</p> <p>Screening for colon cancer: A colonoscopy is an important screening tool for colon cancer, which is one of the leading causes of cancer deaths in the United States. The procedure can help detect the disease in its early stages, when it is most treatable.</p> <p>Evaluating symptoms: A colonoscopy may be performed to evaluate symptoms such as abdominal pain, diarrhea, constipation, rectal bleeding, and unexplained weight loss.</p> <p>Monitoring polyps: If a polyp is found during a colonoscopy, the doctor may remove it during the procedure. Regular follow-up colonoscopies may be recommended to monitor</p>	<p>Physicians use a colonoscopy, the primary and most effective screening technique, to screen patients for colorectal cancer.</p> <p>A colonoscopy is a safe, private and usually painless procedure that takes about 30 minutes to complete. A colonoscopy may be done in a hospital outpatient department, clinic or in a doctor's office.</p> <p>During the procedure, a general surgeon or gastroenterologist (a doctor who specializes in identifying and treating gastrointestinal diseases) uses a colonoscope to see the entire colon. A colonoscope is a long, flexible tube. If a polyp is found, the doctor may remove it during your colonoscopy.</p>	<p>Your doctor may recommend a colonoscopy to:</p> <p>Investigate intestinal signs and symptoms. A colonoscopy can help your doctor explore possible causes of abdominal pain, rectal bleeding, chronic diarrhea and other intestinal problems.</p> <p>Screen for colon cancer. If you're age 45 or older and at average risk of colon cancer — you have no colon cancer risk factors other than age — your doctor may recommend a colonoscopy every 10 years. If you have other risk factors, your doctor may recommend a screen sooner.</p> <p>Colonoscopy is one of a few options for colon cancer screening. Talk with your doctor about the best options for you.</p> <p>Look for more polyps. If you have had polyps</p>	<p>Descending colon. This extends from the transverse colon downward on your left side.</p> <p>Sigmoid colon. This is named because of its S-shape. It extends from the descending colon to your rectum.</p> <p>The rectum joins the anus. This is the opening where stool passes out of your body.</p> <p>Colonoscopy can help your provider look for problems in your colon. These include:</p> <ul style="list-style-type: none"> Any early signs of cancer Red or swollen (inflamed) tissue Open sores (ulcers) Bleeding <p>Colonoscopy is also used to screen for colorectal cancer. Screening means looking for cancer in people who don't have any symptoms of the disease. A colonoscopy may be used to check and if needed treat things such as:</p> <ul style="list-style-type: none"> Colon polyps Tumors Ulceration Redness or swelling (inflammation) Pouches (diverticula) along the colon wall Narrowed areas (strictures) of the colon Any objects that might be in the colon <p>It may also be used to find the cause of unexplained, long-term (chronic) diarrhea or bleeding in the GI (gastrointestinal) tract. It can also be used to check the colon after cancer treatment.</p>

Supplementary Table 3. Continued

CQs	AI1 answer	AI2 answer	Hospital1 answer	Hospital2 answer	Hospital3 answer
<p>colonoscopy may be performed regularly to monitor the progression of the disease and check for any complications.</p>	<p>colonoscopy may be performed regularly to monitor the progression of the disease and check for any complications.</p>	<p>for new polyps or to check for the return of removed polyps.</p> <p>Inflammatory bowel disease: A colonoscopy can help diagnose and monitor conditions such as Crohn’s disease and ulcerative colitis.</p> <p>Irritable bowel syndrome: A colonoscopy may be performed to rule out other causes of symptoms and to help diagnose irritable bowel syndrome.</p> <p>Other conditions: A colonoscopy may also be performed to evaluate other conditions affecting the large intestine, such as diverticulitis, colitis, and strictures.</p> <p>It’s important to discuss the reasons for having a colonoscopy with your doctor to determine the best course of action for your individual circumstances.</p>	<p>During your colonoscopy, the doctor may use the colonoscope to take a biopsy – a small tissue sample – if anything looks abnormal. The tissue sample will be sent to the lab for further testing.</p>	<p>before, your doctor may recommend a follow-up colonoscopy to look for and remove any additional polyps. This is done to reduce your risk of colon cancer.</p> <p>Treat an issue. Sometimes, a colonoscopy may be done for treatment purposes, such as placing a stent or removing an object in your colon.</p>	<p>Colonoscopy may be used when other tests (such as a barium enema, CT colography, tests for blood in stool, stool DNA tests, or sigmoidoscopy) show the need for more testing.</p> <p>Your healthcare provider may have other reasons to advise a colonoscopy.</p>

Supplementary Table 4. Comparison of Reading Levels of Answers From AI and Non-AI (Hospital Webpages)

Source of answers	Words	Flesch-Kincaid Grade Level	<i>P</i> value	Gunning Fog Index	<i>P</i> value
			Versus eighth grade reading level ^a		
AI	175.1 (58.2)	13.1 (2.2)	<.001 ^b	15.7 (2.2)	<.001 ^b
Non-AI	168.6 (81.6)	9.0 (2.1)	.02	11.4 (2.3)	<.001 ^b
			AI vs non-AI ^c		
			<.001 ^b		<.001 ^b
			Versus eighth grade reading level ^a		
AI1	168.9 (57.7)	13.3 (2.8)	<.001 ^b	16.1 (3.0)	<.001 ^b
AI2	181.4 (61.9)	12.9 (1.5)	<.001 ^b	15.4 (1.2)	<.001 ^b
Hospital1	108.8 (45.4)	10.0 (2.0)	<.001 ^b	12.8 (2.0)	<.001 ^b
Hospital2	169.4 (68.2)	9.9 (1.4)	.001 ^b	11.8 (1.5)	<.001 ^b
Hospital3	227.3 (91.2)	7.0 (1.5)	.06	9.6 (2.5)	.08

NOTE. Data are shown as mean (standard deviation). Flesch-Kincaid Grade Level and Gunning Fog Index are well-recognized objective measures of the reading levels of texts, in which the number *x* represents the corresponding *x*th grade reading level.³⁻⁵ Measurements were performed with an online readability tool (<https://readable.com>, accessed on April 12, 2023).

^aStatistical analyses between the reading levels of the answers and the eighth grade reading level were done with 2-sided Student *t* tests.

^bBonferroni-corrected α for multiple comparisons = 0.05/16 comparisons = 0.0031; *P* < .0031 as significant.

^cStatistical analyses between the reading levels of answers from AI vs non-AI sources were done with Mann-Whitney *U* tests.