Training ChatGPT Models in Assisting Urologists in Daily Practice

Ioannis MANOLITSIS[a,1,†], Georgios FERETZAKIS[b,c,†], Lazaros TZELVES[a],
Dimitris KALLES[b], Stamatis KATSIMPERIS[a], Panagiotis ANGELOPOULOS[a],
Athanasios ANASTASIOU[d], Dimitrios KOUTSOURIS[d], Thanos KOSMIDIS[e],
Vassilios S. VERYKIOS[b], Andreas SKOLARIKOS[a] and Ioannis VARKARAKIS[a]

[a]*Second Department of Urology, National and Kapodistrian University of Athens,
Sismanogleio General Hospital, Athens, Greece*
[b]*School of Science and Technology, Hellenic Open University, Patras, Greece*
[c]*Sismanogleio General Hospital, Department of Quality Control, Research and
Continuing Education, Marousi, Greece*
[d]*National Technical University of Athens (NTUA)*
[e]*CareAcross Ltd, London, United Kingdom*
† *Equal contribution (1st authors)*

Abstract. Artificial Intelligence (AI) has shown the ability to enhance the accuracy and efficiency of physicians. ChatGPT is an AI chatbot that can interact with humans through text, over the internet. It is trained with machine learning algorithms, using large datasets. In this study, we compare the performance of using a ChatGPT API 3.5 Turbo model to a general model, in assisting urologists in obtaining accurate, valid medical information. The API was accessed through a Python script that was applied specifically for this study based on 2023 EAU guidelines in PDF format. This custom-trained model leads to providing doctors with more precise, prompt answers about specific urologic subjects, thus helping them, ultimately, providing better patient care.

Keywords. ChatGPT, Large Language Models, Artificial Intelligence, Urology

1. Introduction and Background

Artificial Intelligence (AI) has become an ever-growing factor in our daily lives, as it has demonstrated its utility in many fields [1]. In medicine, it has showcased its potential to enhance the work of physicians, and contribute enormously to their efficiency, accuracy, and ultimately, patient outcomes. As a result, medical training and research are also greatly benefited [2]. An example of an AI system that creates interaction between humans and computers, is a chatbot, a computer program that imitates human language understanding (and generation) through Natural Language Processing (NLP) [3] and can instigate a conversation with human users over the internet. ChatGPT (generative pre-trained transformer) is an AI chatbot built on an

---

[1] Corresponding Author: Ioannis MANOLITSIS, MD, PhD(c), Second Department of Urology, National and Kapodistrian University of Athens, Sismanogleio General Hospital, Athens, Greece; E-mail: giannismanolit@gmail.com.

OpenAI large language model (LLM) that can generate human-like text, using a machine learning algorithm trained on vast datasets [4]. In our current study, we investigate the use of a custom ChatGPT model in comparison with the ChatGPT general model, in providing credible information to urologists and assisting them, in their clinical practice.

2. Methods and Materials

In this study, we used the ChatGPT API 3.5 Turbo, which is a language model developed by OpenAI that can generate natural language text based on prompts provided by the user. The API was accessed through a Python script that was developed specifically for this study. Another key material used was the 2023 EAU guidelines in PDF format [5]. We also used a PDF parsing tool, PyMuPDF [6] library, which is an enhanced Python binding for MuPDF to extract data from PDF files. We process embeddings using the OpenAI tokenizer tiktoken [7] as in this example given by OpenAI [8]. The pdf document that is used in our study, has 34745 characters and contains 9121 tokens, which the OpenAI's GPT-3.5-turbo cannot handle all at once. Thus, we break that document up into multiple sentences that are no longer than 500 tokens [9]. In addition, we have to state that the API was run, based on default settings. The series of steps that this procedure follows are: 1) preprocess the contextual information by splitting it into chunks and creating an embedding vector for each chunk. 2) When a query is received, embed the query in the same vector space as the context chunks and find the context embeddings which are most similar to the query. 3) Prepend the query prompt with the most relevant context embeddings. 4) Submit the question to GPT along with the context that is most relevant, and receive an answer which makes use of the provided contextual information [10].

3. Results

The ChatGPT models were handed with the same specific questions based on the guidelines of the European Association of Urology (EAU) for 2023, regarding the treatment of urinary tract infections [5]. However, the custom model was trained with more specialized content on the given subject (the pdf file of 2023 EAU guidelines), while the general ChatGPT model [10] replied to our questions, based on public datasets already trained on. For example, some of the answers we received were:
Question 1:
"What is the strength rating of D-mannose use in reducing recurrent urinary tract infections according to the EAU guidelines?"
Answer to Question 1of ChatGPT: *"According to the 2020 European Association of Urology (EAU) guidelines on urinary tract infections (UTIs), the strength of recommendation for D-mannose in reducing recurrent UTIs is "weak" (grade C). This means that the recommendation is based on low-quality evidence or conflicting evidence, and the balance between benefits and risks is uncertain. The guidelines state that D-mannose is a sugar that can interfere with the ability of bacteria to adhere to*

the urothelium, which is the inner lining of the urinary tract. Some studies have suggested that D-mannose may be effective in reducing the recurrence of uncomplicated UTIs in women, but the evidence is limited and the optimal dosing and duration of treatment are not well established. The EAU guidelines recommend that D-mannose can be considered as an option for women with recurrent uncomplicated UTIs who prefer non-antibiotic prophylaxis or have adverse reactions to antibiotics. However, the guidelines also note that D-mannose is not a substitute for antibiotic treatment of acute UTIs, and that its long-term safety and efficacy have not been well established."

Answer to Question 1by using Chat GPT API 3.5 Turbo: *"The strength rating of D-mannose use in reducing recurrent urinary tract infections is overall weak and contradictory, according to the provided text."*

Question 2:

"What is the recommendation of the EAU guidelines regarding the use of fluoroquinol ones for prostate biopsy prophylaxis?"

Answer to Question 2 of Chat GPT: "*The European Association of Urology (EAU) guidelines recommend against the routine use of fluoroquinolones for prostate biopsy prophylaxis due to concerns about increasing resistance to these antibiotics. This recommendation is based on the increasing prevalence of fluoroquinolone-resistant bacteria, as well as the potential risks associated with these antibiotics, including tendon rupture, neuropsychiatric events, and aortic aneurysm.Instead, the EAU guidelines recommend alternative antibiotic regimens for prostate biopsy prophylaxis, depending on local resistance patterns and patient factors. For example, in regions with low resistance rates, a single-dose parenteral antibiotic, such as a third-generation cephalosporin or aminoglycoside, may be used. In regions with high resistance rates, a combination of antibiotics, such as a cephalosporin plus an aminoglycoside or an extended-spectrum penicillin plus an aminoglycoside, may be required.In addition to antibiotic prophylaxis, the EAU guidelines also recommend other measures to reduce the risk of infectious complications after prostate biopsy, such as rectal preparation with antiseptic solutions, aseptic technique during the procedure, and patient education on signs and symptoms of infection."*

Answer to Question 2 by using Chat GPT API 3.5 Turbo: *"Do not use fluoroquinolones for prostate biopsy in line with the European Commission final decision on EMEA/H/A-31/1452."*

4. Discussion and Conclusion

The structure of these responses clearly demonstrates that the custom-trained models reply in a shorter and more specific manner. In addition, custom-trained models could be also trained by feeding them with anonymized patient information derived from hospital electronic health records (EHR). Thus, enabling these models to be even more efficient and provide precise, patient-centric information to physicians. Because of these leaping advancements, there has been concern about the potential replacement of physicians by AI. Doctors equipped with such technologies, on the other hand, can

become even more invaluable to their patients, as AI augments their work and efficiency [11]. It is crucial to think about any potential problems before employing LLMs for theoretical applications, to ensure their dependability and safety. When using LLMs for research data extraction, it is possible that the model may provide incorrect or biased conclusions, leading to low-quality data. Similarly, given the ethical and legal ramifications, it is important to approach the use of LLMs in patient management with prudence. It is not clear who is accountable if something goes wrong if clinicians base their choices on the information given by an LLM. The potential safety repercussions of deploying unverified or unvalidated Artificial Intelligent Systems in clinical settings are raised by this lack of accountability [12]. Thus, it is crucial to develop precise standards and rules for their use after LLMs have been adopted in clinical settings. There are two main key take-away messages from the current study. The first is that large language models like ChatGPT indeed have a role in helping physicians, by providing information that is of great significance in their daily practice. The second message is, that by training the models with very specific content on the subjects we inquire about each time, we increase the credibility of the information we receive as feedback from these models, thus contributing, ultimately, to what is important, optimization of healthcare.

References

[1] Adamopoulou E, Moussiades L. An overview of chatbot technology. InArtificial Intelligence Applications and Innovations: 16th IFIP WG 12.5 International Conference, AIAI 2020, Neos Marmaras, Greece, 2020, Proceedings, Part II 16 2020 (pp. 373-383). Springer International Publishing.

[2] Tzelves L, Manolitsis I, Varkarakis I, Ivanovic M, et al. Artificial intelligence supporting cancer patients across Europe-The ASCAPE project. PLoS One. 2022 Apr 21;17(4): e0265127.

[3] 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 Mar 4;47(1):33. doi: 10.1007/s10916-023-01925-4.

[4] Eysenbach G. The Role of ChatGPT, Generative Language Models, and Artificial Intelligence in Medical Education: A Conversation With ChatGPT and a Call for Papers. JMIR Med Educ. 2023 Mar 6;9:e46885

[5] Bonkat G, Bartoletti R, Bruyere F, Cai T, Geerlings SE, Köves B, Schubert S, Wagenlehner F, Mezei T, Pilatz A, Pradere B. EAU guidelines on urological infections. European Association of Urology. 2017 Nov 24;18:22-6. ISBN 978-94-92671-19-6

[6] PyMuPDF, J. X. McKie, Pymupdf: Python bindings for mupdf's rendering library., https://github.com/pymupdf/PyMuPDF, 2012. (accessed on 4 March 2023)

[7] https://github.com/openai/tiktoken (accessed on 4 March 2023)

[8] https://github.com/openai/openai-cookbook/blob/main/examples/Question_answering_using_embeddings.ipynb (accessed on 4 March 2023)

[9] https://github.com/skandavivek/ChatGPT-API (accessed on 20 March 2023)

[10] OpenAI. *ChatGPT*. chat.openai.com/. Accessed 16 March. 2023.

[11] Giordano C, Brennan M, Mohamed B, Rashidi P, et al. Accessing Artificial Intelligence for Clinical Decision-Making. Front Digit Health. 2021 Jun 25;3:645232

[12] Naik N, et al. Legal and ethical consideration in artificial intelligence in healthcare: who takes responsibility?. Frontiers in surgery. 2022:266.