Evaluating Patient and Otolaryngologist Dialogues Generated by ChatGPT, Are They Adequate?

Oguzhan Topsakal  
Florida Polytechnic University

Tahir Cetin Akinci (✉ tahircetin.akinci@ucr.edu)  
University of California, Riverside

Mazhar Celikoyar  
Demiroğlu Bilim University

Research Article

**Keywords:** ChatGPT, Machine Learning, Large Language Models, Ear Nose Throat, Otolaryngology, Patient-Doctor Interview, Patient-Doctor Dialogue

**Posted Date:** April 3rd, 2023

**DOI:** https://doi.org/10.21203/rs.3.rs-2719379/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

AI applications are becoming more and more prevalent each day. ChatGPT is a recent AI tool that has amazed many people with its capabilities. It is expected that large language model solutions like ChatGPT will provide unique solutions and transform many industries. In many medical educational institutions, it is desired that medical students experience simulated patient encounters before meeting with real patients. These simulations can be designed to closely mimic the experience of a real-life patient encounter, allowing students to practice communication and history-taking skills in a realistic setting. Designing dialogues for these simulations is an important and time-consuming challenge. In this study, we evaluate if ChatGPT, an AI tool based on GPT-3, can generate adequate patient-doctor dialogues that can be utilized for medical student training. We analyze patient-doctor dialogues generated by ChatGPT for ten common ENT diseases and discuss the pros and cons of these dialogues. We believe the patient-doctor dialogues provided by ChatGPT can be a good starting point for teaching medical students how to communicate with patients.

I. Introduction & Related Work

AI applications and solutions are becoming a part of daily life [1]. As AI pioneer, Andrew Ng, said, “AI is the new electricity” and will change our lives in the near future [2]. With the public release of ChatGPT many people started to think the future is nearer than we thought [3].

Language modeling studies date back to the 1950s. Artificial intelligence studies have gained momentum and slowed down several times since the 1960s [4]. In the 2000s, many studies were carried out in the field of language modeling, and as a result of these studies, significant progress was made [5]. In the 2010s, a large number of major language models were trained, and the performance of these models improved continuously [6]. The size of the training datasets often consists of billions of sentences resulting in the model learning about language structure and meaning from this data [7]-[9]. In the literature, many studies have been done on the healthcare and medicine of Artificial Intelligence [10]-[16].

A patient-doctor interview is a conversation between a healthcare provider and a patient in which the provider gathers information about the patient’s health, medical history, current medications, and current symptoms in order to make a diagnosis and develop a treatment plan. There are several techniques that doctors may use during a patient-doctor interview in order to effectively gather this information [17]-[19]:

- **Open-ended questioning**: This involves asking broad, open-ended questions that allow the patient to fully describe their symptoms, feelings, and concerns.
- **Active listening**: This involves paying close attention to what the patient is saying, making eye contact, and using verbal and nonverbal cues to show that the doctor is listening and understanding.
- **Clarifying**: This involves asking follow-up questions or repeating back what the patient has said in order to make sure that the doctor fully understands the patient’s concerns.
• **Summarizing**: This involves periodically summarizing what the patient has said in order to ensure that the doctor has accurately captured the key points of the conversation.

• **Using teaching moments**: This involves using the patient-doctor interview as an opportunity to educate the patient about their health, provide information about treatment options, and answer any questions that the patient may have.

A comprehensive patient-doctor interview (patient history-taking) is a vital aspect of the diagnostic process in medicine [20], [21]. Therefore, training in patient history-taking is an essential part of the education of medical students [22], [23].

There are several methods through which medical students can receive training in patient history-taking. These may include simulated patient encounters, clinical rotations, and the use of educational software [24]-[26]. Simulated patient encounters involve students practicing taking a patient history with actors who are trained to present with specific conditions, while clinical rotations provide students with the opportunity to practice taking patient histories under the supervision of a physician or other healthcare professional [27]-[29]. Educational software may include online tutorials, interactive case simulations, and virtual patient encounters. These tools can help students develop their skills in patient history-taking and understand the importance of this aspect of the diagnostic process [30]-[34].

Virtual reality (VR) software refers to computer programs that create a simulated, immersive environment for users to interact with. In the context of education about patient history taking and doctor-patient interviews, VR software is used to provide medical students and other healthcare professionals with realistic, interactive educational experiences [35]-[37]. VR software for patient history-taking training may include virtual patient encounters where students interact with a simulated patient and practice gathering information about their symptoms, past medical history and current medications. These simulations can be designed to closely mimic the experience of a real-life patient encounter, allowing students to practice communication and story-taking skills in a realistic setting [38], [39].

Utilizing virtual patient encounters, VR software teaches medical students the doctor-patient interview process to improve their communication skills, such as how to relate to patients, ask open-ended and closed-ended questions, and interpret non-verbal cues. Students can get hands-on training using VR software and practice these skills in a safe, controlled environment [40], [41]. A simulation has been developed by creating "Virtual Patients" in a software program that simulates patient-doctor interviews at the University of Liverpool and the University of Florida [42]. In this software, a dataset of virtual patient cases has been created for students to practice patient history taking and diagnosing. Another example in this field is the "Simulated Patient Encounter" developed by the American Medical Association.

In addition, many academic studies have been carried out in the literature with AI in generating patient-doctor interviews/dialogues [43]-[47] and training in the field [48]-[50]. More recently, "Deep Learning for Natural Language Processing in Healthcare" is an artificial intelligence tool that uses deep learning algorithms to create patient-doctor conversations based on real-life patient data [43], [45], [50], [51].
Designing dialogues is an important and time-consuming challenge. They need to provide medically reliable information, and the consequences of events in the dialogue must reflect real-life situations. To overcome this challenging task, large language natural language processing (NLP) models, a specialized type of AI model, can be utilized. GPT3 (Generative Pre-training Transformer) is a state-of-the-art large language model developed by OpenAI that can be used for a variety of language tasks, such as language translation, question answering, and text summarization.

In the context of a patient-doctor interview, GPT-3 could be used to generate a conversation between a patient and a doctor based on a set of prompts or input provided by the user. For example, the user could provide information about the patient's symptoms, and the GPT-3 model could generate responses from the doctor, such as questions about the patient's medical history or recommendations for treatment. This could be used as a tool to help people practice and prepare for real-life patient-doctor interviews, or as a way to generate realistic dialogues for research or training purposes [52],[53].

Our goal in this study is to evaluate if ChatGPT, an AI tool based on GPT-3, can generate adequate patient-doctor dialogues that can be utilized for teaching medical students how to have a patient-doctor interview.

II. Methodology

We asked questions (prompts) to ChatGPT to generate patient-doctor dialogues for common ENT diseases. ChatGPT prepared ten dialogues for the following diseases: throat infection, sinus infection, ear infection, allergy symptoms, middle ear infection, esophageal dysphagia, sinusitis, esophagitis, and nasal polyps. The prompts that were asked to ChatGPT are listed in Table 1. When ChatGPT started to provide dialogues for the same disease (throat infection), we adjusted the prompt so that we did not get dialogues for the same disease. For each disease, the prompts listed in Table 1 are repeated. After the response to the first prompt, we asked if ChatGPT could continue the dialogue. On the sixth prompt, we asked, “How the dialogue would be if the test results did not confirm to the initial diagnosis” to see how the dialogue could progress in a different direction.

The responses to the prompts given in Table 1 are provided in the supplementary document (Patient-ENTDoctor-Interview.xlsx).

<table>
<thead>
<tr>
<th>TABLE I</th>
</tr>
</thead>
<tbody>
<tr>
<td>ChatGPT Prompts to Get A Dialogue Between A Patient And An Ent Doctor About A Common Disease</td>
</tr>
</tbody>
</table>
## iii. Analysis Of The Chatgpt Dialogues

In this section, we analyze the responses provided by ChatGPT as sample dialogues between the patient and the ENT-doctor. We performed our analysis in four parts. Each part focuses on one or more responses to prompts (questions) to ChatGPT.

### A. THE FIRST PART (PROVIDE A DIALOGUE BETWEEN A PATIENT AND AN ENT DOCTOR ABOUT A COMMON ENT DISEASE)

The initial dialogue starts with the complaint of the patient and the doctor’s first response. A further explanation of the complaint exists however, it is not a detailed history taking. It lacks past medical history, family history, occupation, medications, age, and gender. Some of the first parts of the dialogues include family history. The doctor’s second response includes a possible diagnosis and a brief plan of the following steps. The doctor suggests having the physical exam next week. However, this is not in conformity with real life as an exam is to follow the history-taking process without delay. The doctor also recommends having some tests performed to confirm the diagnosis. The first part of the dialogue ends with the sentence, “Great. And remember, if you have any questions or concerns, don’t hesitate to reach out to me. I’m here to help you get better.”

### B. THE SECOND PART (HOW CAN THIS DIALOGUE CONTINUE?)
The patient starts by asking, “So if the tests show that I have a [disease name], what can be done about it?” or “Okay, I’ll make sure to follow up with you in a week. Is there anything else I should be doing in the meantime to help with my symptoms?”

The above dialogue is given even for infectious diseases. However, in real life, infectious diseases are diagnosed mostly by clinical findings. No further testings are required. As the dialogue continues, the doctor lists the treatment options for throat, sinus, and ear infections or options to relieve the symptoms. The options provided are generalized information, but they are mostly accurate and in conformity with real-life situations.

C. THE THIRD PART (HOW CAN THIS DIALOGUE CONTINUE FURTHER?)

We wanted to keep the dialogue continuing to see if there is a pattern or similarity in ChatGPT’s responses by asking the prompt, “How can this dialogue continue further?”. The dialogue mostly continued with the patient asking: “Is there anything else I should know about throat-sinus-ear infections?” or “One more thing I wanted to ask about is whether there are any over-the-counter medications that I can take to help with my symptoms. I’ve been using decongestants and pain relievers, but they don’t seem to be providing much relief.” or “One more thing I wanted to ask about is whether there are any lifestyle changes or habits I should be aware of that could help prevent these types of issues in the future.” As a response to these patient questions, the doctor gives information about prevention or other symptoms or pre-existing conditions or chronic ... infections, follow-up care or lifestyle changes, habits, or self-care.

D. THE FOURTH PART (IF TESTS DO NOT CONFIRM THE DIAGNOSIS, HOW WOULD THE DIALOGUE CONTINUE?) AND HOW CAN THIS DIALOGUE CONTINUE FURTHER?)

The doctor explains that the tests are not confirming the diagnosis and makes new suggestions such as diary symptoms and food. Then the doctor provides general information about hygiene. In some of the dialogues, the doctor makes promises by saying, “We will work together to find the cause of your discomfort and develop a treatment plan that is right for you.” In one of the dialogues about nasal polyps, the doctor suggests further testing, such as a CT scan, even though the diagnosis of nasal polyps was not confirmed.

IV. Feedback From Ent Doctors

We asked the following questions to three experienced ENT doctors after sharing the dialogues for throat infection, ear infection, and esophagitis. We did not let the doctors know that the dialogues were generated by ChatGPT before getting their answers.

- Do you find the dialogs accurate? (Yes/No) If not, why?
Do you find the outline of the dialogue appropriate? (Yes/No) If not, why?
Do you find the information given in the dialogue medically reliable? (Yes/No) If not, explain.
These dialogues can help medical students understand how to communicate with a patient better. (Likert Scale: Strongly agree, agree, neutral, disagree, strongly disagree)

All the responders agreed that the dialogues were accurate and that the outlines were appropriate. They also found the information given in the dialogues was medically reliable. They also agreed that the dialogues would help students understand patient-doctor communication better.

V. Discussion & Results
When evaluating the patient-doctor dialogues provided by ChatGPT, we believe there are mainly two important points that need to be considered; the outline of the dialogues and the medical reliability of the information given in the dialogues.

In real life, after a patient is examined, a possible diagnosis would be reached. In cases such as a middle ear infection or a throat infection, the findings would be obvious and typical. In some cases, further testings would be needed. However, by the end of a physical examination, it is expected to see a doctor making explanations about the certain or presumptive diagnosis to the patient. However, ChatGPT always asked for a test to confirm the diagnosis.

The next step would be, depending on the patient's clinical condition, to proceed with treatment without any delay. Such as if the patient is febrile or suffering from pain, that patient may need to receive some medical treatment. If the patient's clinical condition is not as severe, and the patient is willing to have some further testing done as suggested by the doctor, then these testings would be done. They would include some lab tests, radiological examinations, and so on. However, ChatGPT always suggested scheduling a follow-up visit and delayed the treatment.

Moreover, the dialogues seemed very generic and did not show many variations even though we had asked how the dialogue would continue a few times.

VI. Conclusion
In conclusion, the patient-doctor dialogues provided by ChatGPT can be a good starting point for teaching medical students how to communicate with patients. While the information given by the dialogues sounds medically reliable, the outline needs to be reviewed carefully to make sure that the doctor's responses are in conformity with real-life situations.

Declarations

ACKNOWLEDGEMENTS
We thank the ENT doctors who reviewed the patient-doctor dialogues and then responded to our questions.

**Author Contribution**

All authors contributed to the study conception and data analysis. Study design, material preparation, and data collection were performed by the first author. The first draft of the manuscript was written by the first author, and all authors edited and contributed to the preparation of the manuscript. All authors read and approved the final manuscript.

**Funding:** Open access funding provided by University of California Riverside.

**Data Availability:** Not applicable.

**Code Availability:** Not applicable.

**Ethics Approval** Not applicable.

**Competing Interests** The authors declare no competing interests

**References**


**Supplementary Files**

This is a list of supplementary files associated with this preprint. Click to download.

- [SpringerPatientENTDoctorInterview.xlsx](SpringerPatientENTDoctorInterview.xlsx)