

## **POINT-BY-POINT RESPONSE**

**Name of Journal:** *World Journal of Gastroenterology*

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

**Manuscript ID:** 89525

---

### **RESPONSE TO THE MINOR REVISIONS REQUESTED BY THE REVIEWER 1**

**Comment 1:** 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.

**Authors response:** The authors thank the reviewer for the words spent in favour of our work.

**Comment 2:** This reviewer has some minor questions. 1, Introduction, fourth paragraph: what is "hundred and sixty degrees"? Three hundred and sixty degrees?

**Authors response:** The authors apologize for this typo that we've corrected. We mean precisely "three hundred and sixty degrees," as correctly suggested by the reviewer. The authors thank the reviewer for this suggestion, indeed facilitating our manuscript's presentation and scientific quality.

**Comment 3:** 2, Introduction, last sentence: questioned addressed by the patients? Or questions raised/confronted by the patients?

**Authors response:** The authors apologize for this typo that we've corrected. We meant "questions raised by patients". The sentence is now fixed. The authors thank the reviewer for this suggestion, indeed facilitating our manuscript's presentation and scientific quality.

**Comment 4:** 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?

**Authors' response:** Being an evidence review and ChatGPT being a system that users cannot modify; it is difficult to quantify the risk of correct/incorrect information. Similarly, it is impossible to have a clear, detailed list of all LLM mechanisms that lead ChatGPT to choose or not to select specific papers over others, a prior phenomenon that has already occurred in machine learning. Only potential access to this information could weigh the risk of choosing specific sources over others. Unfortunately, the current operation of ChatGPT prevents such a complex analysis. In addition, in this evidence review, the physicians' opinions were not subjective (i.e., satisfaction) but were compared with the evidence in an aseptic manner. We aimed to check whether the outputs gave information in line with or against guidelines or papers of evidence (i.e., metanalysis) by not weighing the data quantitatively but only qualitatively. The authors thank the reviewer for this request for clarification.

**Comment 5:** 4, There are different strengths of evidence for published papers. Does ChatGPT place different weight on those types of evidence?

**Authors response:** No, the outputs provided by ChatGPT did not offer a level of evidence at all for the textual output supplied in response to typed questions. Unfortunately, this is yet another limitation of ChatGPT, and we congratulate the reviewer for bringing out an additional and significant limitation of ChatGPT. Therefore, we decided to include this in the manuscript. We have included an appropriate sentence (see in the conclusion section: "In addition, ChatGPT did not provide evidence levels for the sources employed to generate the outputs, thus removing the reader's ability to discern different degrees of quality for the same").

### **RESPONSE TO THE MINOR REVISIONS REQUESTED BY THE REVIEWER 2**

**Comment 1: 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.**

**Authors' response:** The authors thank the reviewer for thoroughly examining our paper and for the words written in favour of our work.

**Comment 2: 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?**

**Authors response:** The complete description of how these ten questions were selected is included in the section titled "Selection of ChatGPT inputs for evidence review in the scientific literature and major IBD guidelines". Following the reviewer's advice, this section has been expanded to elucidate the selection mechanisms further. The authors thank the reviewer for this suggestion, indeed facilitating our manuscript's presentation and scientific quality.

**Comment 3: Q2: The complexity of Table 2 may be mitigated by exploring alternative modes of representation.**

**Authors' response:** We thank the reviewer for this request. Indeed, Table 2 sets out in a timely and precise manner the exact answers ChatGPT provided to each output for duty and the methodological correctness of this evidence review. At this point, however, the reviewer is correct that it is a mode of exposition that is not immediate. We pondered at length among the authors and decided to invent from scratch a new figure (i.e., Figure 3) that would summarize at a glance the textual representative excerpts of the outputs provided by ChatGPT with a visual stratification question by question. In this way, we hope to have responded to the Reviewer's requests while keeping Table 2 intact, which, in our opinion, should remain in the manuscript for fairness and methodological timeliness for the reader who would like to learn more about the outputs. We hope that the Reviewer can be satisfied with this change. In any case, we thank him for providing precious advice. We

believe that with this interesting modification, the work is now better presented and more accessible in terms of readership.

**Comment 4: Q3:Based on the author's depiction of Q1, it is evident that CHATGPT fails to offer efficacious diagnosis and treatment. Kindly elucidate further.**

**Authors response:** We thank the reviewer for that comment. The section he indicated has been expanded by adding a new paragraph that catalyzes the information requested by the Reviewer. We are of the opinion that this addition was necessary and has increased the quality of our manuscript.

**Comment 5: Q4:Further optimization is still required for language expression.**

**Authors response:** We thank the reviewer for this question. The manuscript was fully revised grammatically and syntactically by an experienced reviewer. We issued a new English certificate attached to the revisions.

#### **EDITORIAL OFFICE'S AND EDITOR COMMENTS**

**Comment 1: The title of the manuscript is too long and must be shortened to meet the requirement of the journal (Title: The title should be no more than 18 words).**

**Authors response:** The authors apologize for submitting a title that is too long. The title has been shortened to fit within the journal limit (i.e., 18 words). The new title is now "May ChatGPT be a tool producing medical information for common inflammatory bowel disease patients' questions? An evidence-controlled analysis" equal to 18 words.