PEER-REVIEW REPORT

Name of journal: *World Journal of Gastroenterology*

Manuscript NO: 89525

Title: May ChatGPT be a tool producing medical information for common inflammatory bowel disease patients' questions? An evidence-controlled analysis

Provenance and peer review: Invited manuscript; Externally peer reviewed

Peer-review model: Single blind

Reviewer's code: 06730456

Position: Peer Reviewer

Academic degree: PhD

Professional title: Research Assistant Professor

Reviewer's Country/Territory: China

Author's Country/Territory: Italy

Manuscript submission date: 2023-11-04

Reviewer chosen by: AI Technique

Reviewer accepted review: 2023-11-04 12:43

Reviewer performed review: 2023-11-07 00:41

Review time: 2 Days and 11 Hours

Scientific quality	[ ] Grade A: Excellent [Y] Grade B: Very good [ ] Grade C: Good [ ] Grade D: Fair [ ] Grade E: Do not publish
Novelty of this manuscript	[ ] Grade A: Excellent [Y] Grade B: Good [ ] Grade C: Fair [ ] Grade D: No novelty
Creativity or innovation of this manuscript	[ ] Grade A: Excellent [Y] Grade B: Good [ ] Grade C: Fair [ ] Grade D: No creativity or innovation

Scientific significance of the conclusion in this manuscript	[ ] Grade A: Excellent [Y] Grade B: Good [ ] Grade C: Fair [ ] Grade D: No scientific significance
Language quality	[ ] Grade A: Priority publishing [Y] Grade B: Minor language polishing [ ] Grade C: A great deal of language polishing [ ] Grade D: Rejection
Conclusion	[ ] Accept (High priority) [ ] Accept (General priority) [Y] Minor revision [ ] Major revision [ ] Rejection
Re-review	[Y] Yes [ ] No
Peer-reviewer statements	Peer-Review: [Y] Anonymous [ ] Onymous
	Conflicts-of-Interest: [ ] Yes [Y] No

SPECIFIC COMMENTS TO AUTHORS

This study investigated the capacity of ChatGPT to offer medical information (MI) pertaining to inquiries frequently posed by patients with inflammatory bowel disease (IBD) to their gastroenterologists. Upon evaluating the outputs generated by ChatGPT, it became evident that this tool exhibits promising potential, albeit with notable constraints in terms of information updating and specificity, as well as occasional inaccuracies. Consequently, additional research endeavors and enhancements to ChatGPT are warranted, potentially involving the alignment of its outputs with the authoritative medical evidence available in reputable databases.This paper is very interesting and some questions still need to be answered Q1: This paper enumerates numerous inquiries, thereby prompting the query: What criteria were employed in the selection of these ten questions? Q2:The complexity of Table 2 may be mitigated by exploring alternative modes of representation. Q3:Based on the author's depiction of Q1, it is evident that CHATGPT fails to offer efficacious diagnosis and treatment. Kindly elucidate further. Q4:Further optimization is still required for language expression.

Scientific quality	[ ] Grade A: Excellent [Y] Grade B: Very good [ ] Grade C: Good [ ] Grade D: Fair [ ] Grade E: Do not publish
Novelty of this manuscript	[Y] Grade A: Excellent [ ] Grade B: Good [ ] Grade C: Fair [ ] Grade D: No novelty
Creativity or innovation of this manuscript	[Y] Grade A: Excellent [ ] Grade B: Good [ ] Grade C: Fair [ ] Grade D: No creativity or innovation

Scientific significance of the conclusion in this manuscript	[ ] Grade A: Excellent [Y] Grade B: Good [ ] Grade C: Fair [ ] Grade D: No scientific significance
Language quality	[ ] Grade A: Priority publishing [Y] Grade B: Minor language polishing [ ] Grade C: A great deal of language polishing [ ] Grade D: Rejection
Conclusion	[ ] Accept (High priority) [ ] Accept (General priority) [Y] Minor revision [ ] Major revision [ ] Rejection
Re-review	[Y] Yes [ ] No
Peer-reviewer statements	Peer-Review: [Y] Anonymous [ ] Onymous
	Conflicts-of-Interest: [ ] Yes [Y] No

SPECIFIC COMMENTS TO AUTHORS

Gravina et al. performed a timely and interesting review on the use of ChatGPT in the management of IBD. This manuscript could be of gigantic impact on GI physicians and patients. This reviewer has some minor questions. 1, Introduction, fourth paragraph: what is "hundred and sixty degrees"? Three hundred and sixty degrees? 2, Introduction, last sentence: questioned addressed by the patients? Or questions raised/confronted by the patients? 3, Is it possible to quantitatively evaluate the performance of ChatGPT? For example, what are the chances that ChatGPT may generate answers deemed satisfactory by physicians, for each of the ten questions? 4, There are different strengths of evidence for published papers. Does ChatGPT place different weight on those types of evidence?