After a few months, what are the uses of OpenAI’s ChatGPT in medicine? A Scopus-based systematic review

Jeel Moya-Salazar 1*, Carmen R Salazar 2*, Stephany S Delzo 3*, Eliane A Goicochea-Palomino 4*, Victor Rojas-Zumaran 3*

1 Digital Transformation Center, Department of Research, Universidad Norbert Wiener, Lima, PERU
2 Department of Pathology, Hospital Nacional Docente Madre Niño San Bartolomé, Lima, PERU
3 Department of Pathology, Hospital Nacional Docente Madre Niño San Bartolomé, Lima, PERU
4 Department of Pathology, Hospital Nacional Docente Madre Niño San Bartolomé, Lima, PERU
*Corresponding Author: jeel.moya@uwiener.edu.pe

Citation: Moya-Salazar J, Salazar CR, Delzo SS, Goicochea-Palomino EA, Rojas-Zumaran V. After a few months, what are the uses of OpenAI’s ChatGPT in medicine? A Scopus-based systematic review. Electron J Gen Med. 2024;21(2):em577. https://doi.org/10.29333/ejgm/14354

INTRODUCTION

The 4.0 Revolution has led humanity to an unprecedented virtual context. Since the explosion of mobile devices, the massification of the Internet, and the expansion of social networks, humans have embraced the digital world, irreversibly changing human aspects [1]. Such is the expansion of virtual environments and cybernetics that medicine has been altered and potentially benefited by technological advances [2]. Currently, surgery processes [3], clinical decisions [5, 6], and medical treatment [7] are being enhanced with technology based on algorithms that enable technological learning [8]. Recently, ChatGPT (generative pre-trained transformer), a chatbot that uses OpenAI’s GPT-3 language model, has been introduced, showing interesting results in various medical fields [9, 10]. As its potential as an aid in medical decision-making and university education grows [11], concerns and risks regarding its use in medicine arise [12]. Additionally, evidence about its use can be generated from home experiments or studies that do not follow strict scientific protocols, making it necessary to recognize the characteristics of its scientific production in recognized databases such as Medline or Scopus [13, 14]. We aimed to conduct a review of the scientific literature in medicine on the use of ChatGPT in Scopus. This review provides a thematic analysis of publications that have used ChatGPT in any medical field, as well as the worldwide distribution of scientific production.

METHODS

Data Source & Search Strategy

We performed a systematic review adhering to the preferred reporting items for systematic reviews and meta-analyses (PRISMA) 2020 guidelines [15]. The review was conducted on the Scopus website between April 25 and April
27, 2023. We used the Boolean operators and keywords to create the following search equation >>TITLE-ABS-KEY ("ChatGPT" OR "artificial intelligence" OR "Machine Learning" AND "medicine")<<. The search was manually and was conducted by two authors simultaneously (JM-S and CRS). There was a discussion in regard to the search and selection in order to have consensual agreements.

Selection Criteria

The inclusion criteria for selecting articles were studies in clinical and surgical medicine, original articles, clinical trials, correspondence articles, letters, and research protocols that used or discussed ChatGPT. Editorial, systematic reviews and meta-analyses, and the position, reflection and historical papers were excluded. Additionally, other studies that did not include ChatGPT in any of its versions as part of their experiments or analyses were excluded (Figure 1).

Screening Study, Data Extraction, & Analysis

Two authors conducted an independent assessment of the abstracts, excluding studies that did not meet the inclusion criteria. Agreement of the selected studies was evaluated with the Kappa test [16]. The final full-texts were reviewed in order to include the analysis of the revision. The data were recorded in a database using MS-Excel 2015 (Microsoft Corp., Redmond, WA, USA) and data extraction was performed through critical appraisal skills program to capture information on this topic [17]. A narrative analysis of the included studies was employed, determining the principal themes and characteristics following previous reports [18]. The descriptive analysis was performed with SPSS v24.0 (Armonk, NY, USA). The construction of maps was performed using Microsoft Bing Maps (Microsoft Corp., Redmond, WA, USA).

RESULTS

In the initial search, 11,219 articles were found in Scopus, and after the screening process, 10,540 articles were excluded. Two additional studies were not retrieved, and 661 articles were excluded for not meeting the inclusion criteria. Finally, 14 studies underwent analysis [19-32]. The two authors had an optimal concordance (k=0.901) in the global search of articles.

Characteristics of Studies

All the studies were published in 2023 and 35.7% (5/14) were editorials. In addition, we had 14.3% (2/14) documents that were opinion or perspective articles. Only one study [19] was cross-sectional and it performed a pilot study about the ChatGPT application in medical diagnosis, showing a good performance in the election of strategies for the diagnosis, although it was not superior to medical diagnosis. Journal of Medical Systems had the highest number of papers on ChatGPT and medicine, with two publications (14.3%). Additionally, 42.8% of the publications were carried out in collaboration with publishers from the Springer or Elsevier group (Table 1).

Based on the affiliation of the corresponding author or the principal researcher, the USA emerged as the leading contributor, accounting for five (53.7%) publications on ChatGPT in medicine. Following closely were Belgium and France, each with two (14.3%) papers (Figure 2).

Table 1. Bibliometric characteristics of research on ChatGPT in medicine available in Scopus

<table>
<thead>
<tr>
<th>Author</th>
<th>Article type</th>
<th>Journal</th>
<th>Editorial</th>
<th>H</th>
<th>Q*</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hirosawa</td>
<td>Original article</td>
<td>Int J Environ Res Public Health</td>
<td>Frontiers Media S. A.</td>
<td>167</td>
<td>Q2</td>
<td>Switzerland</td>
</tr>
<tr>
<td>Alberts</td>
<td>Editorial</td>
<td>Eur J Nucl Med Mol Imaging</td>
<td>Springer Verlag</td>
<td>177</td>
<td>Q1</td>
<td>Germany</td>
</tr>
<tr>
<td>Buvat &amp; Weber</td>
<td>Notes</td>
<td>J Nucl Med</td>
<td>SNMMI**</td>
<td>232</td>
<td>Q1</td>
<td>USA</td>
</tr>
<tr>
<td>Cascella</td>
<td>Letter</td>
<td>J Med Syst</td>
<td>Springer New York</td>
<td>100</td>
<td>Q2</td>
<td>USA</td>
</tr>
<tr>
<td>Salvagno</td>
<td>Perspective</td>
<td>Critical care</td>
<td>BioMed Central Ltd.</td>
<td>200</td>
<td>Q1</td>
<td>UK</td>
</tr>
<tr>
<td>Elali &amp; Rachid</td>
<td>Opinion letter</td>
<td>Patterns</td>
<td>Cell Press</td>
<td>20</td>
<td>Q1</td>
<td>USA</td>
</tr>
<tr>
<td>Grünebaum</td>
<td>Opinion letter</td>
<td>Am J Obstet Gynecol</td>
<td>Mosby Inc.</td>
<td>250</td>
<td>Q1</td>
<td>USA</td>
</tr>
<tr>
<td>Homolak</td>
<td>Editorial</td>
<td>Croat Med J</td>
<td>Medicinska Naklada</td>
<td>63</td>
<td>Q3</td>
<td>Croatia</td>
</tr>
<tr>
<td>Anderson</td>
<td>Editorial</td>
<td>BMJ Open Sport Exerc Med</td>
<td>BMJ Publishing</td>
<td>36</td>
<td>Q1</td>
<td>UK</td>
</tr>
<tr>
<td>Morreel</td>
<td>Letter to editor</td>
<td>Med Teach</td>
<td>Informa Healthcare</td>
<td>131</td>
<td>Q1</td>
<td>UK</td>
</tr>
<tr>
<td>Nguyen &amp; Costedoat-Chalumeau</td>
<td>Debate paper</td>
<td>Revue Med Interne</td>
<td>Elsevier Masson S. R. L.</td>
<td>38</td>
<td>Q3</td>
<td>France</td>
</tr>
<tr>
<td>Mann</td>
<td>Perspective</td>
<td>JACC Basic Trans Sci</td>
<td>Elsevier Inc.</td>
<td>50</td>
<td>Q1</td>
<td>USA</td>
</tr>
<tr>
<td>DiGiorgio &amp; Ehrenfeld</td>
<td>Editorial</td>
<td>J Med Syst</td>
<td>Springer New York</td>
<td>100</td>
<td>Q2</td>
<td>USA</td>
</tr>
<tr>
<td>Ferres</td>
<td>Editorial</td>
<td>Diagn Interv Imaging</td>
<td>Elsevier Masson S. R. L.</td>
<td>51</td>
<td>Q1</td>
<td>France</td>
</tr>
</tbody>
</table>

Note. * Scimago 2023; ** Society of Nuclear Medicine and Molecular Imaging; H: h-index; & Q: Quartile
Main Findings

The available articles highlight the potential and challenges of ChatGPT and artificial intelligence in various healthcare domains (Table 2).

The first study [19] presents findings comparing the diagnostic performance of ChatGPT-3 and physicians. The authors compared the diagnostic accuracy of ChatGPT-3 with that of physicians, revealing lower performance in differential diagnosis (83.3% vs. 98.3%, p=0.030) and important diagnostic tasks (53.3% vs. 93.3%, p<0.001). The study in Germany [20] emphasizes the ground-breaking potential of large language models (LLMs) in nuclear medicine but also analyse ethical concerns such as data privacy and bias.

The studies in USA [21, 22] highlight that clear communication between healthcare professionals and patients can be enhanced with ChatGPT, overcoming barriers such as lack of time and concise explanations. Therefore, both articles consider ChatGPT as a tool for improving patient education and commitment, handling various tasks, and providing accurate information about health and diseases. However, careful evaluation of the ethical, legal, and technical aspects of implementing artificial intelligence (AI) in healthcare is considered crucial.

Other studies [23, 24, 26, 27] explore the role of AI in scientific writing, highlighting its assistance in tasks such as reference verification but emphasizing the importance of

Table 2. Main results of studies included in analysis of this review on ChatGPT in medicine

<table>
<thead>
<tr>
<th>Author</th>
<th>Article type</th>
<th>Issues found or discussed</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hirosawa</td>
<td>Original article</td>
<td>Correct diagnosis by ChatGPT-3 vs. physicians (83.3% vs. 98.3%, p=0.030) for differential diagnosis and (53.3% vs. 93.3%, p&lt;0.001) for principal diagnostic.</td>
<td>[19]</td>
</tr>
<tr>
<td>Alberts</td>
<td>Editorial</td>
<td>Authors argue that LLMs and ChatGPT have the potential to revolutionize nuclear medicine, but also highlight the ethical concerns surrounding the use of AI in healthcare, such as data privacy and bias.</td>
<td>[20]</td>
</tr>
<tr>
<td>Buvat &amp; Weber</td>
<td>Notes</td>
<td>Two nuclear medicine experts highlight importance of clear communication between healthcare professionals &amp; patients &amp; discuss how ChatGPT could improve patient education, engagement, limitations, &amp; biases.</td>
<td>[21]</td>
</tr>
<tr>
<td>Cascella</td>
<td>Letter</td>
<td>ChatGPT was able to handle various tasks &amp; provide accurate &amp; useful information to patients &amp; healthcare providers. But ethical, legal, &amp; technical aspects when implementing AI in healthcare should be evaluated.</td>
<td>[22]</td>
</tr>
<tr>
<td>Salvagno</td>
<td>Perspective</td>
<td>AI can assist in tasks like reference checking &amp; improving readability of manuscripts. But they also caution that AI should not replace human writing and those ethical concerns, such as bias, must be addressed.</td>
<td>[23]</td>
</tr>
<tr>
<td>Elali &amp; Rachid</td>
<td>Opinion letter</td>
<td>There are risks associated with the ease of access to research paper generation models and the authors emphasize the need for strict ethical standards and guidelines to prevent fraudulent practices.</td>
<td>[24]</td>
</tr>
<tr>
<td>Grünebaum</td>
<td>Opinion letter</td>
<td>ChatGPT can aid in patient communication, clinical decision-making, &amp; education in OB/GYN. They also propose future directions for research like developing specialized ChatGPT models for specific topics.</td>
<td>[25]</td>
</tr>
<tr>
<td>Homolak</td>
<td>Editorial</td>
<td>While these AI models can improve scientific research &amp; productivity, there is also a risk of unethical use like fabrication of research articles &amp; plagiarism. Guidelines &amp; regulations are necessary for its responsible use.</td>
<td>[26]</td>
</tr>
<tr>
<td>Anderson</td>
<td>Editorial</td>
<td>The authors conducted experiments to determine if AI-generated texts can deceive current AI text detectors. Results show that with proper tuning, AI can produce manuscripts that can bypass current detectors, which can potentially lead to the widespread use of AI in scientific writing.</td>
<td>[27]</td>
</tr>
<tr>
<td>Morreel</td>
<td>Letter to editor</td>
<td>Researchers tested ChatGPT’s ability to pass a multiple-choice family medicine exam &amp; found that it was able to achieve a passing score, demonstrating its potential as a tool to support medical professionals.</td>
<td>[28]</td>
</tr>
<tr>
<td>Nguyen &amp; Costedoat-Chalumeau</td>
<td>Debate paper</td>
<td>There is a potential for ChatGPT to analyze complex medical information &amp; provide insights related to hydroxychloroquine. Also, authors caution against potential biases &amp; need for human expertise in decision-making.</td>
<td>[29]</td>
</tr>
<tr>
<td>Mann</td>
<td>Perspective</td>
<td>The authors show the potential of AI in translational medicine, discussing various applications such as drug discovery, patient stratification, and personalized medicine. However, the challenges and ethical considerations need to be addressed to ensure the safe and effective integration.</td>
<td>[30]</td>
</tr>
<tr>
<td>DiGiorgio &amp; Ehrenfeld</td>
<td>Editorial</td>
<td>ChatGPT could provide information, answer questions, and even generate patient reports, allowing physicians to focus on critical tasks. However, the authors also acknowledge the need for caution and further evaluation of the technology’s effectiveness in real-world clinical settings.</td>
<td>[31]</td>
</tr>
<tr>
<td>Ferres</td>
<td>Editorial</td>
<td>ChatGPT could enhance patient care and radiology education but also raises concerns about data privacy and bias. The authors suggest that careful consideration of ethical and legal implications, as well as further research, is needed before the widespread adoption of ChatGPT in these fields.</td>
<td>[32]</td>
</tr>
</tbody>
</table>

Note: R: Reference
preserving human involvement (as human writing should not be completely replaced) and addressing ethical concerns, including bias. However, the ease of access to article generation models poses risks such as fabrication and plagiarism, underscoring the need for ethical standards and strict guidelines. It was conducted experiments to evaluate if AI-text generated can deceive existing AI text detectors [27]. The results indicate that, with the adequate adjustment, AI can produce texts that cannot be detected by the detectors, which could lead to a generalized habit in scientific writing, which creates concern in regard to publication practices and scientific dissemination [26].

The capability of ChatGPT in solving medical exams has also been discussed based on early experiments. OpenAI’s ChatGPT has been able to pass a multiple-choice family medicine exam [28], and the authors highlight its potential use as a supportive tool for medical, non-medical professionals, undergraduate and postgraduate students. However, caution regarding biases and the lack of need of human experience. While AI models can enhance scientific research and productivity, there is a risk of unethical use, and guidelines and regulations are needed for responsible implementation [26-28].

ChatGPT is also promising in analysing complex medical information, particularly regarding hydroxychloroquine [29]. The study in [30] has also emphasized the role of AI in translational medicine, focusing on applications such as drug discovery and personalized medicine. However, ethical considerations and challenges are recognized.

Finally, in terms of patient care [32], radiology education [32], and gynaecology [25], ChatGPT could provide key and organized information, answer questions, and generate patient reports, aiding in clinical decision-making and education, allowing healthcare professionals to focus on critical tasks. However, the authors emphasize the importance of caution and further evaluation of technology effectiveness in real-world clinical settings. Data privacy and bias are also raised as concerns to ensure responsible use.

**DISCUSSION**

This systematic review suggests that ChatGPT holds the potential to improve medical care and education across diverse research fields, clinical domains, and patient care contexts. However, the studies identified highlight the importance of careful evaluation and investigation before the adoption of a generalized use of ChatGPT in clinical surroundings, as ethical and legal mistakes can be made through its indiscriminate and widespread use in medicine.

**Strengths**

This is one of the first systematic revisions in Scopus about the scientific output on OpenAI’s ChatGPT. Some recent systematic reviews have not either included the Scopus database [12, 33-35] nor have described a narrative panorama about the limitations and benefits of ChatGPT in medicine [36]. Thus, this revision based on Scopus, one of the integral databases of scientific literature, show some initial studies on the use of this AI in medicine. On the other hand, in this revision we also frame countries with the most significant scientific output on ChatGPT in medicine (Figure 2) highlighting regions such as Africa and Latin America that are regions that are still without scientific production in Scopus. This panorama is key for the evaluation of scientific production, use, or development of AI, particularly of ChatGPT in certain regions in to reduce scientific gaps and the prioritization of activities by continents.

**Main Findings**

According to our results, this LLM is still in its early stages as a medical tool, but the benefits in patient care and diagnostic support are already apparent. In the diagnostic field, ChatGPT has been evaluated with favourable preliminary results [33] in radiology [37], ophthalmology [38, 39], and rare disease diagnosis [40].

There are also experiences in case reporting in medicine [41, 42] and in managing a range of diseases such as cancer [43-45]. Although the evidence is still limited in Scopus, it is possible that ChatGPT can also have an impact as a clinical decision sup-port tool, medical recordkeeping, management, and translation, as well as assistants for telemedicine and remote patient monitoring.

Another key and growing aspect is the generation of research and papers assisted by ChatGPT. This review has identified some papers [23, 24, 26, 27] discussing this topic and highlighting that with proper guidance, OpenAI’s ChatGPT can be a “co-author” of scientific documents.

Other studies have demonstrated its use as a tool in writing and literature review [34] and have described up to seven roles of ChatGPT in medical education (e.g., generating case scenarios, research assistance) that promise to improve scientific production but also raise risks of plagiarism. Most publications on ChatGPT in Scopus have been published in high-impact journals, indicating the im-portance of this LLM in the future of publication and scientific research.

As there are ethical concerns surrounding its use, it is necessary to create surveillance systems for AI-assisted scientific production. This is to prevent misuse and ensure good practices in scientific production [46], as promising as its benefits are, so are its risks [47].

ChatGPT, an AI-powered chatbot, is rapidly changing the field of scientific production and healthcare, and its results will redirect efforts in public health and medical science for the quality of patient care [48]. Currently, there are other platforms like DeepL Write, HuggingChat, Marmof, Bard, OpenAssistant.io, and Perplexity AI that have also emerged as options to ChatGPT and can be medical assistance systems that also need to be evaluated. Just as ethical and legal implications must be carefully considered, the scientific gaps between continents in AI research must also be highlighted. Based on our results in Scopus, we did not identify papers on ChatGPT in regions with Latin American or African affiliation. More research is needed during widespread adoption of this AI to include regions, where use of this technology could improve medical diagnostic processes [49], promote educational and research opportunity in health sciences [50, 51], enhance telemedicine healthcare processes [44], and improve patient care by reducing misinformation and myths about health and disease [52].

**Limitations**

Firstly, the 14 analysed papers were identified in the Scopus database, but there are other databases (including Gray literature) with additional papers discussing the role of ChatGPT in medicine [12, 33-35]. Papers often take time to be
indexed in Scopus, and some are initially registered in PubMed/Medline and Google Scholar before appearing in Scopus [52].

Hence, the number of papers may vary in the coming months. Another limitation is that we analysed the role of OpenAI’s ChatGPT, but other LLMs may generate different evidence in medical sciences. Additional research is necessary to assess the practical application of ChatGPT and other LLMs in healthcare and research.

CONCLUSIONS

We have demonstrated that while ChatGPT has promising benefits, ethical considerations, guidelines, and additional research are essential for responsible and effective integration into healthcare. Although there are hopes and fears about ChatGPT in medicine, the evidence in Scopus is limited. Most studies are discussions about the role of this chatbot in various clinical fields, offering advantages and opportunities to physicians, but also discussing potential risks in scientific research and medical education.

Author contributions: JM-S: provided study concept & design, data curation & formal analysis, & wrote manuscript; CR: provided study concept & design, formal analysis, & wrote manuscript; SD: provided data acquisition, curation, & analysis; EAG-P: provided data acquisition, curation, & performed data management; & VRZ: provided design, formal analysis & visualization, & wrote manuscript. All authors have agreed with the results and conclusions.

Funding: No funding source is reported for this study.

Ethical statement: The authors stated that the study, being a systematic review, has not been reviewed by the Ethics and Research Committee of the universities.

Declaration of interest: No conflict of interest is declared by authors.

Data sharing statement: Data supporting the findings and conclusions are available upon request from the corresponding author.

REFERENCES


