

EDITORIAL



ChatGPT in prostate cancer: myth or reality?

© The Author(s), under exclusive licence to Springer Nature Limited 2023

Prostate Cancer and Prostatic Diseases (2024) 27:9–10; <https://doi.org/10.1038/s41391-023-00750-7>

Chat.OpenAI is an AI-powered platform that provides a natural language processing (NLP) interface for building conversational agents, also known as chatbots. It is a cloud-based service that allows developers to quickly create chatbots that can understand and respond to user queries, allowing for a more natural and engaging user experience. Chat.OpenAI is based on cutting-edge AI research, including deep learning and NLP, and is designed to constantly improve its abilities through machine learning. The platform is highly flexible and customizable, allowing developers to create chatbots for a wide variety of use cases, such as customer service, personal assistants, and more. Since its introduction in March 2023, clinicians have been exploring its possible use in different areas of research. Nowadays, 36 articles have been published, and several studies are ongoing.

Cocci et al. have the merit of publishing the first urological study on ChatGPT [1]. The authors evaluated the capacity of ChatGPT to solve 100 urological cases comprising several different clinical scenarios. The accuracy of ChatGPT responses was evaluated by an experienced board-certified consultant urologist with 5 years of experience. According to their results, only 52% of all responses were deemed appropriate. Interestingly, only 11% of emergency urology cases were correctly managed by ChatGPT. This first experience in urology clearly highlights some of the expected limitations of ChatGPT technology and opens several questions and doubts. In urology, we probably have to learn from other specialties.

Suchman et al. evaluated the ability of ChatGPT to pass the American College of Gastroenterology self-assessment test [2]. According to their results, ChatGPT-3 scored 65.1% on 455 included questions, and GPT-4 scored 62.4%. Also, Gilson et al. evaluated the performance of ChatGPT on questions within the scope of the United States Medical Licensing Examination Step 1 and Step 2 exams, as well as analyzed responses for user interpretability [3]. According to their results, of the four data sets, AMBOSS-Step1, AMBOSS-Step2, NBME-Free-Step1, and NBME-Free-Step2, ChatGPT achieved accuracies of 44% (44/100), 42% (42/100), 64.4% (56/87), and 57.8% (59/102), respectively. Estomba et al. evaluated the potential of ChatGPT as a supportive tool for sialendoscopy clinical decision making. According to their results, the mean level of agreement was 3.4 (SD: 0.69; Min: 2, Max: 4) for ChatGPT's answers, while it was 4.1 (SD: 0.56; Min: 3, Max: 5) for the group of EESS ($p < 0.015$). Standing to the available evidence, ChatGPT should be used with caution given the suboptimal medical performances [4].

The ability of ChatGPT in scientific writing has also been tested. Manohar and Prasads made ChatGPT write a paper on the association between lupus and HIV, and the paper was fairly accurate. When reading the paper, it appears as a consensus paper; it was written in simple language and may not apply to specialist doctors. The style is rather robotic; however, such a style

is often present in scientific literature and, therefore, may be undistinguishable from human writing. Also, Majovski et al. evaluated the ability of artificial intelligence to generate fraudulent but authentic-looking scientific medical articles. Their study clearly highlights that the AI language model can create a highly convincing fraudulent article that resembles a genuine scientific paper in terms of word usage, sentence structure, and overall composition [5]. Last but not least, the process took no more than one hour.

Currently, this software lacks the ability for advanced critical thinking. However, in its present state, it can be valuable for recognizing common ideas from various data and literature sources. It can then present this information to skilled users who can formulate testable hypotheses. Nonetheless, the AI's frequent training may hinder innovation. The publicly accessible versions of this technology were constructed using outdated resources from 5 years ago, and its training only extends until 2021, which means that recently published data are not currently integrated. Thus far, software such as ChatGPT appears to be an intelligent tool, metaphorically speaking, assisting in the dissemination of relatively straightforward and easily understandable information. It serves as a helpful starting point. However, like conducting an immunohistochemistry experiment using an antibody with uncertain specificity, the outputs may hold little value without the involvement of inquisitive, integrative, and non-linear minds to evaluate the content. Science has always been a human endeavor, and instead of being fearful, we should embrace new technologies while preserving the human element in science. It is important to contemplate how we can prepare for the next conceptual phase, where a critically thinking, analytical AI contributes to scientific progress in the future.

An important limitation of the actual versions of ChatGPT is related to privacy and ethical issues. The output of ChatGPT is based on a probabilistic basis and may not consider these issues. The use of ChatGPT by non-expert users may easily lead to wrong information especially if the input is not correct. Unfortunately, ChatGPT suffers the same limitations as "Dr Google;" however, it has at least the ability to give balanced answers to important questions such as: "Should I treat my BPH? As an AI language model, I cannot provide personal medical advice. However, I can offer you some general information about benign prostatic hyperplasia (BPH) that may help you make an informed decision."

The possible future applications of ChatGPT in urology are infinite, especially if used as an assistant and not as a substitute for the clinician. The possible applications include patient information, radiological and histological analysis, prediction models, and therapeutic and prognostic assistance [6, 7]. As with every technological innovation, ChatGPT raised skepticism among scientists and fear of being replaced by technology. The above-mentioned experiences clearly tear away these fears, considering the suboptimal results of ChatGPT alone. However, if adequately trained, ChatGPT may become an excellent assistant. It is important to underline that ChatGPT answers should take into account social context and physician's preferences and

Received: 16 July 2023 Revised: 3 August 2023 Accepted: 18 October 2023
Published online: 10 November 2023

experience. Further studies will better define the role of ChatGPT in Urological clinical and academic practice and if it can overcome some unmet needs in urology. However, our personal impression, as well stated by the authors, is that “ChatGPT will not replace Urologists however Urologists who use ChatGPT will probably replace those who don’t.”

Riccardo Lombardo¹✉, Antonio Cicione¹,
Giuseppe Santoro¹ and Cosimo De Nunzio¹

¹*Ospedale Sant’Andrea, Sapienza University of Rome, Rome, Italy.*
✉email: rlombardo@me.com

REFERENCES

- Cocci A, Pezzoli M, Lo Re M, Russo GI, Asmundo MG, Fode M, et al. Quality of information and appropriateness of ChatGPT outputs for urology patients. *Prostate Cancer Prostatic Dis.* 2023. <https://doi.org/10.1038/s41391-023-00705-y>. In Press.
- Suchman K, Garg S, Trindade AJ. Chat generative pretrained transformer fails the multiple-choice American College of Gastroenterology Self-Assessment Test. *Am J Gastroenterol.* 2023. <https://doi.org/10.14309/ajg.0000000000002320>.
- Gilson A, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA, et al. How does ChatGPT perform on the United States medical licensing examination? The implications of large language models for medical education and knowledge assessment. *JMIR Med Educ.* 2023;9:e45312.
- Chiesa-Estomba CM, Lechien JR, Vaira LA, Brunet A, Cammaroto G, Mayo-Yanez M, et al. Exploring the potential of Chat-GPT as a supportive tool for sialendoscopy clinical decision making and patient information support. *Eur Arch Otorhinolaryngol.* 2023. <https://doi.org/10.1007/s00405-023-08104-8>.
- Májovský M, Černý M, Kasal M, Komarc M, Netuka D. Artificial intelligence can generate fraudulent but authentic-looking scientific medical articles: pandora’s box has been opened. *J Med Internet Res.* 2023;25:e46924.

- De Nunzio C, Presicce F, Lombardo R, Trucchi A, Bellangino M, Tubaro A, et al. Patient centred care for the medical treatment of lower urinary tract symptoms in patients with benign prostatic obstruction: a key point to improve patients’ care – a systematic review. *BMC Urol.* 2018;18:1–12.
- Alarcón-Zendejas AP, Scavuzzo A, Jiménez-Ríos MA, Álvarez-Gómez RM, Montiel-Manríquez R, Castro-Hernández C, et al. The promising role of new molecular biomarkers in prostate cancer: from coding and non-coding genes to artificial intelligence approaches. *Prostate Cancer Prostatic Dis.* 2022;25:431–43. <https://www.nature.com/articles/s41391-022-00537-2>.

AUTHOR CONTRIBUTIONS

RL: conception, design, writing, and approval of the final version. AC: design, revision, and approval of the final version. GS: conception, design, revision, approval of the final version. CDN: conception, revision, approval of the final version.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Correspondence and requests for materials should be addressed to Riccardo Lombardo.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.