



Transitioning from “Dr. Google” to “Dr. ChatGPT”: the advent of artificial intelligence chatbots

Enrico Checcucci¹, Severin Rodler^{2,3,4}, Pietro Piazza⁵, Francesco Porpiglia⁶, Giovanni Enrico Cacciamani^{2,3}

¹Department of Surgery, Candiolo Cancer Institute (FPO-IRCCS), Candiolo, Italy; ²Catherine and Joseph Aresty Department of Urology, Keck School of Medicine, USC Institute of Urology, Los Angeles, CA, USA; ³Artificial Intelligence Center at USC Institute of Urology, Los Angeles, CA, USA; ⁴Department of Urology, University Hospital of Munich (LMU), Munich, Germany; ⁵Division of Urology, IRCCS Azienda Ospedaliero-Universitaria di Bologna, Bologna, Italy; ⁶Department of Oncology, Division of Urology, University of Turin, San Luigi Gonzaga Hospital, Orbassano, TO, Italy

Correspondence to: Enrico Checcucci, MD. Department of Surgery, Candiolo Cancer Institute (FPO-IRCCS), Strada Provinciale 142, km 3.95, 10060 Candiolo, Turin, Italy. Email: checcu.e@hotmail.it.

Comment on: Musheyev D, Pan A, Loeb S, *et al.* How Well Do Artificial Intelligence Chatbots Respond to the Top Search Queries About Urological Malignancies? *Eur Urol* 2024;85:13-6.

Keywords: Artificial intelligence (AI); large language model (LLM); Dr. Google; Dr. ChatGPT

Submitted Dec 08, 2023. Accepted for publication Apr 09, 2024. Published online May 23, 2024.

doi: 10.21037/tau-23-629

View this article at: <https://dx.doi.org/10.21037/tau-23-629>

In the era of digital data, the Internet has emerged as a central source from which people obtain healthcare information. Patients seek information from the Internet both prior to and following the diagnosis of urological conditions, making web-based health data an integral part of the decision-making process (1). This trend of utilizing online health information is on a rapid rise, significantly influencing healthcare choices. In the last 2 years of the previous decade, the unprecedented impact of web-based research on patients' choices and orientation culminated in the term “Dr. Google” (2).

More recently, the advent of the coronavirus disease 2019 (COVID-19) pandemic further increased the digital literacy of both patients and physicians, resulting in an even wider adoption of digital consultations (3,4). Parallely, new social media platforms such as Instagram or Tik Tok emerged, where significant public attention and interaction are experienced, notwithstanding the partially poor quality and low reliability of the content (5,6).

Following this direction, the advent of artificial intelligence (AI) chatbots, based on large language model (LLM), as deep learning algorithm that can perform a variety of natural language processing (NLP) tasks, has gained great recognition even in urology (7,8).

In a recently published paper, Musheyev *et al.* (9) undertook an analysis to assess the information quality and identify instances of misinformation regarding prostate, bladder, kidney, and testicular cancers. They conducted this analysis using four AI chatbots: ChatGPT, Perplexity, Chat Sonic, and Microsoft Bing AI. As input for these AI chatbots, the study used the top five search queries related to prostate, bladder, kidney, and testicular cancers as determined by Google Trends data spanning from January 2021 to January 2023. Responses were evaluated for quality by the DISCERN scale, understandability by the Patient Education Material Assessment Tool for Printable Materials (PEMAT-P), actionability by the PEMAT-P actionability tool and readability of responses was assessed using the Flesch-Kincaid test. The findings of this study showed that AI chatbots produce information that is generally accurate (median DISCERN score 4 out of 5, range 2–5) and of moderately high quality in response to popular Google searches about urological cancers. However, their responses are fairly difficult to read, are moderately hard to understand, and lack clear instructions for users to act on (PEMAT-P 66.7%, range 44.4–90.9%). No misinformation was recorded in this study.

When discussing the role of LLM in patient education,

it is essential to weigh both its' merits and shortcomings, while considering the broader allegations of AI in healthcare information diffusion. The study's comprehensive approach is praiseworthy. It does not just focus on a single AI chatbot or a specific type of urological malignancy; instead, it spans across various AI tools and cancer types. This broad coverage provides a more holistic view of the AI chatbots' capabilities and limits in the medical information domain. In addition, the use of validated tools such as DISCERN and PEMAT-P adds value to the assessment performed, ensuring that the evaluation of chatbots is grounded in established healthcare communication standards. However, a concern arises from the study's disclosure about the lack of actionable guidance in the chatbot answers. In the area of healthcare, where information serves as a cornerstone to decision-making, the inability of AI chatbots to provide clear, actionable steps could significantly limit their role. Patients seeking information are often in search of not just understanding their condition but also seeking clear guidance on what to do next (10). If AI chatbots fail in this aspect, their role in patient education and support becomes limited, as already stated by the World Health Organization (<https://www.who.int/news/item/16-05-2023-who-calls-for-safe-and-ethical-ai-for-health>). Furthermore, the study highlights a notable gap in the understandability of the information provided. The fact that responses are often provided at a challenging reading level is a critical drawback, especially when considering the various educational levels of the potential users. However, this issue can be easily solved as already demonstrated (11) and might be due to suboptimal prompts used in the described experiments that did not include layperson translation. The absence of misinformation in AI chatbot responses, as noted in the study, is however a positive note—assuming this finding holds further validation. In an era where misinformation can spread rapidly, especially in health-related topics, it would be reassuring if these AI tools are not contributing to this issue. However, this finding should be seen with caution and is partially contradicted by similar studies (12). This is especially important, as phenomena as hallucinations are a known issue of LLM (13) and LLM can actually add to the dangerous infodemic regarding medical content (14).

In recent years, “Dr. Google” has been critically viewed as an unreliable information source (2). During online searches, although the information provider may vary in credibility, patients have the ability to discern the source and navigate to different websites for verification, as the source is transparent. In contrast, AI-powered chatbots

provide output without explicit references, and their information is not inherently filtered, potentially leading to untrue healthcare information. With no visible source, users are presented with a single, unverified option. Thus, a careful evaluation of AI-powered chatbots is imperative. It is crucial to involve physicians in shaping the medical knowledge base of these technologies. Physicians should actively participate in the development process rather than passively adopting these tools after their development.

Limitations, potentialities, and future direction of LLM

The publication clearly elucidates the limitations inherent in current LLMs associated to the acquisition and dissemination of healthcare information. It also highlights the constraints faced by end-users, particularly in terms of technological literacy, which impede their effective utilization of LLMs. For the first time in history, a significant number of laypersons are engaging with generative AI models (<https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/>). Traditionally, the use of LLMs was predominantly confined to high-level researchers, owing to the specialized knowledge required to access and utilize these programs. However, with the advent of commercially available LLMs, there is a growing inclination to apply these models for patients as end-users (12,15).

Consequently, there is a risk that LLMs may be misapplied or used for inappropriate tasks, as layperson patients may lack the necessary preparation or education to exploit the full potential of these models. As mentioned above, the quality of input and the prompts used for a LLM are critical and may have presented a limitation in this study. However, these initial experiments pose a significant question: Are LLMs, in their unrefined state, able to adequately respond to layperson inquiries? It is improbable that patients will directly interact with unrefined LLMs like ChatGPT, Perplexity, Chat Sonic, or Microsoft Bing AI in the future for medical queries. More likely, future adaption will include the use of tailored models or the implementation of output control by healthcare professionals.

To mitigate the inaccuracies of LLMs in responding to medical questions, technical solutions such as instruction fine-tuning and adjustments to the knowledge base can be employed. These interventions aim to minimize the dissemination of non-trustworthy information and enhance

the quality of communication delivered to patients (15,16). At present, these measures appear inadequate for fully addressing the concerns associated with the unsupervised use of LLMs. The complexity of ensuring accurate and reliable responses in a medical context underscores the need for further advancements and oversight in this domain.

In order to disseminate only accurate information, output control by an expert as a urologist for medical questions might be an option to deal with current limitations of LLM. In general, the less clear the knowledge base of a model is, the more the output should be controlled by experts. As for ChatGPT almost the entire Internet is the knowledge base, caution in unreflect use is paramount (12) and physicians might be involved in control of the output. It has recently been shown that this use of AI applications within healthcare has the highest trust level by patients and exceeds the use of AI alone (17). LLM can thereby become the perfect copilot for urologists and researchers alike and might increase efficacy. A potentially safe use includes summarization of information and dissemination in easy-to-understand ways as already demonstrated (11,12).

As of today, the urologist is not a tech-proven end-user as well. Urology associations are advised to proactively put the use of LLMs into their agenda. This includes the early adoption, utilization, and incorporation of LLMs into the training curricula for students, residents, and at an expert level for members of these societies. Similar to other technological innovations, quick adoption of new technologies in training is key (18). Currently, there is a notable and rapid uptake of LLMs by urologists, primarily for research purposes but also increasingly in clinical practice (19). Therefore, clear guidelines for use and application are required (20,21).

For future use, LLM can produce as output not only text, but also images or audio (22). Those features for illustration might enhance patient understanding in e. g. preoperative counseling of complex treatment situations (23). This flexibility of LLM to provide a wide range of output can make them a crucial interface between urologists, patients and advanced technologies as other AI-based models.

The capability of LLM is quickly evolving, and urologists start to embrace LLM in clinical practice. Current versions of LLMs still reveal a limited performance that has to be either controlled by experts or require technical refinements when used in the field of medicine. Therefore, it will be highly important to closely observe the changes and adoptions in technological feasibility of those technologies

in the future.

From a patient perspective, we would like to emphasize that every patient is a distinct individual with different comorbidities, social status, and support system. Adapting care to patients requires more than one-size-fits-all solutions and generic recommendations. The patient-centered urologist remains the essential part of this technological integration to analyze and frame the information provided by AI into clinical treatments and recommendations that are adapted to every patient. While AI is here to stay, it should not be replacing the fiduciary connection that the patient values.

Acknowledgments

Funding: None.

Footnote

Provenance and Peer Review: This article was commissioned by the editorial office, *Translational Andrology and Urology*. The article has undergone external peer review.

Peer Review File: Available at <https://tau.amegroups.com/article/view/10.21037/tau-23-629/prf>

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://tau.amegroups.com/article/view/10.21037/tau-23-629/coif>). S.R. receives consultancy fees from Merck, MSD and Novartis and has equity in Rocketlane Medical Ventures GmbH. The other authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

- Cacciamani GE, Bassi S, Sebben M, et al. Consulting "Dr. Google" for Prostate Cancer Treatment Options: A Contemporary Worldwide Trend Analysis. *Eur Urol Oncol* 2020;3:481-8.
- Cacciamani GE, Dell'Oglio P, Cocci A, et al. Asking "Dr. Google" for a Second Opinion: The Devil Is in the Details. *Eur Urol Focus* 2021;7:479-81.
- Porpiglia F, Amparore D, Checcucci E, et al. The revolution of congress meetings and scientific events: how to navigate among their heterogeneous modalities? *Minerva Urol Nephrol* 2021;73:3-5.
- Amparore D, Campi R, Checcucci E, et al. Patients' perspective on the use of telemedicine for outpatient urological visits: Learning from the COVID-19 outbreak. *Actas Urol Esp (Engl Ed)* 2020;44:637-8.
- Wang S, Malik RD. Social Media and Apps in Urology. *Curr Surg Rep* 2023. [Epub ahead of print]. doi: 10.1007/s40137-023-00366-9.
- Xue X, Yang X, Xu W, et al. TikTok as an Information Hodgepodge: Evaluation of the Quality and Reliability of Genitourinary Cancers Related Content. *Front Oncol* 2022;12:789956.
- Checcucci E, De Cillis S, Granato S, et al. Applications of neural networks in urology: a systematic review. *Curr Opin Urol* 2020;30:788-807.
- Checcucci E, Verri P, Amparore D, et al. Generative Pre-training Transformer Chat (ChatGPT) in the scientific community: the train has left the station. *Minerva Urol Nephrol* 2023;75:131-3.
- Musheyev D, Pan A, Loeb S, et al. How Well Do Artificial Intelligence Chatbots Respond to the Top Search Queries About Urological Malignancies? *Eur Urol* 2024;85:13-6.
- Meyrowitsch DW, Jensen AK, Sørensen JB, et al. AI chatbots and (mis)information in public health: impact on vulnerable communities. *Front Public Health* 2023;11:1226776.
- Eppler MB, Ganjavi C, Knudsen JE, et al. Bridging the Gap Between Urological Research and Patient Understanding: The Role of Large Language Models in Automated Generation of Layperson's Summaries. *Urol Pract* 2023;10:436-43.
- Davis R, Eppler M, Ayo-Ajibola O, et al. Evaluating the Effectiveness of Artificial Intelligence-powered Large Language Models Application in Disseminating Appropriate and Readable Health Information in Urology. *J Urol* 2023;210:688-94.
- Singhal K, Azizi S, Tu T, et al. Large language models encode clinical knowledge. *Nature* 2023;620:172-80.
- De Angelis L, Baglivo F, Arzilli G, et al. ChatGPT and the rise of large language models: the new AI-driven infodemic threat in public health. *Front Public Health* 2023;11:1166120.
- Singhal K, Azizi S, Tu T, et al. Publisher Correction: Large language models encode clinical knowledge. *Nature* 2023;620:E19.
- Benary M, Wang XD, Schmidt M, et al. Leveraging Large Language Models for Decision Support in Personalized Oncology. *JAMA Netw Open* 2023;6:e2343689.
- Rodler S, Kopliku R, Ulrich D, et al. Patients' Trust in Artificial Intelligence-based Decision-making for Localized Prostate Cancer: Results from a Prospective Trial. *Eur Urol Focus* 2023. [Epub ahead of print]. doi: 10.1016/j.euf.2023.10.020.
- Rodler S, Bujoreanu CE, Baekelandt L, et al. The Impact on Urology Residents' Learning of Social Media and Web Technologies after the Pandemic: A Step Forward through the Sharing of Knowledge. *Healthcare (Basel)* 2023;11:1844.
- Eppler M, Ganjavi C, Ramacciotti LS, et al. Awareness and Use of ChatGPT and Large Language Models: A Prospective Cross-sectional Global Survey in Urology. *Eur Urol* 2024;85:146-53.
- Cacciamani GE, Chu TN, Sanford DI, et al. PRISMA AI reporting guidelines for systematic reviews and meta-analyses on AI in healthcare. *Nat Med* 2023;29:14-5.
- Cacciamani GE, Collins GS, Gill IS. ChatGPT: standard reporting guidelines for responsible use. *Nature* 2023;618:238.
- Introduction to Generative AI. Available online: https://www.cloudskillsboost.google/course_templates/536
- Rodler S, Kidess MA, Westhofen T, et al. A Systematic Review of New Imaging Technologies for Robotic Prostatectomy: From Molecular Imaging to Augmented Reality. *J Clin Med* 2023;12:5425.

Cite this article as: Checcucci E, Rodler S, Piazza P, Porpiglia F, Cacciamani GE. Transitioning from "Dr. Google" to "Dr. ChatGPT": the advent of artificial intelligence chatbots. *Transl Androl Urol* 2024;13(6):1067-1070. doi: 10.21037/tau-23-629