Communication

Can Artificial Intelligence Treat My Urinary Tract Infections?—Evaluation of Health Information Provided by OpenAI™ ChatGPT on Urinary Tract Infections

Kevin Yinkit Zhuo [1,2,3,*] , Paul Kim [1,2,3] , James Kovacic [1,2], Venu Chalasani [1,2,3,4] , Krishan Rasiah [1,2,3,4] , Stuart Menogue [2,3] and Amanda Chung [1,2,3,5]

1 Department of Urology, Royal North Shore Hospital, St Leonards, NSW 2065, Australia; venu.chalasani@sydney.edu.au (V.C.)
2 North Shore Urology Research Group (NSURG), St Leonards, NSW 2065, Australia
3 Department of Urology, Northern Beaches Hospital, Frenchs Forest, NSW 2086, Australia
4 Faculty of Medicine and Health, University of Sydney, Sydney, NSW 2008, Australia
5 Department of Urology, Macquarie University Hospital, Macquarie Park, NSW 2113, Australia
* Correspondence: k.zhuo2@gmail.com

Abstract: Urinary tract infections (UTIs) are highly prevalent and have significant implications for patients. As internet-based health information becomes more relied upon, ChatGPT has emerged as a potential source of healthcare advice. In this study, ChatGPT-3.5 was subjected to 16 patient-like UTI queries, with its responses evaluated by a panel of urologists. ChatGPT can address general UTI questions and exhibits some reasoning capacity in specific contexts. Nevertheless, it lacks source verification, occasionally overlooks vital information, and struggles with contextual clinical advice. ChatGPT holds promise as a supplementary tool in the urologist's toolkit, demanding further refinement and validation for optimal integration.

Keywords: humans; teaching; internet use; urinary tract infections; delivery of health care

1. Introduction

Urinary tract infections (UTIs) affect 50% to 60% of adult women during their lifetime and account for 1% to 6% of all medical visits, imposing substantial burdens on both individuals and society [1]. As such, UTIs are frequently overseen within the community through the collaborative efforts of primary care physicians guided by urologists.

However, there is an increasing demand for primary care physician appointments to the point that demand outstrips supply, with a recent report by the Royal Australian College of General Practitioners noting a decline in the availability of free public clinics [2]. Similarly, there is a trend whereby patients are increasingly relying on the internet as a primary source of health-related information [3], with social media platforms such as YouTube™ becoming prominent outlets of health information. However, the information available on these platforms may not always reflect safe or current clinical practices [4].

OpenAI™'s ChatGPT-3.5 (ChatGPT) is a conversational artificial intelligence model available for use by the public on the internet. It is a large language model with artificial intelligence (LLMAI) designed to generate text responses that mimic human language, which employs a deep machine learning algorithm leveraging an extensive database collected by a neural network for pattern recognition and word association [5,6]. Whilst the complete extent of its utility is to be explored, there is notable interest from clinicians and patients concerning its application within healthcare [6,7].

A paucity of literature exists regarding the outcomes produced by ChatGPT in the context of patients seeking healthcare advice within the community. This study aims to evaluate the healthcare advice generated by ChatGPT regarding UTIs and assess its accuracy and safety for dissemination amongst patients.

2. Methods

A total of 16 questions mirroring common questions asked by patients were generated based upon common internet search engine queries, our clinical experience, and patient education guidelines. Information regarding known risk factors, screening and prevention options, and treatment guidelines for UTIs in the outpatient community cohort were incorporated into these questions (Table 1). These questions were graded as either elementary or advanced based upon the difficulty in answering them in a clinical setting.

Table 1. Questions with corresponding difficulty levels and scores used to assess the performance of ChatGPT.

	Question	Score
Elementary 1	What are urinary tract infections and what cause them?	5
Elementary 2	How do I know if I have a urinary tract infection?	1
Elementary 3	How can I prevent urinary tract infections?	4
Elementary 4	How many is too many urinary tract infections?	3
Elementary 5	How do you treat urinary tract infections?	1
Elementary 6	When should I see my doctor to find out if I have a urinary tract infection?	5
Elementary 7	Do urinary tract infections cause cancer?	4
Elementary 8	How much water do I need to drink to prevent urinary tract infections?	4
Advanced 9	What is the best at preventing urinary tract infections? Ural, cranberry juice, D-mannose, Hiprex or probiotics?	1
Advanced 10	I am about to start chemotherapy for my lung cancer. Should I be on prophylactic antibiotics during my course of chemotherapy for urinary tract infections?	5
Advanced 11	I got a random urine culture performed and it came back as positive for E. Coli, but I do not have any symptoms. Should I get antibiotic treatment for this, and if so, which antibiotic is best for me?	1
Advanced 12	I am a spinal cord injury patient and I perform intermittent self-catheterisation. However, I keep developing urinary tract infections despite the educator's saying I have excellent sterile intermittent self-catheterisation technique. What are the other causes for me to have these urinary tract infections?	2
Advanced 13	My doctor said my urine culture was contaminated. What does this mean?	3
Advanced 14	I am a teenager with recurrent urinary tract infections. Will this continue as I grow older?	3
Advanced 15	I have a long-term indwelling catheter and keep getting recurrent urinary tract infections. Why do not they prescribe me prophylactic antibiotics to prevent urinary tract infections?	4
Advanced 16	I am about to undergo a cystoscopy and left ureteropyeloscopy and laser lithotripsy next week. I have a stent in now and need to urinate all the time. How do I know if I have a urinary tract infection or not?	2

These questions were entered into ChatGPT, and the initial ensuing outputs were systematically collated, thereby mirroring the likely output that patients would receive (Supplementary S1). These outputs were then independently assessed by a panel of 3 urology residents and 4 practicing urologists. Each response was rated using a quantitative grading scale ranging from 1 to 5, which was predicated using criteria based on their accuracy, comprehensiveness, and patient-friendly intelligibility. A rating of 1 was indicative of responses characterised by substandard or potentially perilous information, whilst a rating of 3 denoted responses deemed reasonably satisfactory. A rating of 5 signified responses deemed to be gold standard. The ratings were collated, and any observed disparities were discussed amongst reviewers, with a new rating being assigned once consensus was achieved. Data were analysed using IBM® SPSS® Statistics V28. $p < 0.05$ were considered statistically significant.

3. Results

The median (range) length of responses was 332.5 (182 to 445) words. No statistical significance was found between the lengths of elementary and advanced question responses ($p = 0.87$). The mean (range) rating for responses was 3 (1 to 5), with ChatGPT reasonably answering 63% of questions (10/16) (Supplementary S1). A notable proportion of the responses exhibited issues. A total of 50% of responses (8/16) provided low-level evidence information, whilst 44% (7/16) provided responses that lacked detail. Elementary questions received more accurate (6/8 vs. 4/8, $p = 0.30$) responses, whilst advanced question responses were rated lower overall (2.63 vs. 3.38, $p = 0.34$) with more generic information (2/8 vs. 5/8, $p = 0.13$) and low-level evidence (5/8 vs. 3/8, $p = 0.32$). In all the responses, ChatGPT responsibly recommended consulting medical professionals for further advice, acknowledging its limitations in healthcare interactions and advocating for patient safety.

4. Discussion

Interest in LLMAIs such as ChatGPT has increased in recent years [8]. This study is one of the first to assess the information quality produced by LLMAIs such as ChatGPT in relation to UTIs as experienced by patients [5,6]. As expected, the predominant issue identified in ChatGPT-generated responses was that of inadequate comprehensiveness due to the omission or inaccuracy of specific details. This is principally due to ChatGPT's training regimen not incorporating medical databases and is further compounded by the knowledge base restriction of data up to 2021 only [5]. If an LLMAI model trained specifically using medical databases could be developed, the specific detail inaccuracies found in ChatGPT's responses could be reduced and the response quality enhanced.

Furthermore, ChatGPT's responses primarily focused on addressing queries towards female patients, as evidenced in questions 3, 4, and 14. This inclination, whilst reflective of internet trends [9], raises concerns regarding the potential exclusion of a substantial cohort of male patients who are susceptible to complex UTIs. This gender-skewed responsiveness highlights the need for improved inclusivity in online medical information to cater to a comprehensive patient demographic.

Finally, it is essential to assess ChatGPT's performance in comparison with human medical expertise. Whilst approaching a level of proficiency reminiscent of human experts, a study by Thirunavukurasu et al. determined that ChatGPT fell short of achieving the mean passing mark in the Royal College of General Practitioners' Applied Knowledge Test, a barrier fellowship exam for primary health care trainees [10]. Furthermore, ChatGPT is unable to empathetically comfort patients, thereby reaffirming the importance of emphasizing the role of seeking professional medical advice despite the advancements in AI-driven information provision. Whilst ChatGPT may serve as a valuable resource in the realm of health information dissemination, it is incumbent upon patients to recognize the limitations of AI and prioritise consultation with healthcare practitioners for comprehensive and accurate medical guidance.

5. Conclusions

ChatGPT is capable of dealing with generalised questions regarding UTIs and demonstrates a degree of reasoning aptitude in specific scenarios. Notably, it possesses the capability to disseminate information regardless of patient geographical or socioeconomical restrictions, potentially ameliorating waiting times. However, ChatGPT remains an imperfect substitute for doctors as it does not verify its source information, omits key points, and has deficiencies in delivering contextually appropriate clinical counsel. It cannot pose diagnostic questions autonomously nor empathetically comfort patients in its current iteration. Despite these limitations, ChatGPT has potential for involvement in the urologist's clinical toolkit.

Abbreviations

LLMAI   Large Language Model with Artificial Intelligence
UTI   Urinary Tract Infection

References

1. Medina, M.; Castillo-Pino, E. An introduction to the epidemiology and burden of urinary tract infections. *Ther. Adv. Urol.* **2019**, *11*, 1756287219832172. [CrossRef] [PubMed]
2. Royal Australian College of General Practitioners. General Practice Health of the Nation 2022. Updated 2022. Available online: https://www.racgp.org.au/getmedia/80c8bdc9-8886-4055-8a8d-ea793b088e5a/Health-of-the-Nation.pdf.aspx (accessed on 22 September 2023).
3. Wong, C.; Harrison, C.; Britt, H.; Henderson, J. Patient use of the internet for health information. *Aust. Fam. Physician* **2014**, *43*, 875–877. [CrossRef] [PubMed]
4. Huang, Z.; Jeong, R.; Lee, E.; Ruan, H.; Cohen, T.M.; Kim, J. MP02-06 urinary tract infection on social media: Examining youtube content as a source of patient educational information. *J. Urol.* **2022**, *207* (Suppl. S5). [CrossRef]
5. Zhou, Z.; Wang, X.; Li, X.; Liao, L. Is ChatGPT an Evidence-based Doctor? *Eur. Urol.* **2023**, *84*, 355–356. [CrossRef] [PubMed]
6. Gabrielson, A.T.; Odisho, A.Y.; Canes, D. Harnessing Generative Artificial Intelligence to Improve Efficiency among Urologists: Welcome ChatGPT. *J. Urol.* **2023**, *209*, 827–829. [CrossRef] [PubMed]
7. Cascella, M.; Montomoli, J.; Bellini, V.; Bignami, E. Evaluating the Feasibility of ChatGPT in Healthcare: An Analysis of Multiple Clinical and Research Scenarios. *J. Med. Syst.* **2023**, *47*, 33. [CrossRef] [PubMed]
8. Beam, A.L.; Drazen, J.M.; Kohane, I.S.; Leong, T.Y.; Manrai, A.K.; Rubin, E.J. Artificial intelligence in medicine. *N. Engl. J. Med.* **2023**, *388*, 1220–1221. [CrossRef] [PubMed]
9. Tam, J.; Porter, E.K.; Lee, U.J. Examination of information and misinformation about urinary tract infections on TikTok and YouTube. *Urology* **2022**, *168*, 35–40. [CrossRef] [PubMed]
10. Thirunavukarasu, A.J.; Hassan, R.; Mahmood, S.; Sanghera, R.; Barzangi, K.; El Mukashfi, M.; Shah, S. Trialling a Large Language Model (ChatGPT) in General Practice with the Applied Knowledge Test: Observational Study Demonstrating Opportunities and Limitations in Primary Care. *JMIR Med. Educ.* **2023**, *9*, e46599. [CrossRef] [PubMed]