

Original Article

Role of ChatGPT and Gemini in the Urology Field: A Case-Based Study

Rawa Bapir^{1,2,3}, Ahmed Mohammed Abdalqadir⁴, Kamran Hassan Bhatti⁵, Nadeem Sohail⁵, Bryar Othman Muhammed⁶, Ameer M. Salih⁷, Hunzalah Hassan Bhatti⁸, Hawbash M. Rahim^{9*}, Jihad Ibrahim Hama¹⁰, Hemin S. Mohammed¹, Muhammed Bag A. Ali⁶, Sabah Jalal Hasan¹, Ismaeel Aghaways⁴, Yousif M. Mahmood¹, Hussein M. Hamasalih¹

1. Smart Health Tower, Madam Mitterrand Street, Sulaymaniyah, Iraq
2. Kscien Organization for Scientific Research (Middle East office), Hamdi Street, Azadi Mall, Sulaymaniyah, Iraq
3. Department of Urology, Sulaimani Surgical Teaching Hospital, Sulaymaniyah, Iraq
4. College of Medicine, University of Sulaimani, Madam Mitterrand Street, Sulaymaniyah, Iraq
5. Urology Department, Hamad Medical Corporation, Al Khor Hospital, Qatar
6. Smart Health Tower (Raparin Branch), Karux Street, Rania, Iraq
7. Civil Engineering Department, College of Engineering, University of Sulaimani, Sulaymaniyah, Iraq
8. College of Engineering, Department of Computer Science, University of Qatar, Doha, Qatar
9. Department of Medical Laboratory Science, College of Health Sciences, University of Human Development, Sulaimani, Iraq
10. Research Center, University of Halabja, Halabja, Iraq

* **Corresponding author:** hawbash.mhamad96@gmail.com (H.M. Rahim). Mamostayan City, Building C7, Apartment 54, Zip code: 46001, Sulaymaniyah, Iraq



Keywords:

ChatGPT
Artificial intelligence
Gemini
Urology
Medical Decision-Making

Received: July 7, 2024
Revised: July 20, 2024
Accepted: July 27, 2024
First Published: August 2, 2024

Copyright: © 2024 Bapir et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Citation: Bapir R, Abdalqadir AM, Bhatti KH, Sohail D, Muhammed BO, Salih AS et al. Role of ChatGPT and Gemini in the Urology Field: A Case-Based Study. Barw Medical Journal. 2024;2(4):9-13.
<https://doi.org/10.58742/bmj.v2i3.116>

Abstract

Introduction

The healthcare sector is witnessing a transformation with the advent of artificial intelligence (AI), exemplified by ChatGPT and Gemini AI. These AI systems emulate human conversation and provide accurate medical responses. This study explores their integration into medical decision-making in the urology field.

Methods

The study presented a collection of 20 medical case scenarios, carefully crafted and revised by a team of authors in the field of urology. Each case was presented to ChatGPT and Gemini in September of 2023, and their responses were recorded and analyzed.

Results

Both AI tools displayed varying accuracy in diagnoses and management recommendations. ChatGPT failed in identifying congenital penile curvature, while Gemini succeeded. Conversely, ChatGPT excelled in recommending a management plan for renal artery aneurysms. Gemini outperformed in explaining iodinated contrast material toxicity. Both struggled with a bladder prolapse prevention question.

Conclusion

AI integration in urology is promising but has limitations. AI provides valuable insights but cannot replace human expertise. Research is vital to improve AI's role in urology. Clinicians should view AI suggestions as supplements to their judgment, fostering collaborative healthcare decisions.

1. Introduction

The healthcare sector is subjected to significant transformation with the advent of artificial intelligence (AI), which holds the potential to revolutionize patient care and outcomes. In November 2022, OpenAI unveiled an innovative natural language model known as Chat Generative Pre-Trained Transformer (ChatGPT), renowned for its ability to generate responses that closely resemble human interaction across various tasks. Additionally, Gemini AI serves as a text-based AI conversational tool, leveraging machine learning and natural language understanding to address complex inquiries. These models have the ability to produce new data by discerning structures and patterns from pre-existing data. They demonstrate versatility in generating content across diverse domains. Generative AI models rely on deep learning techniques and neural networks to scrutinize, comprehend, and produce content that closely emulates human-generated results. These two tools, ChatGPT and Gemini, have rapidly gained worldwide recognition due to their remarkable capacity to emulate human conversation and cognitive abilities. Moreover, their accuracy and reliability in delivering medical responses have been rigorously examined and confirmed [1-3].

One noteworthy advantage of ChatGPT in the realm of medical decision-making lies in its adeptness at analyzing intricate medical data. It serves as a valuable resource for healthcare professionals, offering rapid insights derived from patient records, medical research, and clinical guidelines [1,4]. Furthermore, ChatGPT can play a pivotal role in the differential diagnosis process by synthesizing information from symptoms, medical history, and risk factors, comprehensively processing this data to present a range of potential medical conditions, thus aiding medical practitioners in their assessments. This has the potential to enhance diagnostic precision, minimizing instances of misdiagnosis or delays [4].

The integration of ChatGPT and Gemini into the medical decision-making landscape has garnered attention across various medical specialties. Numerous disciplines have published articles highlighting the significance and potential applications of ChatGPT and Gemini in their respective fields [2,5].

This research aims to compare the decision-making patterns of humans and AI, exploring the potential benefits and limitations of incorporating AI tools into healthcare decision-making on a case-by-case basis.

2. Methods

A collection of case scenarios consisting of 20 inquiries and their optimal responses were crafted by a urologist. Subsequently, these cases and responses underwent revisions by additional authors. Each case was individually presented to ChatGPT and Gemini on the fourteenth of September 2023 in the same way by directly copy-pasting the cases from a Word document, and if an answer was not given directly a specific prompt by writing "Please give the most possible diagnosis for the above question." Or "Please give the best next step for the above question."

according to the question format that was given to put emphasis on the question at the end of the case and get an answer from the AI systems. The outcomes were documented in a tabulated format. The findings were subjected to scrutiny and evaluation by a group of four board-certified urologists, with two having more than twenty years of experience and the remaining two having more than five years of experience in the field of urology. The data resources and the answers for each question from ChatGPT and Gemini are summarized in (Table 1) and illustrated in (Figure 1).

3. Results

Both ChatGPT and Gemini reported a relatively low rate of precision in the spot diagnosis and proposing management plan for the questions that have been asked (7 questions, 35% for ChatGPT) and (9 questions, 45% for Gemini). ChatGPT failed to recognize a case of congenital penile curvature, while Gemini was able to answer correctly. Both ChatGPT and Gemini were able to propose the correct management plan for a case of urothelial bladder carcinoma. However, they both answered a question about preventing bladder prolapse after vesicostomy incorrectly and chose circumferential suturing of the bladder to rectus fascia rather than placing the stoma at the dome of the bladder. Gemini was more capable of answering a question about the mechanism of toxicity of iodinated contrast material, while ChatGPT answered incorrectly. ChatGPT was able to answer a question about the management plan for a 30-year-old woman with a 2 cm, circumferentially calcified saccular aneurysm of renal artery correctly, while Gemini failed to answer the question accurately (see Table 1 and Figure 1).

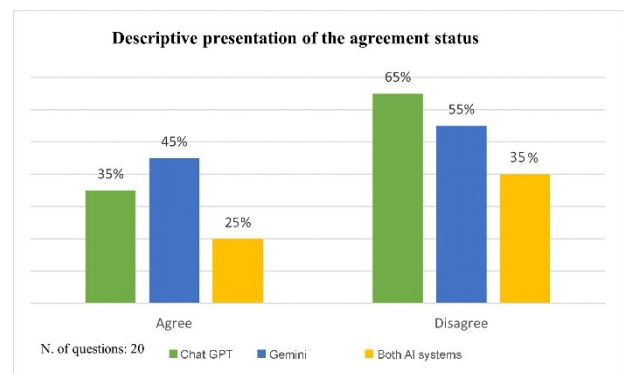


Figure 1. Description of ChatGPT and Gemini answers.

4. Discussion

The integration of AI, specifically ChatGPT and Gemini, into the field of urology, has shown promise and challenges, as evidenced by the results obtained in this study. Herein, we focused on the performance of these AI tools in urology, highlighting their strengths, weaknesses, and potential implications for clinical practice. Doddi and associates found that both ChatGPT and Gemini exhibited a relatively low rate of precision when it came to recommending the most suitable imaging modality for various urological cases [6,7]. In

Table 1. Description of The Cases and Corresponding Recommendations

Questions	ChatGPT	Gemini	Surgeon's decision	Agreements	
				ChatGPT	Gemini
A 22-year-old man states that his partner has noticed significant left lateral deviation of his penis on erection. He had no penile trauma and had not previously noticed the curvature. The most likely diagnosis is	Peyronie's disease.	Congenital penile curvature	Congenital penile curvature	Disagree	Agree
A 40-year-old man with azoospermia had testicular biopsies that revealed Sertoli cell-only histology. He and his partner prefer to have a biological child. His serum hormone testing is normal except for an elevated FSH level. He has a 46, XY karyotype and no Y chromosome microdeletions. The next step is	Microdissection testicular sperm extraction (micro-TESE).	Microdissection testicular sperm extraction (micro-TESE).	Microdissection testicular sperm extraction (micro-TESE).	Agree	Agree
A 57-year-old man completes induction and the first maintenance course of intravesical BCG for cT1 urothelial bladder carcinoma with CIS. Bladder biopsies two months after his last BCG course reveal CIS. CT scan is normal. The next step is	Radical cystectomy	Radical cystectomy	Radical cystectomy	Agree	Agree
A 30-year-old man has a gunshot wound in the suprapubic area. He is hemodynamically stable and has gross hematuria. An arterial phase CT scan shows heterogeneous fluid in the pelvis. CT cystogram reveals extraperitoneal contrast extravasation. The best treatment option is?	Exploratory laparotomy and cystorrhaphy	Suprapubic tube placement	Exploratory laparotomy and cystorrhaphy	Agree	Disagree
A 40-year-old man with a two-year history of infertility consistently has normal ejaculate volume cryptozoospermia, with rare, non-motile sperm found in the centrifuged semen pellet. Vitality testing reveals 100% viable sperm. His serum FSH level is 20 mIU/mL. The next step is	Microdissection testicular sperm extraction (micro-TESE).	Microdissection testicular sperm extraction (micro-TESE).	In vitro fertilization with intracytoplasmic sperm injection (IVF/ICSI).	Disagree	Disagree
A 70-year-old man seeks surgical correction of refractory LUTS, after failing medical therapy. He takes apixaban for atrial fibrillation. TRUS reveals a 125-gram prostate and an intravesical lobe. The best treatment option is	TURP (Transurethral Resection of the Prostate)	TURP (Transurethral Resection of the Prostate)	Laser prostatectomy	Agree	Agree
A seven-year-old boy with hematuria has a 2 cm non-obstructing right distal ureteral stone. The best treatment option is	Swl (shock wave lithotripsy).	Ureteroscopy and lithotripsy	Open ureterolithotomy	Disagree	Disagree
A 55-year-old woman developed a vesicovaginal fistula nine days after a vaginal hysterectomy. Cystoscopy shows a 5 mm area of erythema just cephalad to the trigone. Bilateral retrograde pyelograms are normal. Although vaginal leakage of urine persists, the majority of urine is voided normally. The next step is	Immediate fistula repair	Urethral catheter drainage	Urethral catheter drainage	Disagree	Agree
A 60-year-old man previously treated with pelvic XRT for colon cancer undergoes radical prostatectomy. Four weeks after surgery, a 2 cm rectourethral fistula is noted. The next step is catheter drainage and	Fecal diversion and staged fistula repair	Fecal diversion and staged fistula repair	Fecal diversion and staged fistula repair	Agree	Agree
A 35-year-old man with fever of 39.8° C undergoes incision and drainage of a perineal abscess and broad spectrum IV. antibiotics administration. The next morning, urine starts to drain from the wound with voiding. The next step is	Suprapubic cystostomy	Urethral catheter	Suprapubic cystostomy	Disagree	Agree
The action that will most effectively lower the patient's radiation exposure, when performing fluoroscopy, is to	Use tight collimation	Use tight collimation	Move the image intensifier closer to the patient	Disagree	Disagree

Table 1. Continued...

Four months after a complex ureteroscopic stone extraction, a 13-year-old boy developed a 1 cm lower ureteral stricture. The next step is	Balloon dilation and stent placement	Balloon dilation and stent placement	Ureteral reimplantation	Disagree	Disagree
A 30-year-old woman has a 2 cm, circumferentially calcified saccular aneurysm on renal arteriography. Her blood pressure is 130/82 mmHg and her creatinine is 1.1 mg/dL. She is considering pregnancy. The next step is	Surgical repair	Serial imaging	Surgical repair	Agree	Disagree
Four weeks after starting hydrochlorothiazide for hypercalcemia due to renal calcium leak, a 40-year-old calcium stone former reports fatigue. The next step is	Check serum calcium and phosphorous	Check serum sodium and potassium	Check serum sodium and potassium	Disagree	Agree
The mechanism of action of Alpha-mercaptopyronylglycine to prevent cystine stone formation is	Decreasing cystine excretion	Forming drug-cystine complexes	Forming drug-cystine complexes	Disagree	Agree
A 14-year-old boy with a history of CAH has painful bilateral testicular masses confirmed on ultrasound. The next step is	Bilateral partial orchiectomy	Fine needle aspiration of testis	Increase corticosteroids.	Disagree	Disagree
The most common pure stone composition in patients with gout treated with allopurinol is	Xanthine	Uric acid	Calcium oxalate monohydrate.	Disagree	Disagree
Opioid abuse may interfere with male fertility through	Secondary hypogonadism	Secondary hypogonadism	Secondary hypogonadism	Agree	Agree
The renal toxicity of IV. Iodinated contrast material is due to the mechanism of renal toxicity of iodinated IV contrast is due to	Afferent arteriolar constriction	Intrarenal vasoconstriction and tubular necrosis	Intrarenal vasoconstriction and tubular necrosis	Disagree	Agree
The best way to prevent vesicostomy prolapse is by	Circumferential suturing of bladder to rectus fascia	Circumferential suturing of bladder to rectus fascia	Placing the stoma at the bladder dome	Disagree	Disagree

accordance, the findings in this study confirmed the previous finding. This underscores the complexity of medical decision-making and the limitations of current AI models in handling nuanced clinical scenarios. The relatively low accuracy rates of 35% for ChatGPT and 45% for Gemini suggest that these AI tools are not yet reliable substitutes for human clinicians in urological diagnosis and treatment planning.

Cakir et al. found that ChatGPT performed impressively in responding to urolithiasis-related questions. It achieved a high accuracy rate of 94.6% in answering frequently asked questions (FAQs) and consistently provided correct responses in areas such as general information, diagnosis, and ureteral stones. Additionally, when evaluated against European Association of Urology (EAU) guidelines, ChatGPT maintained a strong accuracy rate, with 83.3% of questions receiving the highest rating of "completely correct." The chatbot's ability to consistently provide accurate and reliable information highlighted its potential to support patients and healthcare professionals in the context of urolithiasis [8].

In this study, the discussion revolved around the variations in performance between ChatGPT and Gemini on specific medical

questions. While ChatGPT excelled in some instances, correctly answering complex medical queries, it also exhibited limitations in recognizing certain conditions. Conversely, Gemini provided accurate responses in situations where ChatGPT faltered. This comparison underscored the importance of carefully choosing AI tools based on their domain-specific capabilities, as the choice of AI system can significantly impact the accuracy and reliability of responses in medical contexts. The current study revealed a trend in responses where both ChatGPT and Gemini indicated a comparatively low level of accuracy in spot diagnosis of urological cases. Furthermore, they both provided inaccurate responses to a question regarding preventing bladder prolapse after vesicostomy. Instead of suggesting placing the stoma at the dome of the bladder, they incorrectly recommended circumferential suturing of the bladder to the rectus fascia. Pointing to the limitations of AI systems in understanding the subtleties and nuances of clinical guidelines and best practices. However, more research is needed to confirm and understand the trends these AIs give regarding their answers in the field of urology. A plausible explanation for why these systems chose wrong answers can be attributed to insufficient exposure to diverse or specific datasets on certain subjects, leading to inaccurate responses. Furthermore, they often fall short in

possessing the common-sense reasoning and intuition that humans inherently have [9].

While AI has shown promise in assisting healthcare professionals, it is essential to recognize that AI algorithms are only as good as the data they are trained on; for instance, ChatGPT-3 underwent training on an extensive corpus of textual data, encompassing web pages, books, and various written materials, through a language modelling task. This makes it stand out as one of the most extensive and robust language models ever devised, incorporating 175 billion parameters [3]. Due to their training on vast datasets from the internet, the potential for these AIs to internalize and perpetuate biases inherent in the training data is a valid concern. To address this issue, it is imperative to formulate strategies for mitigating biases in AI models and incorporate algorithms that prioritize fairness awareness during development, which is especially important in the field of medicine to prevent biases regarding patient care and the decision-making capabilities of these AIs from being a more reliable resource in the future for medical professionals. In addition, AI may lack the contextual understanding and clinical experience that human experts possess. Hence, the integration of AI into healthcare should be approached with caution, with clinicians using their expertise to validate and refine AI-generated suggestions to ensure the highest quality of care [3,10]. There are ethical concerns regarding AI; for example, ChatGPT has the potential to be misused for malicious reasons, such as disseminating misinformation and creating fake news. Furthermore, they can generate content that is not only highly persuasive but also emotionally resonant. Indeed, it is crucial to address these risks proactively and uphold responsible and ethical usage of sophisticated language models [3].

One of the limitations of this study is that the questions were not tested for consistency by asking the AI systems the same question multiple times and to check if the answers differ between the first and later responses. The results of this study emphasize the importance of considering AI as a complementary tool rather than a replacement for human expertise in urology. AI can serve as a valuable resource for information retrieval and decision support, potentially speeding up the decision-making process and reducing the cognitive load on clinicians.

5. Conclusion

AI integration in urology, exemplified by ChatGPT and Gemini, holds promise but also has limitations. These tools offer valuable insights, yet their diagnostic and management capabilities still fall short of human clinicians'. Ongoing research is vital to improving AI in urology, and AI can be considered a complementary tool rather than a replacement for human expertise in the field.

Declarations

Conflicts of interest: The author(s) have no conflicts of interest to disclose.

Ethical approval: Not applicable.

Patient consent (participation and publication): Not applicable.

Funding: The present study received no financial support.

Acknowledgements: None to be declared.

Authors' contributions: RB was a major contributor to the conception of the study and the literature search for related studies. NS, BOM, AMS, HHB, JIH, and HSM were involved in the literature review, the study's design, and the critical revision of the manuscript, and they participated in data collection. HMR, AMA, and KHB were involved in the literature review, study design, and manuscript writing. MAA, SJH, IA, YMM, and HMH Literature review, final approval of the manuscript, and processing of the tables. BRB and HHB confirm the authenticity of all the raw data. All authors approved the final version of the manuscript.

Use of AI: AI was not used in the drafting of the manuscript, the production of graphical elements, or the collection and analysis of data.

Data availability statement: Not applicable.

References

1. Malik P, Pathania M, Rathaur VK. Overview of artificial intelligence in medicine. *Journal of family medicine and primary care*. 2019;8(7):2328-31. [doi:10.4103/jfmpe.jfmpe.440.19](https://doi.org/10.4103/jfmpe.jfmpe.440.19)
2. Aydin Ö. Google Bard Generated Literature Review: Metaverse. *Journal of AI*. 2023;7(1): 1-14. [doi:10.61969/jjai.1311271](https://doi.org/10.61969/jjai.1311271)
3. Ray PP. ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems*. 2023 Apr 14. [doi:10.1016/j.iotcps.2023.04.003](https://doi.org/10.1016/j.iotcps.2023.04.003)
4. 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 Medical Education*. 2023;9:e45312 [doi:10.2196/45312](https://doi.org/10.2196/45312)
5. Ayers JW, Zhu Z, Poliak A, Leas EC, Dredze M, Hogarth M, et al. Evaluating Artificial Intelligence Responses to Public Health Questions. *JAMA Network Open*. 2023;6(6):e2317517. [doi:10.1001/jamanetworkopen.2023.17517](https://doi.org/10.1001/jamanetworkopen.2023.17517)
6. Doddi S, Hibshman T, Salichs O, Bera K, Tippareddy C, Ramaiya N, Tirumani SH. Assessing Appropriate Responses to ACR Urologic Imaging Scenarios using ChatGPT and Bard. *Current Problems in Diagnostic Radiology*. 2023 Oct 20. [doi:10.1067/j.cpradiol.2023.10.022](https://doi.org/10.1067/j.cpradiol.2023.10.022)
7. Caruccio L, Cirillo S, Polese G, Solimando G, Sundaramurthy S, Tortora G. Can ChatGPT provide intelligent diagnoses? A comparative study between predictive models and ChatGPT to define a new medical diagnostic bot. *Expert Systems with Applications*. 2023;121186. [doi:10.1016/j.eswa.2023.121186](https://doi.org/10.1016/j.eswa.2023.121186)
8. Cakir H, Caglar U, Yildiz O, Meric A, Ayranci A, Ozgor F. Evaluating the performance of ChatGPT in answering questions related to urolithiasis. *International Urology and Nephrology*. 2023;1-5. [doi:10.5152/iud.2023.23171](https://doi.org/10.5152/iud.2023.23171)
9. Harika J, Baleeshwar P, Navya K, Shanmugasundaram H. A review on artificial intelligence with deep human reasoning. In 2022 International Conference on Applied Artificial Intelligence and Computing (ICAIC) 2022 9 (pp. 81-84). IEEE. [doi:10.1109/ICAIC53929.2022.9793310](https://doi.org/10.1109/ICAIC53929.2022.9793310)
10. Wang F, Casalino LP, Khullar D. Deep learning in medicine—promise, progress, and challenges. *JAMA Internal Medicine*. 2019;179(3):293-4. [doi:10.1001/jamainternmed.2018.7117](https://doi.org/10.1001/jamainternmed.2018.7117)