https://doi.org/10.1590/1806-9282.20260381

Comment on "Examining the effectiveness of ChatGPT responses to frequently asked questions by individuals with postural disorders"

Ismail Sivri[1]*, Esra Babaoglu[2], Gamze Gul[1], Tuncay Colak[1]

Dear Editor,

We read with great interest the recent article by Dursun et al. In this study, the authors aimed to evaluate the quality and readability of ChatGPT 4.0 responses to commonly asked questions about postural disorders. Ten frequently asked questions were selected from a list generated by ChatGPT and were posed to the model without follow-up prompts. The responses were independently assessed by five experts from relevant clinical disciplines using a four-grade evaluation system, and inter-rater reliability was calculated using intraclass correlation coefficients. In addition, readability was analyzed using the Flesch-Kincaid Grade Level. The authors further compared ChatGPT's responses with peer-reviewed literature to examine consistency with scientific sources. The findings suggested that most responses were rated as excellent or satisfactory, with good inter-rater reliability and an average readability level above the recommended sixth-grade standard for patient education materials[1].

However, we would like to raise a methodological concern regarding the selection of the questions. The ten questions were generated by ChatGPT itself and then selected by the researchers. They were not obtained from real patient consultations, clinical guidelines, structured surveys, or search engine data such as Google Trends. Therefore, the evaluated questions may not reflect actual patient information needs. This approach limits the study's real-world relevance and may reduce its validity. For example, in a study, the 20 most frequently searched toothache-related queries identified through Google Trends were submitted to ChatGPT[2]. Furthermore, in another study, 100 real patient records were used to generate the queries, reflecting authentic patient concerns[3].

Additionally, previous studies have shown that the readability scores of the Large Language Models' responses can vary significantly when role-based or audience-specific prompts are used. In particular, instructing the model to "explain to a medical layperson" or "explain at a 6th-grade reading level" has been associated with improved readability outcomes[4,5]. In this context, incorporating such prompts may help generate responses that are more accessible and better aligned with patient-centered communication goals.

In conclusion, while this study provides valuable insight into the potential of ChatGPT in patient education on postural disorders, methodological refinements in question selection and prompt design may enhance its real-world applicability. Future research incorporating authentic patient-derived queries and audience-specific prompting strategies could yield more robust, patient-centered evidence.

DECLARATION OF GENERATIVE AI

During the preparation of this work, the author(s) used Google Gemini 3 Pro to improve language clarity and readability. After using this tool, the author(s) reviewed and edited the content as needed and took full responsibility for the final version of the manuscript.

AUTHORS' CONTRIBUTIONS

IS: Conceptualization, Investigation, Methodology, Writing – original draft, Writing – review & editing. **EB:** Conceptualization, Investigation, Writing – review & editing. **GG:** Conceptualization, Investigation, Writing – review & editing. **TC:** Supervision, Writing – review & editing.

DATA AVAILABILITY STATEMENT

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

[1]Kocaeli University, Umuttepe Campus, Faculty of Medicine, Department of Anatomy – Izmit, Turkey.
[2]Zonguldak Bülent Ecevit University, Faculty of Medicine, Department of Anatomy – Zonguldak, Turkey.
*Corresponding author: ismail.sivri@kocaeli.edu.tr

Conflicts of interest: the authors declare there is no conflicts of interest. Funding: none.
Received on March 02, 2026. Accepted on March 21, 2026.
Scientific Editor: José Maria Soares Júnior

REFERENCES

1. Dursun B, Torlak MS, Tufekci O. Examining the effectiveness of ChatGPT responses to frequently asked questions by individuals with postural disorders. Rev Assoc Med Bras (1992). 2025;71(12):e20250750. https://doi.org/10.1590/1806-9282.20250750
2. Degirmencioglu D, Temel AN. Evaluation of the quality and readability of ChatGPT responses to toothache queries: a study based on Google trends. Cureus. 2025;17(11):e96436. https://doi.org/10.7759/cureus.96436
3. Kahan R, Shen C, Wellborn P, Lauder A, Berchuck S, Javeed H, et al. Artificial intelligence in triaging patient questions: an evaluation of a large language model for distal radius fractures. J Am Acad Orthop Surg. 2026;34(1):e106-15. https://doi.org/10.5435/JAAOS-D-25-00456
4. Eid K, Eid A, Wang D, Raiker RS, Chen S, Nguyen J. Optimizing ophthalmology patient education via ChatBot-generated materials: readability analysis of AI-generated patient education materials and The American Society of Ophthalmic Plastic and Reconstructive Surgery patient brochures. Ophthalmic Plast Reconstr Surg. 2024;40(2):212-6. https://doi.org/10.1097/IOP.0000000000002549
5. Fazilat AZ, Brenac C, Kawamoto-Duran D, Berry CE, Alyono J, Chang MT, et al. Evaluating the quality and readability of ChatGPT-generated patient-facing medical information in rhinology. Eur Arch Otorhinolaryngol. 2025;282(4):1911-20. https://doi.org/10.1007/s00405-024-09180-0