LETTER TO THE EDITOR EDİTÖRE MEKTUP

DOI: 10.5336/urology.2025-112890

Letter to the Editor: "Assessing the Performance of ChatGPT in Addressing Questions Regarding Urethral Strictures a Cross-Sectional Study"

Editöre Mektup: "Üretral Darlıklarla İlgili Sorulara Yanıt Vermede ChatGPT'nin Performansının Değerlendirilmesi Üzerine Kesitsel Bir Çalışma"

Ahmet Burak YILMAZ[a]

[a]Sincan Training and Research Hospital, Clinic of Urology, Ankara, Türkiye

Dear editor,

I read with interest the recent article by Ayrancı et al., titled "Assessing the Performance of ChatGPT in Addressing Questions Regarding Urethral Strictures a Cross Sectional Study".[1] As large language models such as ChatGPT are increasingly integrated into healthcare settings, particularly for patient education, studies assessing their validity and limitations are essential.

The methodological framework of the study is particularly noteworthy. The authors generated their question set using both frequently asked questions from reliable institutional websites and evidence-based clinical guidance from the 2022 European Association of Urology Guidelines. This dual approach allowed them to assess ChatGPT's performance across both layperson-oriented and clinician-level content. The responses were rated independently by 2 urologists using a predefined 4-point scale, with a 3rd expert resolving discrepancies—an approach that enhances the objectivity and reproducibility of the assessment.

According to the results, ChatGPT demonstrated high accuracy for general and patient-centered questions (89.5%), while its performance declined for guideline-based clinical questions (75.6%), particularly those involving treatment strategies. These findings align with previous studies suggesting that while artificial intelligence-based models perform adequately on broad medical topics, they often lack precision when addressing more complex, treatment-specific content.[2,3]

The authors also evaluated intra-model consistency by repeating the same questions across different ChatGPT sessions. The relatively stable responses, especially for basic information categories, are encouraging in terms of reliability. However, important limitations remain. As the authors

Correspondence: Ahmet Burak YILMAZ
Sincan Training and Research Hospital, Clinic of Urology, Ankara, Türkiye
E-mail: abyilmaz05@gmail.com

Peer review under responsibility of Journal of Reconstructive Urology.

Received: 19 Jun 2025 ***Accepted:*** 18 Jul 2025 ***Available online:*** 21 Aug 2025

appropriately highlight, ChatGPT does not provide source citations, is not updated in real time, and may occasionally generate inaccurate or fabricated content—a phenomenon known as hallucination.[4] These shortcomings underscore the importance of clinician oversight when artificial intelligence tools are used in healthcare communication.

From a broader perspective, this study offers valuable insights for clinicians, health communicators, and developers involved in artificial intelligence implementation. While ChatGPT and similar models may serve as cost-effective tools to support health education, they are not substitutes for professional medical advice. Their integration should be approached cautiously, particularly in contexts that require individualized clinical judgment.[2-4]

In conclusion, the study by Ayrancı et al. adds to the growing evidence base regarding the capabilities and limitations of artificial intelligence-driven conversational agents in urology.[1] It supports the potential utility of ChatGPT in patient education, provided that its content is critically evaluated and supplemented with up-to-date medical oversight. Future research incorporating patient feedback and multilingual assessments would be valuable in further determining the role of these technologies in diverse healthcare settings.

Source of Finance

During this study, no financial or spiritual support was received neither from any pharmaceutical company that has a direct connection with the research subject, nor from a company that provides or produces medical instruments and materials which may negatively affect the evaluation process of this study.

Conflict of Interest

No conflicts of interest between the authors and / or family members of the scientific and medical committee members or members of the potential conflicts of interest, counseling, expertise, working conditions, share holding and similar situations in any firm.

Authorship Contributions

This study is entirely author's own work and no other author contribution.

REFERENCES

1. Ayrancı A, Çağlar U, Yıldız O, Meriç A, Sarılar Ö, Özgör F. Assessing the performance of ChatGPT in addressing questions regarding urethral strictures: a cross-sectional study. J Reconstr Urol. 2024;14(1):16-22. doi: 10.5336/urology.2024-101725
2. Gabriel J, Shafik L, Alanbuki A, Larner T. The utility of the ChatGPT artificial intelligence tool for patient education and enquiry in robotic radical prostatectomy. Int Urol Nephrol. 2023;55(11):2717-32. PMID: 37528247
3. Yeo YH, Samaan JS, Ng WH, Ting PS, Trivedi H, Vipani A, et al. Assessing the performance of ChatGPT in answering questions regarding cirrhosis and hepatocellular carcinoma. Clin Mol Hepatol. 2023;29(3):721-32. PMID: 36946005; PMCID: PMC10366809
4. Alkaissi H, McFarlane SI. Artificial hallucinations in ChatGPT: implications in scientific writing. Cureus. 2023;15(2):e35179. PMID: 36811129; PMCID: PMC9939079