

The Utility and Limitations of Artificial Intelligence-Powered Chatbots in Healthcare

Review began 10/15/2024

Review ended 11/01/2024

Published 11/06/2024

© Copyright 2024

Hayat et al. This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY 4.0., which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

DOI: 10.7759/cureus.73127

Jafar Hayat ¹, Mohammad Lari ², Mohammad AlHerz ³, Ali Lari ²

¹. Department of Surgery, Jaber Al-Ahmad Al-Sabah Hospital, Kuwait City, KWT ². Department of Orthopedic Surgery, Al-Razi National Orthopedic Hospital, Kuwait City, KWT ³. Department of Anatomy, Trinity College Dublin, Dublin, IRL

Corresponding author: Jafar Hayat, jafarhayat@hotmail.com

Abstract

At the intersection of artificial intelligence (AI) and healthcare, it is essential that clinicians grasp the ability of chatbots. AI-powered chatbots such as ChatGPT are being explored for their potential benefits by both individuals and institutions. The utility of ChatGPT (OpenAI) in various scenarios was explored through a series of recorded prompts and responses. In the clinical aspects, the chatbot facilitated tasks such as triage, patient consultation, diagnosis, and administrative responsibilities. Their capacity to translate and simplify intricate medical topics was also evaluated. For research purposes, the chatbots' abilities to suggest ideas, prepare protocols, assist in manuscript writing, guide statistical analyses, and recommend suitable journals were assessed. In the educational domain, chatbots were tested for simplifying complex subjects, reviewing procedural steps, generating clinical scenarios, and formulating multiple-choice questions. A comprehensive literature review was also conducted across Medline, Embase, and Web of Science. Chatbots, when optimally employed, can serve as invaluable resources in healthcare, spanning clinical, research, and educational domains. Their potential lies in enhancing efficiency, guiding decision-making, and facilitating patient care and education. However, their application requires a nuanced understanding and caution regarding their limitations.

Categories: Other, Healthcare Technology, Health Policy

Keywords: artificial intelligence, chatbot, chatgpt, healthcare, language-models

Introduction

At the intersection of artificial intelligence (AI) and healthcare, it is vital for providers to grasp the currently transformative ability of chatbots. These tools can generate context-specific solutions across a broad spectrum of scenarios, thereby enhancing daily workflow and efficiency [1]. This article offers a look into the versatile applications of chatbots across various clinical, academic, and administrative duties and presents a perspective on how these AI tools, with proper prompting, can evolve into invaluable resources in healthcare.

Materials And Methods

This methodology aimed to draw a comprehensive view of current knowledge and practices around AI chatbots in healthcare. The prompts and questions were posed to the free 3.5 version of the AI chatbot, ChatGPT (OpenAI). The first responses obtained from the chatbot were recorded and discussed below. The primary outcome measure of this study was to assess and discuss the role of ChatGPT in several clinical scenarios highlighted in the figures below in the results section, and the secondary outcome measure involved the discussion of the role of chatbots in various aspects of healthcare in general, stratified by several subheadings. Findings were discussed descriptively and categorized into appropriate subheadings. No comparator or control was used for this observational study, and interventions provided by ChatGPT were compared to present standards of care found in the available literature.

For the literature review, a comprehensive search in relevant databases including Medline, Embase, and Web of Science was performed using keywords that included; "Chatbots", "Language models", "Artificial intelligence", and "ChatGPT". These were coupled with terms that investigate their medical application; "Research", "Clinical", "Education", "Triage", "Examination", "Administration", "Translation", "Documentation" and "Healthcare". An initial investigation into the applications of chatbots was conducted to identify common themes and trends in the recent application of language models, such as the GPT series by OpenAI, BERT by Google, and Claude. Thus, the selection criteria involved articles that directly addressed the use of AI-powered chatbots in healthcare, whether in patient care, administrative efficiency via patient scheduling, streamlined data collection or interpretation, billing, resource allocation, quality improvement, or educational support for medical professionals.

Results

AI-powered chatbots demonstrated significant utility in clinical practice and research. In clinical settings,

How to cite this article

Hayat J, Lari M, AlHerz M, et al. (November 06, 2024) The Utility and Limitations of Artificial Intelligence-Powered Chatbots in Healthcare . Cureus 16(11): e73127. DOI 10.7759/cureus.73127

chatbots assisted in triage and patient consultations by providing structured frameworks for history-taking (Figure 1) and suggesting differential diagnoses based on symptoms.

I am the emergency physician on-call covering the triage; I have a 50-year-old female patient that is hypertensive presenting with 3 days history of progressive shortness of breath. Can you briefly list pertinent points in the history and examination along with an appropriate differential diagnosis. Be concise and use bullet-points.

Absolutely, here are key points to cover during your history, examination, and the differential diagnosis.

History:

- Characterize the shortness of breath: Onset, duration, progression, any associated symptoms such as chest pain, cough, wheezing, or fever.
- Cardiac history: History of hypertension, other cardiovascular diseases, dyslipidemia, or diabetes mellitus.
- Respiratory history: History of asthma, COPD, pulmonary fibrosis, or other lung diseases.
- Social history: Smoking status, occupational exposure, recent travel.
- Family history: cardiovascular diseases, asthma, or other respiratory diseases.
- Review of systems: Recent weight gain, leg swelling (heart failure), palpitations (arrhythmia), anxiety or panic attacks.

Physical Examination:

- Vital signs: Blood pressure, heart rate, respiratory rate, oxygen saturation, temperature.
- Respiratory: Inspect for use of accessory muscles, cyanosis, audible wheezes, percuss and auscultate the lungs.
- Cardiovascular: Inspect and palpate for jugular venous distension, palpate the pulse, auscultate the heart (murmurs, extra heart sounds).
- Lower extremities: Look for edema which might suggest heart failure.
- Neurological: Check for any signs of stroke.
- Abdomen: Hepatomegaly and hepatic enlargement (congestive heart failure).

Differential Diagnosis:

- Acute decompensated heart failure: This may present with hypertension, especially in a patient with known history of hypertension.
- Acute coronary syndrome: This might present with dyspnea, especially in a hypertensive patient.
- Pulmonary embolism: Particularly if there is a history of recent immobilization, surgery, or hypercoagulable state.
- Pneumonia: Especially if there are signs of infection such as fever or productive cough.
- Chronic obstructive pulmonary disease (COPD) or Asthma exacerbation: If the patient has a history of either.
- Anxiety or panic disorder: May present with shortness of breath, but usually in absence of other significant physical findings.

FIGURE 1: A prompt and a response from ChatGPT suggesting pertinent points in the history and examination.

They recommended relevant investigations (Figure 2) and facilitated the patient consent process by generating clear, jargon-free forms, helping patients understand procedures, risks, and benefits more effectively (Figure 3).

Can you list the important tests, laboratory tests, and radiologic investigations?

Sure, based on your patient's presentation, the following tests and investigations could be considered:

Bedside Tests:

- **ECG:** To assess for ischemia, infarction, arrhythmias, or signs of pulmonary hypertension.
- **Pulse Oximetry:** To check the oxygen saturation level.
- **Spirometry (if available and patient is able):** Can aid in diagnosing obstructive diseases like asthma or COPD.

Laboratory Tests:

- **Complete Blood Count (CBC):** To check for signs of infection, anemia, or polycythemia.
- **Cardiac Enzymes (Troponin I or T):** To rule out myocardial infarction.
- **B-type Natriuretic Peptide (BNP):** This can help differentiate cardiac vs. non-cardiac causes of dyspnea.
- **D-Dimer:** Can be used to rule out pulmonary embolism in low-risk patients.
- **Arterial Blood Gas (ABG):** To assess for hypoxia or hypercapnia and acid-base status.
- **Basic Metabolic Panel:** To assess electrolytes, kidney function, and glucose level.
- **Liver Function Tests (LFTs):** If suspicion of right-sided heart failure (for assessing hepatic congestion).
- **Thyroid Function Tests:** If suspicion of hyperthyroidism.

Radiological Investigations:

- **Chest X-Ray:** To identify signs of heart failure, pneumonia, pneumothorax, or other lung diseases.
- **CT Pulmonary Angiogram:** If high suspicion for pulmonary embolism and patient is stable.
- **Echocardiogram:** To assess cardiac function, wall motion, and valvular disease. Especially useful if suspecting heart failure.
- **V/Q scan:** Alternative to CT pulmonary angiogram in certain cases for diagnosis of pulmonary embolism.

Remember, the choice of investigations should be guided by clinical judgement, and the patient's history and examination findings. It's also important to consider the patient's preferences and overall health status before ordering investigations.

FIGURE 2: A follow-up prompt inquiring into the relevant investigations for the same patient.

I am the surgical trainee on-call, a patient of mine has been diagnosed with appendicitis and my consultant has asked me to take consent for a laparoscopic appendectomy. Can you formulate a consent form that includes all possible complications?

Patient Consent

1. **Understanding the procedure:** I understand that a laparoscopic appendectomy involves removing my appendix using small incisions and special surgical instruments. The surgeon will inflate my abdomen with gas and insert a laparoscope to visualize the surgical area.
2. **Benefits:** The main benefit of this procedure is to relieve my symptoms and prevent potential life-threatening complications of appendicitis. Laparoscopic surgery may lead to a faster recovery, less post-operative pain, and a shorter hospital stay compared to open surgery.
3. **Risks and Complications:** I understand that every surgical procedure has potential risks and complications. These can include but are not limited to:
 - General anesthesia complications (<1%).
 - Pain, bruising, and swelling at the incision sites (common).
 - Infection (2-3%).
 - Bleeding, which may require a blood transfusion or additional surgery (<1%).
 - Damage to surrounding structures including bowel, bladder, blood vessels, and nerves (<1%).
 - Development of a hernia at the incision site (1-5%).
 - Conversion to an open procedure if the laparoscopic approach is not safe or feasible (2-4%).
 - Formation of blood clots (deep vein thrombosis) that can migrate to the lungs (pulmonary embolism) (<1%).
 - Post-operative ileus or temporary pause in bowel function (common, but usually resolves within a few days).
 - Risks related to the appendicitis itself, such as abscess or peritonitis, if the appendix has ruptured (10-20% if appendix has ruptured).
 - Complications related to potential undiagnosed conditions, like appendiceal tumors (<1%).
4. **Alternatives:** I have been informed about alternative options, including not having the operation, antibiotic treatment, or open appendectomy, and I understand their potential risks and benefits.
5. **Consent:** I voluntarily give my consent to undergo a laparoscopic appendectomy. I understand the procedure and its risks, and my questions have been satisfactorily answered.

FIGURE 3: A prompt asking for the details regarding patient consent for a laparoscopic appendectomy procedure.

Chatbots further simplified medical language for diverse populations, aiding patient comprehension of diagnoses and treatments, especially for individuals of different ages and education levels (Figure 4). Additionally, chatbots reduced administrative burdens by aiding with documentation, dictation, and ensuring accurate record-keeping.

I am a practicing doctor, I have to explain to a 5th grader their recent diagnosis of Crohn's disease. Can you explain this disease in a way that they can understand it, using simpler lay terms.

Absolutely! Let's imagine that your body is like a big, long tunnel where the food you eat goes on a journey. This tunnel is called your digestive system. It starts at your mouth, goes down through your stomach and intestines, and ends at the place where you poop.

Now, in some people, parts of this tunnel can sometimes get angry and swollen, kind of like how your skin turns red and puffy if you get a bug bite. This is what's happening inside your body when you have something called Crohn's disease.

Crohn's disease usually gets upset in the last part of your tunnel, the part right before where you poop, but it can happen anywhere along the way. When it's upset, you might feel pain in your tummy, you might not feel like eating, and you might need to use the bathroom a lot. Sometimes, you might even feel very tired.

The tricky thing about Crohn's is that it likes to play hide and seek. It can make you feel yucky for a while and then disappear, making you feel all better. But then, it can come back again. Doctors call these times when you feel sick "flare-ups", and times when you feel good "remission".

The good news is, we have lots of tools to help manage Crohn's disease. You'll have a team of doctors and nurses helping you, and they might give you some special medicine to keep the tunnel calm and less upset. You may also need to eat certain foods that are easier for your tunnel to handle.

And remember, even though Crohn's can be a big nuisance, it doesn't stop you from doing most of the things you love. With the right care and support, you can still go to school, play sports, and spend time with friends. You're just as special and capable as everyone else - you just have a tunnel that needs a little extra care and attention.

FIGURE 4: A prompt generating a response that uses metaphors and lay terms to easily explain Crohn's disease to a child.

In research, chatbots supported idea generation by identifying potential research areas and guiding clinicians in brainstorming systematic reviews on trending topics. They assisted in protocol preparation, generating comprehensive research protocols using predefined templates, and improved manuscript writing by refining structure and correcting language errors. Chatbots also guided researchers on statistical analysis and power sampling, though caution is necessary due to potential inaccuracies (Figure 5). Moreover, chatbots provided journal recommendations by matching the manuscript's content with relevant journals (Figure 6).

Can you perform a rough power sample analysis for my study:
An RCT comparing postoperative pain after procedure A and B.
At a 95% confidence level, 0.8 effect size/power at a 0.05 significance level.

To calculate the sample size for a randomized controlled trial (RCT), the effect size, power, and significance level are crucial parameters. You mentioned that the effect size (Cohen's d) is 0.8 (large effect), the power ($1 - \beta$) is 0.80 (80% chance of detecting an effect if there is one), and the significance level (α) is 0.05 (5% chance of falsely rejecting the null hypothesis).

Assuming equal sample sizes in each group, a two-sided test, and using the parameters you provided, the total sample size required for this study is approximately 52 subjects, with 26 in each group (Procedure A and Procedure B).

This estimate is a general guideline. The actual sample size you need might be larger to account for potential dropouts or other sources of variability.

For a precise calculation and to adjust for other factors such as loss to follow-up, it's recommended to consult with a statistician. Also, different statistical software packages might provide slightly different estimates due to variations in the algorithms they use to calculate sample size.

FIGURE 5: A brief power analysis performed by the chatbot.

The chatbot often guides the user if information is missing.

I have completed a randomized controlled trial comparing the outcomes of antibiotic coated nails versus external circular frames for the treatment of infected non-unions of the tibia.

Can you recommend 5 appropriate journals to submit to? Two of them should be open access.

Absolutely, based on the specifics of your study, here are five journals that might be suitable for your submission:

1. **Journal of Bone and Joint Surgery (JBJS)**: A highly respected, general orthopedic journal that frequently publishes clinical trials. Its broad readership could give your study a wide impact.
2. **Clinical Orthopaedics and Related Research (CORR)**: This journal is a premiere location for high-quality orthopedic research. It's well known for publishing in-depth clinical studies and trials.
3. **Bone & Joint Journal (BJJ)**: This is another highly respected journal in the field of orthopedic surgery and is recognized for its high standard of research publication.

Open Access Journals:

1. **Journal of Orthopaedic Surgery and Research (JOSR)**: An open-access, peer-reviewed journal that is known for publishing high-quality clinical research in the field of orthopedics.
2. **BMC Musculoskeletal Disorders**: A reputable open-access journal that covers all aspects of the prevention, diagnosis, and management of musculoskeletal and associated disorders.

Remember, each journal has its own submission guidelines and criteria that you will need to thoroughly review before submitting your work. Additionally, be aware that open-access journals typically require an article processing charge (APC).

FIGURE 6: Suggestions for journals based on the brief information given to the chatbot.

In education, chatbots efficiently summarized medications, including their mechanisms, doses, and potential interactions (Figure 7). They also reviewed procedures, offering detailed step-by-step explanations to help both beginners and experienced clinicians. Chatbots further generated clinical scenarios and multiple-choice questions (MCQs), enhancing interactive learning and self-assessment.

Can you explain to me why a carotid endarterectomy is contraindicated if there is 100% occlusion as opposed to 99%. This is confusing to me

Carotid endarterectomy (CEA) is a surgical procedure that's performed to remove atherosclerotic plaque from the carotid artery. It's commonly used to reduce the risk of stroke in patients with significant carotid artery stenosis.

However, when a carotid artery is 100% occluded (completely blocked), it means the blood flow through this artery has already been fully cut off. At this point, the brain region that the occluded carotid artery was supplying has likely already found alternative sources of blood, a process known as collateral circulation.

Here are some reasons why CEA is generally not performed in these cases:

1. **Risk vs. Benefit:** Since the blood flow is already completely blocked and alternative sources have likely been established, the potential benefit of re-opening the occluded carotid artery is generally small and may not outweigh the risk of the procedure.
2. **Surgical difficulty and risk:** An artery that is fully occluded is typically filled with a thrombus (blood clot) or highly calcified plaque, both of which can make the surgery technically challenging and increase the risk of complications, including stroke and damage to the artery itself.
3. **Risk of Embolization:** There is a risk during surgery that fragments of the clot or plaque might break off and travel to the brain, potentially causing a stroke.
4. **Lack of reversibility of brain damage:** If the occlusion has been present for a long time, it's possible that irreversible brain damage has occurred in the area of the brain that was supplied by the artery. In this scenario, re-opening the artery won't reverse this damage.

In contrast, in patients with less than 100% occlusion, such as 70-99% occlusion, CEA can potentially prevent a stroke by removing the plaque before total occlusion and subsequent stroke occur. The degree of stenosis at which the benefits of CEA outweigh the risks can depend on several factors, and clinical guidelines generally recommend considering CEA for symptomatic patients with stenosis of 70-99%.

Please note that these are general considerations and the specifics can vary between patients. Each patient's case should be individually evaluated by a team of healthcare professionals. It's also important to note that there are ongoing advancements in endovascular therapies, and treatment approaches may evolve.

FIGURE 7: A question and answer displaying the response of the chatbot explaining the concept of why carotid endarterectomies are contraindicated beyond a certain degree of occlusion.

While chatbots offer valuable support, they can occasionally generate inaccurate information, necessitating human oversight for critical decisions.

Discussion

What are chatbots?

Chatbots are programs that simulate human-like conversations. They utilize AI technology, natural language processing (NLP), and machine learning (ML) to assess the user input and provide an appropriate output (response) [1, 2]. When a chatbot is AI-powered, it is capable of handling complex conversations and can learn from different user interactions. This enables them to make sense of various types of unstructured input. This is in contrast to the earlier rule-based systems, which were only able to respond if a certain structured input was used. Voice-based chatbots represent AI-powered chatbots with the feature of voice-based interactions, allowing for a more natural conversation without the need for typing.

How do they work?

While chatbots have certain limitations (limited diagnostic capabilities, a lack of empathy and human touch, an inability to conduct and appraise physical examinations, and technical limitations), users can achieve optimal responses by understanding the algorithmic flow of chatbot conversations. The user initiates the interaction by providing input, referred to as a prompt. This prompt serves as the main query and establishes the context for the chatbot's output. It is important to note that chatbot conversations are iterative, involving multiple interactions. This means that building context through multiple prompts can lead to the desired response. This has given attention to what has been termed "Prompt engineering," which refers to refining instructions or prompts that effectively obtain various responses from chatbots [3]. Furthermore, chatbots can be fine-tuned to generate responses based on specific datasets and data domains, such as incorporating medical data into their database. By extracting responses from specialized data, chatbots ensure that their responses align with current medical information.

Clinical utility

In the clinical setting, chatbots serve as invaluable tools, offering clinicians a plethora of advantages. These virtual assistants, akin to advanced search engines, seamlessly retrieve medical information, propose treatment algorithms, aid in patient counseling, and alleviate the burdens of administrative tasks. Particularly for junior doctors or clinicians in training, chatbots enable the creation of comprehensive tables to succinctly outline the pros and cons of various investigations, guiding the selection of the optimal next step in patient workups and presenting diverse treatment algorithms. This article, albeit briefly, will explore some of the rudimentary applications of chatbots, showcasing their transformative potential in healthcare.

Triage and Initial Patient Consultation, an On-call Assistant

At an individual level, clinicians can utilize chatbots to analyze patient symptoms and facilitate the process of history taking. These chatbots intelligently examine the symptoms and offer clinicians a structured framework for conducting a thorough patient history and examination by using natural language processing, contextual analysis and symptom triaging, pattern recognition and matching with medical databases, and via the integration of clinical guidelines and best practices [4]. By referring to a list of potential differential diagnoses, the chatbot suggests pertinent questions to ensure important details are not overlooked [5, 6] (Figure 1). This interactive process allows clinicians to have ongoing conversations with the chatbot, incorporating new information and prompting the chatbot to provide more detailed responses. This collaborative approach combines the clinician's expertise with the analytical capabilities of the chatbot. However, there are instances where its application in emergency triage settings may be limited at this stage [7, 8].

Aiding Diagnoses and Providing General Work-up

Healthcare professionals can employ chatbots to assist in the diagnostic process and subsequent patient work-up. Through dynamic interaction, clinicians can utilize chatbots' access to timely information to arrive at a tailored differential diagnosis. Additionally, clinicians can provide updated patient data, prompting the chatbot to suggest appropriate diagnoses and recommend relevant investigations [9]. For example, if a patient presents with respiratory symptoms, the chatbot can offer a list of pertinent questions and potential differential diagnoses [10]. Upon request, the chatbot can propose investigations to help narrow down the diagnosis and rule out more severe conditions (Figure 2). This dynamic process aids clinicians in structuring their approach and ensures comprehensive assessments during the initial patient consultation.

An Outline for Patient Consent

Obtaining patient consent is a crucial aspect of a clinician's daily work, particularly for junior healthcare professionals. In this regard, chatbots can play a supportive role by assisting in the formulation of consent forms that utilize clear language and avoid medical jargon (Figure 3). These chatbots have the capability to search for relevant information related to a specific procedure, including its risks, benefits, and alternatives, and generate concise paragraphs that are easily understandable by patients [11]. This functionality offers two significant benefits. Firstly, it aids clinicians in effectively communicating the information using plain language, ensuring patients comprehend the important aspects of the procedure. Secondly, patients themselves can engage with the chatbot in an interactive dialogue, enabling them to actively participate in the consent process [11, 12]. Moreover, chatbots can easily create tailored consent forms for different procedures, saving time and effort for both clinicians and patients.

Administrative Burden

The day-to-day work of healthcare professionals can be demanding, particularly due to administrative tasks and clinical documentation [13]. Chatbots can act as virtual assistants, assisting in dictation and reducing the time and effort required for manual data entry [14]. They also offer the benefit of grammar and language error checking, ensuring accurate and well-structured documentation [15]. Chatbots have recently been assessed for their ability to write discharge summaries and operative notes, yielding encouraging results

[16]. Additionally, chatbots can be prompted with voice commands to obtain information such as drug dosage or clarify confusing topics.

Explaining and Translating

Chatbots are capable of using simpler language to accommodate for different ages and levels of education [17]. This is useful when trying to explain diagnoses, treatment, medication and prognosis in simpler terms (Figure 4) [18]. Further, chatbots are efficient translators and are capable of translating different text to many different languages in a clear and concise manner [19].

Research utility

The utility of AI-powered language models has garnered significant attention in recent times. It is clear that chatbots will have a significant role to play in future research, especially considering the rapid advancements in language models. However, there exist numerous legitimate concerns among researchers regarding the reliability and potential misuse of these language models in research [20, 21]. While there are various ways chatbots can assist in research on a larger scale, this article will primarily focus on discussing a selection of methods employed at the individual level.

Suggesting Ideas

Chatbots can assist clinicians in identifying areas of medical interest. By engaging with chatbots, clinicians can brainstorm ideas and discover potential research areas that require further investigation, such as conducting systematic reviews on current trending topics [22]. However, it's worth noting that the availability of the latest online medical databases may limit the scope of this feature, potentially restricting it from serving as a source of inspiration.

Preparing Protocols

Chatbots can be valuable in aiding the preparation of research protocols. By utilizing predefined templates, clinicians can input their research ideas and methodology, allowing the chatbot to generate a comprehensive protocol. This functionality can be particularly helpful in ensuring completeness and saving time that would otherwise be spent on manual data entry [23].

Research Writing and Language

Chatbots can serve as helpful guides for structuring and formatting manuscripts. They can assist in editing the overall layout of the document while identifying and correcting linguistic and grammatical errors [20, 24]. Their efficiency in language processing enables them to provide prompt feedback on language-related issues. Chatbots can also be asked to formulate language in different writing styles. However, their ability to accurately write entire manuscripts while maintaining context appears to fall short [25].

Statistical Analysis and Power Sampling

Chatbots can provide guidance on various statistical analysis techniques, presenting explanations in simpler terms for better understanding. They can assist researchers in selecting appropriate analytic tests based on the research design and offer information on general statistical queries [26]. Additionally, chatbots have the capability to perform power analyses when provided with the necessary information, aiding researchers in determining the appropriate sample size for their studies (Figure 5). However, there have been reports of inaccuracies in generating power samples [27]. Thus, a person with statistical experience should be involved in overseeing analytical suggestions by the chatbot.

Journal Recommendation

Chatbots can assist researchers in finding suitable journals for their manuscripts by analyzing the scopes of various journals and matching them with the title and premise of the article (Figure 6). By considering the content and focus of the research, chatbots can provide recommendations on journals that align well with the manuscript. Moreover, they can offer guidance on effective strategies for searching and identifying journals that are a good fit for the research topic and methodology.

Studying and education utility

Chatbots can aid in studying and reviewing information, particularly for medical students and clinicians in training. Often, users rely on standard search engines to gather information, but this can be overwhelming as they need to search through multiple sources to find concise summaries [28]. In contrast, chatbots have the ability to provide personalized responses that are brief and easily comprehensible, making the learning process more efficient and effective for users.

Simplifying complex topics can save time and effort while also potentially aiding in information retention [29]. Chatbots play a crucial role in providing quick summaries of various medications, including their doses, mechanisms of action, and safety profiles, even when considering potential interactions with other drugs [30]. This makes chatbots an excellent tool for extracting concise information rapidly (Figure 7).

Additionally, chatbots can assist in reviewing the steps involved in procedures and surgeries. They can provide detailed explanations of each step and the underlying reasons behind them. This functionality proves particularly useful for beginners, as it helps them understand basic concepts and review the procedure [29, 31]. Similarly, experienced clinicians can also benefit from chatbots by using them to review the steps involved in procedures they are already familiar with.

Chatbots possess the ability to formulate a wide range of clinical questions and patient scenarios, making them a valuable resource for both clinicians and students. By generating realistic scenarios across various topics, chatbots simulate real-world encounters and offer users an opportunity to apply their knowledge in practical settings [32]. These scenarios can be accompanied by pertinent questions encouraging users to think critically and provide accurate responses.

Moreover, chatbots can generate multiple-choice questions (MCQs) at varying levels of difficulty. This feature enables users to assess their understanding and reinforce key concepts through self-testing [33]. By offering a diverse range of clinical questions and patient scenarios, chatbots serve as an interactive learning tool that may enhance clinical reasoning and knowledge retention.

Limitations

While AI-powered chatbots offer convenience and impressive language comprehension, it is important to be aware of their limitations. Chatbots primarily function as advanced search engines, interpreting language at a complex level 1. However, they can occasionally misinterpret medical queries, lacking a complete understanding of the nuances that may result in inaccurate responses. This notable issue, termed as "hallucinations", involves these AI systems producing information that might not be based on the actual input data, potentially leading to misleading or false conclusions [34–36]. Clinicians should exercise caution when relying solely on chatbot-generated answers. Furthermore, it is important to note that chatbots do not have real-time access to the latest medical literature or complex online medical libraries. Therefore, their responses may not always reflect the most up-to-date information available. However, it is likely that this drawback will be addressed in the future. Another limitation of this narrative review is the narrow scope of replies this article covers, leaving room for variance in AI-generated results to other clinical scenarios. The risk of error should always be considered in developing technologies, and it is important to not rely on such technologies exclusively for both clinicians and patients alike.

To maintain a balanced approach, healthcare professionals should use chatbots as supplementary tools rather than relying solely on them. The expertise and clinical judgment of clinicians should not be overshadowed by the capabilities of chatbots. By recognizing these limitations, healthcare professionals can effectively utilize chatbots to enhance their practice without compromising patient care. Similarly, while the utility of chatbots in research is promising, addressing their limitations and concerns is essential. Ethical considerations arise regarding the use of chatbots in research and while there is no consensus on this matter currently, many journals require transparency in chatbot usage [37]. To mitigate the unregulated use of chatbots, chatbot detection programs exist, however their accuracy is uncertain.

Moreover, chatbots rely on preexisting data up to a certain point in time, which means they cannot access the latest evidence without software extensions. Finally, researchers must independently verify the data extracted by chatbots and avoid routine unregulated use that may lead to inaccuracies.

Chatbot Extensions and Other Chatbots

To address the current limitations of chatbots, various internet browser software offers features that enhance their capabilities. Specific software extensions can enable chatbots like ChatGPT to access real-time data and up-to-date information that reflects the latest developments in various fields. Additionally, there are extensions available that can generate concise summaries of articles or videos, condensing them into smaller paragraphs or bullet points. Furthermore, integrating chatbots with web browsers through an extension allows for the integration of the standard search engine with AI-powered language comprehension. These extensions may enhance the functionality and performance of chatbots. However, even in the absence of an extension, newer versions of ChatGPT are likely to add additional features to enhance their online outreach (GPT-4).

The AI chatbot landscape in healthcare features several advanced platforms. ChatGPT (OpenAI Incorporated, Mission District, San Francisco, United States), which is discussed in this study, Google Bard (Alphabet Inc., CA, US), and Microsoft Bing (Microsoft Corporation, WA, US). Google's Bard is tailored to generating contextually rich responses, potentially improving patient engagement and understanding in healthcare interactions. Microsoft's Bing search capabilities offer a broad spectrum of current medical

knowledge, which is useful for up-to-date clinical decision-making [38]. However, direct comparisons are scarce, and results from their performance in exams are similar [39, 40]. In patient-specific healthcare, Ada Health and Babylon Health are often cited [41]. Ada Health's strength lies in its ability to conduct detailed symptom analyses and guide initial patient assessments using a vast medical database [42, 43]. Babylon Health, on the other hand, specializes in virtual healthcare delivery, offering remote consultations and personalized health advice, aligning with the trend toward telemedicine [41]. Nevertheless, the majority of surfacing research has focused on ChatGPT, likely due to the ease of access.

Conclusions

Exploring AI-powered language chatbots like ChatGPT underscores their potential in healthcare and research. These digital tools can assist in clinical work, reduce administrative burden, and enhance educational platforms. Nevertheless, the adoption of such technology necessitates a balanced view, acknowledging their current limitations, especially in context-specific situations requiring a creative approach to problem-solving. Looking forward, the evolution of chatbots could see them becoming more integral to personalized healthcare, with advancements potentially enabling these tools to offer more nuanced, patient-specific insight. Thus, knowledge and proficiency in chatbot usage are likely to offer an advantage to the daily work of the healthcare provider. The rapid pace and trajectory of AI language model development suggest that chatbots might soon evolve beyond their current role as supplementary tools in healthcare. Although there are significant barriers to the widespread implementation of AI technology in healthcare, chatbots have already proven to be useful to healthcare providers in various ways. Clearly following the trajectory of previous digital innovations that enhanced healthcare efficiency, chatbots have a promising future. Despite existing obstacles and some resistance, it's crucial to implement chatbots cautiously to ensure safety and avoid potential pitfalls.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Jafar Hayat, Mohammad Lari, Mohammad AlHerz, Ali Lari

Acquisition, analysis, or interpretation of data: Jafar Hayat, Ali Lari

Drafting of the manuscript: Jafar Hayat, Mohammad Lari, Mohammad AlHerz

Critical review of the manuscript for important intellectual content: Jafar Hayat, Mohammad Lari, Ali Lari

Supervision: Ali Lari

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue.

Animal subjects: All authors have confirmed that this study did not involve animal subjects or tissue.

Conflicts of interest: In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

References

1. Adamopoulou E, Moussiades L: An Overview of Chatbot Technology. Artificial Intelligence Applications and Innovations. 2020, 584:373-83. [10.1007/978-3-030-49186-4_31](https://doi.org/10.1007/978-3-030-49186-4_31)
2. Jung S: Semantic vector learning for natural language understanding. Comput Speech Lang. 2019, 56:130-145. [10.1016/j.csl.2018.12.008](https://doi.org/10.1016/j.csl.2018.12.008)
3. Meskó B: Prompt engineering as an important emerging skill for medical professionals: tutorial . J Med Internet Res. 2023, 25:e50638. [10.2196/50638](https://doi.org/10.2196/50638)
4. Co M, John Yuen TH, Cheung HH: Using clinical history taking chatbot mobile app for clinical bedside teachings: a prospective case control study. Heliyon. 2022, 8:e09751. [10.1016/j.heliyon.2022.e09751](https://doi.org/10.1016/j.heliyon.2022.e09751)
5. Bhattaram S, Shinde VS, Khumujam PP: ChatGPT: The next-gen tool for triaging? . Am J Emerg Med. 2023, 69:215-17. [10.1016/j.ajem.2023.05.027](https://doi.org/10.1016/j.ajem.2023.05.027)
6. Tsui JC, Wong MB, Kim BJ, Maguire AM, Scoles D, VanderBeek BL, Brucker AJ: Appropriateness of ophthalmic symptoms triage by a popular online artificial intelligence chatbot. Eye (Lond). 2023, 37:3692-3. [10.1038/s41433-023-02556-2](https://doi.org/10.1038/s41433-023-02556-2)
7. Knebel D, Priglinger S, Scherer N, Klaas J, Siedlecki J, Schworm B: Assessment of ChatGPT in the

- prehospital management of ophthalmological emergencies: an analysis of 10 fictional case vignettes. *Klin Monbl Augenheilkd.* 2024; 241:675-81. [10.1055/a-2149-0447](https://doi.org/10.1055/a-2149-0447)
8. Fraser H, Crossland D, Bacher I, Ranney M, Madsen T, Hilliard R: Comparison of diagnostic and triage accuracy of ada health and webmd symptom checkers, ChatGPT, and physicians for patients in an emergency department: clinical data analysis study. *JMIR Mhealth Uhealth.* 2023; 11:e49995. [10.2196/49995](https://doi.org/10.2196/49995)
 9. Rao A, Pang M, Kim J, et al.: Assessing the utility of ChatGPT throughout the entire clinical workflow [PREPRINT]. *medRxiv.* 2023; 2:21. [10.1101/2023.02.21.23285886](https://doi.org/10.1101/2023.02.21.23285886)
 10. Hiroswa T, Kawamura R, Harada Y, et al.: Chatgpt-generated differential diagnosis lists for complex case-derived clinical vignettes: diagnostic accuracy evaluation. *JMIR Med Inform.* 2023; 11:e48808. [10.2196/48808](https://doi.org/10.2196/48808)
 11. Decker H, Trang K, Ramirez J, et al.: Large language model-based chatbot vs surgeon-generated informed consent documentation for common procedures. *JAMA Netw Open.* 2023; 6:e2336997. [10.1001/jamanetworkopen.2023.36997](https://doi.org/10.1001/jamanetworkopen.2023.36997)
 12. Savage SK, LoTempio J, Smith ED, et al.: Using a chat-based informed consent tool in large-scale genomic research. *J Am Med Inform Assoc.* 2024; 31:472-78. [10.1093/jamia/ocad181](https://doi.org/10.1093/jamia/ocad181)
 13. Ali SR, Dobbs TD, Hutchings HA, Whitaker IS: Using ChatGPT to write patient clinic letters. *Lancet Digit Health.* 2023; 5:179-181. [10.1016/S2589-7500\(23\)00048-1](https://doi.org/10.1016/S2589-7500(23)00048-1)
 14. Jeblick K, Schachtner B, Dexl J, et al.: ChatGPT makes medicine easy to swallow: an exploratory case study on simplified radiology reports. *Eur Radiol.* 2024; 34:2817-25. [10.1007/s00350-023-10213-1](https://doi.org/10.1007/s00350-023-10213-1)
 15. Robinson A, Aggarwal S Jr: When precision meets penmanship: ChatGPT and surgery documentation. *Cureus.* 2023; 15:e40546. [10.7759/cureus.40546](https://doi.org/10.7759/cureus.40546)
 16. Singh S, Djalilian A, Ali MJ: Chatgpt and ophthalmology: exploring its potential with discharge summaries and operative notes. *Semin Ophthalmol.* 2023; 38:503-7. [10.1080/08820538.2023.2209166](https://doi.org/10.1080/08820538.2023.2209166)
 17. Teixeira da Silva JA: Can ChatGPT rescue or assist with language barriers in healthcare communication?. *Patient Educ Couns.* 2023; 115:107940. [10.1016/j.pec.2023.107940](https://doi.org/10.1016/j.pec.2023.107940)
 18. Zhang Y, Chen R, Nguyen D, et al.: Assessing the ability of an artificial intelligence chatbot to translate dermatopathology reports into patient-friendly language: A cross-sectional study. *J Am Acad Dermatol.* 2024; 90:397-99. [10.1016/j.jaad.2023.09.072](https://doi.org/10.1016/j.jaad.2023.09.072)
 19. Rainey JP, Blackburn BE, McCutcheon CL, Kenyon CM, Campbell KJ, Anderson LA, Gililand JM: A multilingual chatbot can effectively engage arthroplasty patients who have limited english proficiency. *J Arthroplasty.* 2023; 38:S78-83. [10.1016/j.arth.2023.04.014](https://doi.org/10.1016/j.arth.2023.04.014)
 20. Salvagno M, Taccone FS, Gerli AG: Can artificial intelligence help for scientific writing?. *Crit Care.* 2023; 27:75. [10.1186/s13054-023-04380-2](https://doi.org/10.1186/s13054-023-04380-2)
 21. Huang J, Tan M: The role of ChatGPT in scientific communication: writing better scientific review articles. *Am J Cancer Res.* 2023; 13:1148-54.
 22. Gupta R, Park JB, Bisht C, et al.: Expanding cosmetic plastic surgery research with ChatGPT. *Aesthet Surg J.* 2023; 43:930-37. [10.1093/asj/sjad069](https://doi.org/10.1093/asj/sjad069)
 23. Sallam M: ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. *Healthcare (Basel).* 2023; 11:887. [10.3390/healthcare11060887](https://doi.org/10.3390/healthcare11060887)
 24. Hutson M: Could AI help you to write your next paper?. *Nature.* 2022; 611:192-3. [10.1038/d41586-022-03479-w](https://doi.org/10.1038/d41586-022-03479-w)
 25. Májovský M, Černý M, Kasal M, Komarc M, Netuka D: Artificial intelligence can generate fraudulent but authentic-looking scientific medical articles: pandora's box has been opened. *J Med Internet Res.* 2023; 25:e46924. [10.2196/46924](https://doi.org/10.2196/46924)
 26. Macdonald C, Adeloye D, Sheikh A, Rudan I: Can ChatGPT draft a research article? An example of population-level vaccine effectiveness analysis. *J Glob Health.* 2023; 13:01003. [10.7189/jogh.13.01003](https://doi.org/10.7189/jogh.13.01003)
 27. Methnani J, Latiri I, Dergaa I, Chamari K, Ben Saad H: ChatGPT for sample-size calculation in sports medicine and exercise sciences: a cautionary note. *Int J Sports Physiol Perform.* 2023; 18:1219-23. [10.1123/ijsspp.2023-0109](https://doi.org/10.1123/ijsspp.2023-0109)
 28. Kaur A, Singh S, Chandan JS, Robbins T, Patel V: Qualitative exploration of digital chatbot use in medical education: A pilot study. *Digit Health.* 2021; 7:20552076211038151. [10.1177/20552076211038151](https://doi.org/10.1177/20552076211038151)
 29. Lee H: The rise of ChatGPT: Exploring its potential in medical education. *Anat Sci Educ.* 2024; 17:926-31. [10.1002/ase.2270](https://doi.org/10.1002/ase.2270)
 30. Li YS, Lam CS, See C: Using a machine learning architecture to create an ai-powered chatbot for anatomy education. *Med Sci Educ.* 2021; 31:1729-30. [10.1007/s40670-021-01405-9](https://doi.org/10.1007/s40670-021-01405-9)
 31. Kung TH, Cheatham M, Medenilla A, et al.: Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLOS Digit Health.* 2023; 2:e0000198. [10.1371/journal.pdig.0000198](https://doi.org/10.1371/journal.pdig.0000198)
 32. Moldt JA, Festl-Wietek T, Madany Mamlouk A, Nieselt K, Fuhl W, Herrmann-Werner A: Chatbots for future docs: exploring medical students' attitudes and knowledge towards artificial intelligence and medical chatbots. *Med Educ Online.* 2023; 28:2182659. [10.1080/10872981.2023.2182659](https://doi.org/10.1080/10872981.2023.2182659)
 33. Gilson A, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA, Chartash D: How does ChatGPT perform on the united states medical licensing examination (USMLE)? The implications of large language models for medical education and knowledge assessment. *JMIR Med Educ.* 2023; 9:e45312. [10.2196/45312](https://doi.org/10.2196/45312)
 34. Metze K, Morandin-Reis RC, Lorand-Metze I, Florindo JB: Bibliographic research with ChatGPT may be misleading: the problem of hallucination. *J Pediatr Surg.* 2024; 59:158. [10.1016/j.jpedsurg.2023.08.018](https://doi.org/10.1016/j.jpedsurg.2023.08.018)
 35. Giuffrè M, You K, Shung DL: Evaluating ChatGPT in medical contexts: the imperative to guard against hallucinations and partial accuracies. *Clin Gastroenterol Hepatol.* 2024; 22:1145-6. [10.1016/j.cgh.2023.09.035](https://doi.org/10.1016/j.cgh.2023.09.035)
 36. Sharun K, Banu SA, Pawde AM, Kumar R, Akash S, Dhama K, Pal A: ChatGPT and artificial hallucinations in stem cell research: assessing the accuracy of generated references - a preliminary study. *Ann Med Surg (Lond).* 2023; 85:5275-78. [10.1097/MS9.0000000000001228](https://doi.org/10.1097/MS9.0000000000001228)
 37. Meskó B, Topol EJ: The imperative for regulatory oversight of large language models (or generative AI) in healthcare. *NPJ Digit Med.* 2023; 6:120. [10.1038/s41746-023-00873-0](https://doi.org/10.1038/s41746-023-00873-0)
 38. Dhanvijay AK, Pinjar MJ, Dhokane N, Sorte SR, Kumari A, Mondal H: Performance of Large Language Models

- (ChatGPT, Bing Search, and Google Bard) in solving case vignettes in physiology. *Cureus*. 2023, 15:e42972. [10.7759/cureus.42972](https://doi.org/10.7759/cureus.42972)
- 39. Stephens LD, Jacobs JW, Adkins BD, Booth GS: Battle of the (Chat)Bots: comparing Large Language Models to practice guidelines for transfusion-associated graft-versus-host disease prevention. *Transfus Med Rev*. 2023, 37:150753. [10.1016/j.tmr.2023.150753](https://doi.org/10.1016/j.tmr.2023.150753)
 - 40. Thibaut G, Dabbagh A, Liverneaux P: Does Google's Bard Chatbot perform better than ChatGPT on the European hand surgery exam?. *Int Orthop*. 2024, 48:151-58. [10.1007/s00264-023-06034-y](https://doi.org/10.1007/s00264-023-06034-y)
 - 41. Gilbert S, Mehl A, Baluch A, et al.: How accurate are digital symptom assessment apps for suggesting conditions and urgency advice? A clinical vignettes comparison to GPs. *BMJ Open*. 2020, 10:e040269. [10.1136/bmjopen-2020-040269](https://doi.org/10.1136/bmjopen-2020-040269)
 - 42. Faqar-Uz-Zaman SF, Anantharajah L, Baumartz P, et al.: The diagnostic efficacy of an app-based diagnostic health care application in the emergency room: eradar-trial. A prospective, double-blinded, observational study. *Ann Surg*. 2022, 276:935-42. [10.1097/SLA.0000000000005614](https://doi.org/10.1097/SLA.0000000000005614)
 - 43. Gilbert S, Fenech M, Upadhyay S, Wicks P, Novorol C: Quality of condition suggestions and urgency advice provided by the Ada symptom assessment app evaluated with vignettes optimised for Australia. *Aust J Prim Health*. 2021, 27:377-81. [10.1071/PY21032](https://doi.org/10.1071/PY21032)