Review

A SWOT (Strengths, Weaknesses, Opportunities, and Threats) Analysis of ChatGPT in the Medical Literature: Concise Review

Daniel Gödde¹, MD, Dr med; Sophia Nöhl², BSc; Carina Wolf²; Yannick Rupert²; Lukas Rimkus²; Jan Ehlers³, Prof Dr; Frank Breuckmann^{4,5}, MD, Prof Dr; Timur Sellmann^{6,7}, MD, Dr med, MHBA

¹Department of Pathology and Molecularpathology, Helios University Hospital Wuppertal, Witten/Herdecke University, Witten, Germany

²Faculty of Health, Witten/Herdecke University, Witten, Germany

Corresponding Author:

Daniel Gödde, MD, Dr med Department of Pathology and Molecularpathology Helios University Hospital Wuppertal Witten/Herdecke University Alfred-Herrhausen-Straße 50 Witten, 58455 Germany Phone: 49 202 896 2541 Fax: 49 202 896 2739 Email: daniel.goedde@helios-gesundheit.de

Abstract

Background: ChatGPT is a 175-billion-parameter natural language processing model that is already involved in scientific content and publications. Its influence ranges from providing quick access to information on medical topics, assisting in generating medical and scientific articles and papers, performing medical data analyses, and even interpreting complex data sets.

Objective: The future role of ChatGPT remains uncertain and a matter of debate already shortly after its release. This review aimed to analyze the role of ChatGPT in the medical literature during the first 3 months after its release.

Methods: We performed a concise review of literature published in PubMed from December 1, 2022, to March 31, 2023. To find all publications related to ChatGPT or considering ChatGPT, the search term was kept simple ("ChatGPT" in AllFields). All publications available as full text in German or English were included. All accessible publications were evaluated according to specifications by the author team (eg, impact factor, publication modus, article type, publication speed, and type of ChatGPT integration or content). The conclusions of the articles were used for later SWOT (strengths, weaknesses, opportunities, and threats) analysis. All data were analyzed on a descriptive basis.

Results: Of 178 studies in total, 160 met the inclusion criteria and were evaluated. The average impact factor was 4.423 (range 0-96.216), and the average publication speed was 16 (range 0-83) days. Among the articles, there were 77 editorials (48,1%), 43 essays (26.9%), 21 studies (13.1%), 6 reviews (3.8%), 6 case reports (3.8%), 6 news (3.8%), and 1 meta-analysis (0.6%). Of those, 54.4% (n=87) were published as open access, with 5% (n=8) provided on preprint servers. Over 400 quotes with information on strengths, weaknesses, opportunities, and threats were detected. By far, most (n=142, 34.8%) were related to weaknesses. ChatGPT excels in its ability to express ideas clearly and formulate general contexts comprehensibly. It performs so well that even experts in the field have difficulty identifying abstracts generated by ChatGPT. However, the time-limited scope and the need for corrections by experts were mentioned as weaknesses and threats of ChatGPT. Opportunities include assistance in formulating medical issues for nonnative English speakers, as well as the possibility of timely participation in the development of such artificial intelligence tools since it is in its early stages and can therefore still be influenced.

RenderX

³Department of Didactics and Education Research in the Health Sector, Faculty of Health, Witten/Herdecke University, Witten, Germany

⁴Department of Cardiology and Vascular Medicine, West German Heart and Vascular Center Essen, University Duisburg-Essen, Essen, Germany

⁵Department of Cardiology, Pneumology, Neurology and Intensive Care Medicine, Klinik Kitzinger Land, Kitzingen, Germany

⁶Department of Anaesthesiology I, Witten/Herdecke University, Witten, Germany

⁷Department of Anaesthesiology and Intensive Care Medicine, Evangelisches Krankenhaus BETHESDA zu Duisburg, Duisburg, Germany

Conclusions: Artificial intelligence tools such as ChatGPT are already part of the medical publishing landscape. Despite their apparent opportunities, policies and guidelines must be implemented to ensure benefits in education, clinical practice, and research and protect against threats such as scientific misconduct, plagiarism, and inaccuracy.

(J Med Internet Res 2023;25:e49368) doi: 10.2196/49368

KEYWORDS

ChatGPT; chatbot; artificial intelligence; education technology; medical education; machine learning; chatbots; concise review; review methods; review methodology; SWOT

Introduction

ChatGPT is a 175-billion-parameter natural language processing model that can generate conversation-style responses to user input. Parameters are the numerical values that determine how a neural network processes input data and produces output data. They are learned from data during the training process, and they encode the knowledge and skills of the model [1,2]. Since its launch by OpenAI, ChatGPT has stimulated widespread conversation and momentum across various human sciences, including medicine, as demonstrated by over 1250 publications related to ChatGPT indexed in PubMed in 2023 so far. General interest is reflected by the rapid increase in user numbers within a very short time-after just 1 week, ChatGPT had reached the 1 million-user mark, and by January 2023, the number of users had exceeded 100 million [1,3]. It enables humans to interact and discuss with an artificial intelligence (AI) chatbot on a broad range of topics. Since it became known that ChatGPT was able to pass the United States Medical Licensing Examination, performing at a level comparable to a third-year medical student [4], it has garnered even more interest. Preliminary evidence shows that ChatGPT has promising applications across the clinical workflow, especially regarding clinical decision-making, narrowing the final diagnosis, and clinical management [5]. Moreover, ChatGPT and similar generative AI tools have specific applications within medical education, including clinical vignette generation and communication training with AI virtual patients [5-7]; however, it poses several challenges that universities need to carefully address [8]. Tested on the verbal scale of the Wechsler Adult Intelligence Scale (WAIS), an intelligence quotient of 155 was demonstrated for ChatGPT [9]. When the authors asked ChatGPT to explain its role in medical literature," it responded as follows:

As an AI language model, ChatGPT's role in medical literature is to provide information and insights related to health care and medicine based on its knowledge of the subject matter. ChatGPT can help researchers and health care professionals by providing quick access to information on a wide range of medical topics, including symptoms, diagnoses, treatments, and outcomes. ChatGPT can also assist in generating summaries and abstracts of medical articles and papers, making it easier for health care professionals to keep up with the latest developments in their field. Additionally, ChatGPT can assist in developing natural language processing applications that can improve the efficiency and accuracy of medical data analysis and interpretation. However, it's important to note that ChatGPT is not a substitute for medical professionals and should not be used to make medical decisions. It is always recommended to consult with a licensed health care provider for medical advice and treatment. [ChatGPT response]

Although these features are promising, the technology is still vulnerable to misuse and abuse due to the lack of regulation. It can be tempting to save time by creating manuscripts using ChatGPT. However, there is a risk that false information will be published or used, which may compromise patient safety. Since its release, over 200 papers have addressed issues using the search term "ChatGPT." This review aimed to analyze the role of ChatGPT in the medical literature during the first 3 months after its release. The main focus is on how ChatGPT can help authors with research, data analysis, and even writing medical papers, although articles, publishing journals, and authors were studied in more detail. To provide an overview, uncover more information about its potential, and create hypotheses for further research, we undertook this concise review on the role of ChatGPT including a SWOT (strengths, weaknesses, opportunities, and threats) analysis to define its potential, especially for medical publishing.

Methods

Study Design: Search Strategy and Selection Criteria

To extract all publications related to or considering ChatGPT from December 1, 2022, to March 31, 2023, we searched PubMed for the Medical Subject Headings (MeSH) term "ChatGPT" in AllFields on April 1, 2023. All full-text publications in German or English published during this period were included. In the next step, we excluded articles that were not written about the use of ChatGPT but where ChatGPT acted as coauthor or where ChatGPT was used to support the text generation (Textbox 1). The final reference list was generated based on relevance to the broad scope of this review.



Textbox 1. Inclusion and exclusion criteria for the literature search.

Inclusion criteria

- Fully retrievable articles
- Published in English and German
- Published between December 1, 2022, and March 31, 2023
- Human author(s)

Exclusion criteria

- Nonretrievable articles
- Published in languages other than English or German
- Published before December 1, 2022, and after March 31, 2023
- Authored by ChatGPT

Data Analysis

All accessible publications were evaluated according to the following specifications by the authors: title, publication date, PubMed ID, author (including publication experience, specialization), journal title (including specialization), impact factor, publication modus, article type, study type, publication speed, type of ChatGPT integration, content analysis (evidence for strengths, weaknesses, opportunities, and/or threats), and further comments. For evaluation, the full texts were divided between the authors. Each publication was first evaluated and categorized by an author concerning the given criteria. The results were documented in a collaborative table. To increase reliability and validity, the results were then reviewed by a second author. In case of differences in the evaluation, a discussion between the 2 authors took place until a consensus was reached. If it could not be reached, the evaluation was performed by a third author, and a majority decision was made.

Articles

Publications were primarily classified according to the specifications of PubMed. For better comprehensibility, a "studies" category was created, defined as "a method of research in which a problem is identified, relevant data are gathered, a hypothesis is formulated, and the hypothesis is empirically tested." This category allowed for differentiation from nonempirical publications such as editorials. All identified articles were scanned for qualitative (ie, a collection of text-based data, such as interviews and focus groups, usually hypothesis-generating) versus quantitative (ie, a collection of number-based data, such as measurements, questionnaires with associated statistics, usually hypothesis-testing) content. We also chose to discriminate between mixed methods research (ie, a combination of qualitative and quantitative content) and reviews and meta-analysis. Article contents were analyzed in reporting on the use or actual, partial, or full use of ChatGPT in the drafting of the article. In this context, attention was also paid to the correlation between the share of ChatGPT in the preparation of the manuscript (not at all, partially, completely) and the achievable impact factor. The conclusions of the articles were entered in the table in the short form to allow for later SWOT analysis. The publishing speed and the presence of preprint were documented. To better compare the course of the

https://www.jmir.org/2023/1/e49368

number of actual published papers on ChatGPT, an article count was displayed by week and compared to weekly article releases during the COVID-19 outbreak, an impactful event in current medical history.

Journals

Journals publishing articles on ChatGPT were evaluated regarding title, discipline of natural science, actual impact factor, and open access versus traditional publishing. The publishing speed (including preprint servers) was only included in the analysis if given in the paper itself or alongside the date of submission and publication.

Authors

To obtain further information about the authors publishing on the topic, the number of first and last authorships of each first author in PubMed was determined for the years between 2020 and 2022. In addition, the specialty of each first author was also determined via the affiliations in the paper, PubMed, or via the ORCID (Open Researcher and Contributor ID).

SWOT Analysis

The authors broke into groups of two to work on the first part of the SWOT analysis. They examined all the documented comments to determine if they fit into their area. A consensus had to be reached between the authors. If this was not possible, it was noted. Then, the domains were swapped (strengths versus threats and weaknesses versus opportunities), and the previous divisions were reviewed. Particular attention was paid to comments that did not previously reach consensus. Each comment was thus assigned by 4 authors. If consensus decisions could not be reached, the majority decision counted.

Statistical Analysis

The primary endpoint was strengths, weaknesses, opportunities, and threats of ChatGPT use in medical literature, with subcategories of author, article, and journal type. All data were analyzed on a descriptive basis. Data were presented as means and SD unless otherwise stated. Statistical analysis was performed descriptively using Microsoft Excel for Office 365 (Microsoft Corp) and PSPP (GNU Project, Free Software Foundation). The Student *t* test, Levene test, and Mann-Whitney

U test were applied as appropriate. P<.05 was considered to represent statistical significance.

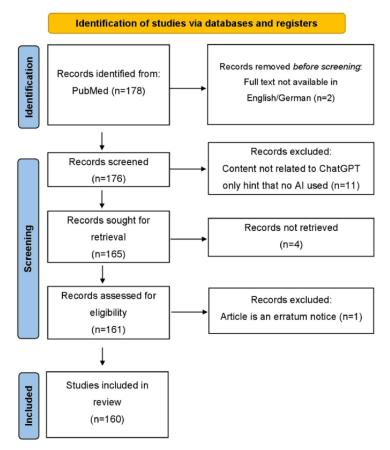
Results

Overview

From December 1, 2022, to March 31, 2023, a total of 178 papers using the search term "ChatGPT" were published in PubMed; among them, 6 (3.4%) were published in December

2022, 16 (9%) in January 2023, 68 (38.2%) in February 2023, and 88 (49.4%) in March 2023. After a thorough human review, 18 papers had to be excluded, 11 because they were written with but not about ChatGPT, 4 that were not retrievable as full text, and 2 that were neither written in English nor German. One paper was just an erratum note. Figure 1 shows the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of ChatGPT-related publications (modified after [10]).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of studies related to ChatGPT.



Articles

The majority of the articles were brief statements like editorials, or letters to the editor (n=77, 48.1%). Essays or commentaries (n=43, 26.9%) represented the second largest portion of the articles. Research studies not further specifiable (n=21, 13.3%), reviews, news, case reports (each n=6, 3.8%), and meta-analyses (n=1, 0.6%) were less frequent. No randomized controlled trial (RCT) could be identified. Table 1 shows the distribution of

article types according to the specifications of PubMed publication type. Of all the articles, 80% (n=128) contained nonempirical data and 20% (n=32) contained empirical data. Among these, 6.9% (n=11) were qualitative, 8.8% (n=14) were quantitative, and 1.9% (n=3) were mixed methods. Regarding the proportion of ChatGPT within the article, 11.9% (n=19) of all articles were written at least partially with ChatGPT. The average impact factors are displayed in Table 2.



Table 1. Distribution of article types according to PubMed publication type (N=160)^a.

Article type	Values, n (%)	
Editorial	77 (48.1)	
Opinion/essay	43 (26.9)	
Study	21 (13.1)	
Case report	6 (3.8)	
News	6 (3.8)	
Review	6 (3.8)	
Meta-analysis	1 (0.6)	

^aThe name of the publication was adopted analogously to the terminology in PubMed. "News" in this context means information or reports about recent events, whereas "study" means an organized experiment.

Table 2. Average impact of articles written with or without ChatGPT.

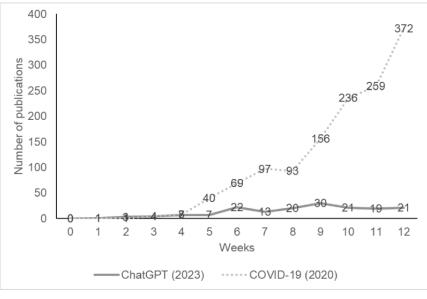
Articles with an impact factor	Values, n (%)	Median (IQR)	Range
Written without ChatGPT	120	5.622 ^a (3.282-13.89)	0.646-96.216
Written with ChatGPT	18	4.403 ^a (3.499-7.188)	1.15-29.983

^aIndicates nonsignificant *P* values. The Levene test (P=.001) indicated that the *t* test, although primarily significant (P=.003), was not robust, so statistical significance was also calculated using the Mann-Whitney *U* test (P=.4).

To illustrate scientific interest in the topic, as measured by the number of publications, Figure 2 shows the comparison to the

number of COVID-19–related publications during the first 12 weeks of 2020 [11].

Figure 2. Weekly publications on ChatGPT compared to COVID-19. Data for COVID-19–related publications were taken from Kambhampati et al [11]. During the first 4 weeks, there was no marked differences in publications (ChatGPT/COVID-19, week 1: 1/1, week 2: 3/1; week 3: 4/4; week 4: 7/8).



Journals

The papers were published in journals covering a wide range of scientific disciplines. Table 3 shows an overview of the specialty distribution of journals published on ChatGPT. The current impact factor of the represented journals ranged from 0 to 96.216, with a median of 5.144 (IQR 3.352-11.325). Overall, 45.6% (n=73) of all articles were published "traditionally" in contrast to 54.4% (n=87) that were published as open access. Of those, 5% (n=8) of "open access" publications were provided on preprint servers in advance. Data on publication speed were accessible in 33.1% (n=53) of all evaluated articles. The average time to publication was 16 days, ranging from 4 to 83 days.



Table 3. Disciplines of journals published on ChatGPT (N=160).

Discipline of first authors	Values, (%)
Clinical medicine	72 (45)
Other	44 (27.5)
Education	19 (11.9)
Theoretical medicine	19 (11.9)
Nature sciences	6 (3.8)

Authors

The authors of the reviewed papers had a median of 5 (IQR 1-12; range 0-94) first and a median of 1 (IQR 0-6; range 0-61) last authorships in the years spanning 2020 to 2022. Their area

of expertise spanned all medical specialties, including science journalism, bioinformatics, nursing, humanities, economics, and law. Table 4 gives an overview of the specialty distribution of the first authors.

Table 4. Disciplines of first authors publishing on ChatGPT (N=160).

Journal discipline	Values, n (%)
Clinical medicine	82 (51.3)
Other	33 (20.6)
Theoretical medicine	23 (14.4)
Nature sciences	10 (6.3)
Not determined	7 (4.4)
Education	5 (3.1)

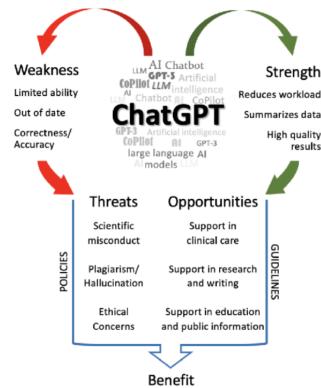
SWOT Analysis

We were able to detect 408 quotes (in the 160 papers included) that provided information on strengths, weaknesses, opportunities, and threats. Of those, most were related to weaknesses (n=142, 34.8 %) and least to opportunities (n=68, 16.7%). Quotes on strengths (n=117, 28.7%) and threats (n=81, 19.9%) were mentioned less frequently. Among the most prevalently cited weaknesses were limited abilities [12,13], lack of accuracy/correctness [14,15], citation problems [16,17], and the need for verification [18,19]. Limited abilities in the context of ChatGPT meant not being able to provide reliable facts and sources [12], limitations in understanding complex scientific concepts, or a limited scope of expertise and a lack of accountability [13]. In addition, limited abilities included, for example, the inability to think like a human and evaluate

interpersonal aspects, as well as recognize emotions and act on them. Strengths, on the other hand, included reduced workload [14,20-22], data summarization [23], and results rated as positive and high-quality by the authors themselves [15,24-28]. Among the threats captured most frequently were plagiarism (meaning not providing sources), hallucination (meaning making up sources), scientific misconduct, and ethical concerns [25,29-31], whereas major opportunities were seen in supporting different faculties [24,32,33]. For example, Goodman et al [32] mentioned that ChatGPT could improve patient education by personalizing and targeting information to the patient's education level. Due to the variability in the mentions, we decided to carry out a semiqualitative SWOT analysis. Figure 3 shows the results, conclusions, and suggestions. Raw data on all the aforementioned information can be found in Multimedia Appendix 1.



Figure 3. Results, conclusions, and suggestions of SWOT (strengths, weaknesses, opportunities, and threats) analysis of ChatGPT in the medical literature.



Discussion

Principal Results

To our knowledge, this is the first review of ChatGPT including a SWOT analysis to date, summarizing all articles published in PubMed since ChatGPT launched in November 2022 through the end of March 2023. In addition to a whole series of metric results, the experiences of authors writing about ChatGPT were also critically reviewed in the context of the SWOT analysis. The density and quality of data available at the time of data collection did not allow the use of further, more specific methods and tools to assess the risk of bias. To the best of our knowledge, no similarly comprehensive study on the topic exists to date.

Concerning the article types, it was interesting to see that so far, only 1 RCT has been published about ChatGPT [34]. RCTs on ChatGPT are potentially ambitious and partially difficult to accomplish, but their data are needed to gain more evidence. In this review, the majority of the articles were of a shorter nature (ie, editorials, letters, features, essays, or commentaries).

Journals from the ranks of clinical medicine have published the most articles on ChatGPT, followed by education and others. This resembles the results from authorship. Both aspects (ie, authorship and journal) show the wide application potential for ChatGPT across many specialist areas, as would be expected from a large language model (LLM). When considering the impact factor of the journals, it is interesting to note that some articles were published in journals with no impact factor, although even highly reputable fundamental research journals, such as Science or Nature, and clinical journals, such as the British Medical Journal (BMJ) or the Lancet, have covered the topic. This proves the importance and relevance of ChatGPT

```
https://www.jmir.org/2023/1/e49368
```

in science, education, and clinical work. Out of the 160 papers that were evaluated, 138 (86.3%) were published in journals with an impact factor. Whether a paper was or was not at least partially written with ChatGPT did not have a significant influence on the journal's impact factor.

Despite the extensive application possibilities of ChatGPT in both medical and nonmedical fields, the publication frequency has not kept pace with the surge seen in COVID-19–related research. This is somewhat surprising considering the significant media attention ChatGPT has received due to remarkable achievements such as passing medical licensing exams, assisting in radiologic decision-making interpretation, and even generating patient clinic letters [4,25,26].

Because ChatGPT is also an event of global significance, we deliberately chose the pandemic as a point of comparison. However, the global health crisis likely served as a stronger catalyst to address the issue, although no relevant difference was seen during the first 4 weeks after the onset of the pandemic compared to the appearance of ChatGPT. Interestingly, it is worth noting that publication speed, if quantifiable, averaged 16 days (ranging from 4 to 83), which is significantly faster than what has been reported in other studies published in biomedical journals [35-37]. However, it is important to recognize that the majority of papers in our review were editorials or commentaries. The quicker publication speed is not as surprising in this context, as these types of publications are easier to produce than larger studies. Furthermore, it is important to note that this statement needs some qualification since corresponding data were only available for about one-third of the publications. This necessitates a more nuanced and less definitive discussion. Alongside the spectrum of journals

containing ChatGPT publications, the proportion of preprint and open-access articles should also be considered as influential factors. A multivariate analysis showed that web-based publishing is strongly associated with reduced submission-to-publication time [37]. It must be emphasized that data on submission speed were only available in about one-third of all articles, which is a limiting factor. However, in combination with the higher proportion of quantitative and nonempirical data, we assume that open access and preprints contributed to the fast publication times.

It is difficult to create a comprehensive author profile in this area due to distribution patterns and publication frequencies. However, most authors seem to come from the "clinical medicine" field, which entails working directly with patients. Surprisingly, education, despite being frequently mentioned and an obvious ChatGPT application area, had a lower representation.

Notably, many authors were not newcomers to publishing. Increased public interest in ChatGPT's medical application seemed to have given them an incentive to conduct research and publish quicker than before. The worth of engaging with ChatGPT is evident from the average impact factor that could be achieved with a publication on this topic, and whether or not an article was written with the help of ChatGPT did not seem to affect the outcome. The median impact of 5.144 (regardless of ChatGPT usage) falls within a range where only 7.9% (n=979) of other journals in a comparison of 13,000 selected scientific journals across 27 major prominent research categories were situated [38].

In addition, despite the allure of using AI in manuscript creation, only slightly under 12% (18/160) of authors used AI, or at least indicated they had. However, since there are no official guidelines for declaring ChatGPT usage in new research, it could be possible that some authors did not disclose its usage. Acknowledging AI use for writing assistance was among the most frequently cited SWOT (which will be discussed later in this section). Thus far, however, ChatGPT usage is not clearly superior or inferior in terms of the impact achieved.

Interestingly, in our SWOT analysis, ChatGPT weaknesses were identified more prominently than strengths, which came in second. It must be said, however, that for the sake of comparability, we weighed the strengths, weaknesses, opportunities, and threats listed by authors equally, regardless of the type of article they were listed in. This is important because essays and opinion pieces were the prevailing types of articles in our analysis, outweighing a more nuanced representation of advantages and disadvantages based on actual trials and scientific research. Given the sheer number of articles that we analyzed, this approach allowed us to make fair comparisons at the time.

When examining the frequency of the SWOT citations, it appeared that many authors provided descriptive accounts of weaknesses and strengths but offered fewer perspectives or ideas for further handling or development of ChatGPT developed from their findings. This is reinforced by the fact that threats were only cited in 19.9% (n=81) papers, and

XSL•FO

opportunities were mentioned in 16.7% (n=68) compared to weaknesses (n=142, 34.8%).

The SWOT analysis, originally defined as a "strategic planning and strategic management technique used to help a person or organization identify strengths, weaknesses, opportunities, and threats related to business competition or project planning" [39], is used to identify the favorable and unfavorable internal and external factors that impact the objectives of a venture. It is valued for its usability and status as a "tried-and-true tool" for strategic analysis. However, points of criticism include limitations such as potential bias, inconsistency in analysis compliance, and the dominance of certain team members [40-43]. To address some of these shortcomings, we used a modified Delphi process to analyze the quotes. Furthermore, we designed our SWOT analysis as a starting point for discussion, considering it a suitable tool for analyzing ChatGPT in its early stages and generating some ideas for moving forward, particularly in a rapidly changing environment.

Thus far, ChatGPT has been used to write essays, pass exams, translate knowledge for various peer groups, and generate comments on a wide array of topics. However, it has become clear during its application that ChatGPT is, at least until 2021, "apparently "knowledge limited." This limitation means that it often generates information and facts that are fictional, detectable only by experts with the relevant expertise.

The existing publications on the topic serve a dual purpose. On the one hand, they contribute to the ongoing improvement of AI, making it safer for use in various contexts. On the other hand, they have identified fields where ChatGPT can presumably be applied safely. These applications include summarizing large data sets and producing easily understandable text.

Nonetheless, caution must be exercised when using ChatGPT, as in several cases, sources have been freely invented (hallucination) [14,25] or copied (plagiarism) [16,29,30,44]; therefore, the accurateness of content created by ChatGPT must always be questioned.

Areas unsuitable for ChatGPT's application include writing scientific papers with references, composing resumes, or writing speeches. In these domains, it has already been shown that ChatGPT can create partially or completely fictitious passages [16,29,30,44], which cannot withstand critical review.

After our concise review, it is clear that ChatGPT serves more as an exploratory tool than a reliable tool for scientific work. This is not inherently negative; playful interactions with ChatGPT can identify strengths and areas for improvement, which developers and programmers can then address. In any case, it is crucial to avoid monopolization (eg, through the displacement of competitors), as it could lead to issues like a lack of transparency in data sources [45] or global commercialization with ethical-economic imbalances [29]. So far, over 30 alternatives for ChatGPT exist, including OpenAI Playground, Jasper Chat, Bard, and Bing AI [46]. Ideally, such large-scale software would be open source.

Another area of major concern is the ability to detect AI-generated scientific output. Existing AI detection software, such as GPTZero, or related products like GLTR, GPTKit,

OpenAI, and Output Detector are based on scanning for perplexity (typically lower in AI) and burstiness (typically higher in AI) [47]. Perplexity and burstiness are key AI metrics, aiding our understanding of how LLMs write and how humans can detect AI-generated content. Perplexity evaluates LLM performance, while burstiness is linked to the SD of sentence lengths, making it a measure of text efficiency. Their most obvious, clear limitation is that texts are not analyzed for context but only for writing patterns, potentially allowing AI to go undetected.

Initial data from a comparative study of AI output detectors, plagiarism detectors, and blinded human reviewers show promising results. In that study, AI output detectors successfully identified most AI-generated abstracts, with an area under the receiver operating characteristic (AUROC) curve of 0.94 [48]. However, it is vital to keep pace with the rapid development of new AI interactions like GPT-3 or GPT-4, all while maintaining an alert, skeptical perspective to uphold data integrity.

The AI Phenomenon Worldwide

ChatGPT, specializing in written conversations, has broad applications across countless domains, underlining the absence of any concrete regulation. As of April 2023, ChatGPT is unavailable in China and other countries with heavy internet censorship, including North Korea, Iran, and Russia. While it is not officially blocked, OpenAI does not allow users from these countries to sign up. Interestingly, several large tech companies in China are developing alternatives [49]. Italy became the first Western country to ban ChatGPT, and various other Western governments, such as Germany, the United Kingdom, and Canada, are currently exploring ways to regulate AI use. The United States has not yet proposed any formal rules to bring oversight to AI technology [49].

The AI Phenomenon in Medicine

During the review, the threat of being unable to detect the use of ChatGPT in the writing process and the associated lack of reproducibility of its written content were mentioned [33]. For example, Aubignat and Diab [16] cited a study that examined the ability of both a plagiarism detector and a group of reviewers to accurately identify abstracts written by ChatGPT across various journals. Unsurprisingly, they struggled to correctly determine whether the content was original or plagiarist, leading to the conclusion that ChatGPT can write "credible scientific abstracts" [16]. Suggested solutions include consistently mentioning ChatGPT's possible contribution [24], albeit not as an author [44,50] but more as an "acknowledgement" [51].

Recently, the World Association of Medical Editors (WAME) clearly stated that "AI cannot be an author" and emphasized the responsibility and reproducibility of human authors [52]. A similar stance is also reflected in the criteria of the International Committee of Medical Journal Editors (ICMJE) [53]. Major publishers have started to integrate these recommendations into their policies [54]. Other sources have suggested the inclusion of AI output detectors in the editorial process and clear disclosure if these technologies are used [48].

AI has always fired the human imagination, as can be seen from famous movies like *Star Trek, Star Wars, Terminator*, and *Aliens*, always accompanied by a resonating, undefined fear that AI may "overtake" us one day—with potentially deleterious consequences. Despite these easily visualizable and seemingly apocalyptic dangers, one should not dismiss the sheer unlimited and fascinating possibilities offered by AI. Nevertheless, after comparing and evaluating the various strengths, weaknesses, opportunities, and threats posed by ChatGPT, we, the authors, firmly believe that strict and clear regulations on many levels are necessary to fully leverage its potential. In our humble judgment, it seems wise to keep a low profile right now, as ChatGPT itself points out at least some of its weaknesses when asked:

As an artificial intelligence language model, I do not have a role in the discussion about ChatGPT in the medical literature. However, I can provide information and answer questions related to my capabilities and limitations as a language model, as well as share insights on how natural language processing technology is being applied in health care and medical research. [ChatGPT response]

Acknowledgments

The authors would like to thank Dr Christian Burisch (private researcher, Essen, Germany) for his help with the statistics and drafting of the manuscript.

Disclaimer

All listed authors declare that all 4 criteria for authorship in the International Committee of Medical Journal Editor (ICMJE) recommendations were met individually. All authors confirm that they had full access to all the data in the study and accept responsibility for submitting the data for publication. All authors claim responsibility and accountability for the originality, accuracy, and integrity of the work. Notably, artificial intelligence (AI) and AI-assisted technologies were used twice as examples and clearly marked as such in the manuscript.

Data Availability

Without exception, all data collected for this study, including individual participant data and a data dictionary defining each field in the set, will be made available to others. This also includes additional, related documents that will be published and made available upon request by email to the corresponding author.



```
https://www.jmir.org/2023/1/e49368
```

Authors' Contributions

DG, SN, CW, YR, LR, JE, FB, and TS were involved in the concept and design of the above review. All authors made major contributions to literature search, data collection, data analysis, and data interpretation. DG, SN, CW, YR, LR, JE, FB and TS wrote and edited the manuscript. DG and TS created all the figures. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist and raw data from the review. [DOCX File, 113 KB-Multimedia Appendix 1]

References

- 1. Shevchuk V. GPT-4 parameters explained: everything you need to know. Level Up Coding. URL: <u>https://levelup.gitconnected.com/gpt-4-parameters-explained-everything-you-need-to-know-e210c20576ca</u>, [accessed 2023-09-26]
- Meffert K. Künstliche intelligenz glossar die wichtigsten begriffe. DSGVO Blog. URL: <u>https://dr-dsgvo.de/kuenstliche-intelligenz-glossar-die-wichtigsten-begriffe</u> [accessed 2023-09-26]
- 3. Dennean K, Gantori S, Limas DK, Pu A, Gilligan R. Let's chat about ChatGPT. UBS. URL: <u>https://www.ubs.com/us/en/</u> wealth-management/insights/market-news/article.1585717.html [accessed 2023-09-26]
- Gilson A, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA, et al. How Does ChatGPT perform on the United States Medical Licensing Examination? The implications of large language models for medical education and knowledge assessment. JMIR Med Educ 2023 Feb 08;9:e45312 [FREE Full text] [doi: 10.2196/45312] [Medline: 36753318]
- Rao A, Pang M, Kim J, Kamineni M, Lie W, Prasad AK, et al. Assessing the utility of ChatGPT throughout the entire clinical workflow. medRxiv. Preprint posted online Feb 26, 2023 [FREE Full text] [doi: 10.1101/2023.02.21.23285886] [Medline: 36865204]
- 6. Benoit J. ChatGPT for clinical vignette generation, revision, and evaluation. medRxiv. Preprint posted online Feb 8, 2023 [doi: 10.1101/2023.02.04.23285478]
- Shorey S, Ang E, Yap J, Ng ED, Lau ST, Chui CK. A virtual counseling application using artificial intelligence for communication skills training in nursing education: development study. J Med Internet Res 2019 Oct 29;21(10):e14658 [FREE Full text] [doi: 10.2196/14658] [Medline: 31663857]
- Sinhaliz S, Burd L, Du PJ. How ChatGPT could revolutionize academia the AI chatbot could enhance learning, but also creates some challenges. IEEE Spectrum. URL: <u>https://spectrum.ieee.org/how-chatgpt-could-revolutionize-academia</u> [accessed 2023-09-26]
- 9. Roivainen E. I gave ChatGPT an IQ test. Here's what I discovered. Scientific American. URL: <u>https://www.scientificamerican.com/article/i-gave-chatgpt-an-iq-test-heres-what-i-discovered/</u>, [accessed 2023-01-05]
- Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: <u>10.1136/bmj.n71</u>] [Medline: <u>33782057</u>]
- Kambhampati SB, Vaishya R, Vaish A. Unprecedented surge in publications related to COVID-19 in the first three months of pandemic: A bibliometric analytic report. J Clin Orthop Trauma 2020 May;11(Suppl 3):S304-S306 [FREE Full text] [doi: 10.1016/j.jcot.2020.04.030] [Medline: 32405191]
- 12. Wen J, Wang W. The future of ChatGPT in academic research and publishing: A commentary for clinical and translational medicine. Clin Transl Med 2023 Mar 20;13(3):e1207 [FREE Full text] [doi: 10.1002/ctm2.1207] [Medline: 36941774]
- 13. Biswas SS. Potential use of ChatGPT in global warming. Ann Biomed Eng 2023 Jun 01;51(6):1126-1127 [doi: 10.1007/s10439-023-03171-8] [Medline: 36856927]
- 14. Shen Y, Heacock L, Elias J, Hentel KD, Reig B, Shih G, et al. ChatGPT and other large language models are double-edged swords. Radiology 2023 Apr 01;307(2):e230163 [doi: 10.1148/radiol.230163] [Medline: 36700838]
- 15. Zhu J, Jiang J, Yang M, Ren ZJ. ChatGPT and environmental research. Environ Sci Technol 2023 Mar 21 [doi: 10.1021/acs.est.3c01818] [Medline: 36943179]
- Aubignat M, Diab E. Artificial intelligence and ChatGPT between worst enemy and best friend: The two faces of a revolution and its impact on science and medical schools. Rev Neurol (Paris) 2023 Jun;179(6):520-522 [doi: <u>10.1016/j.neurol.2023.03.004</u>] [Medline: <u>36959064</u>]
- 17. Benichou L, ChatGPT. The role of using ChatGPT AI in writing medical scientific articles. J Stomatol Oral Maxillofac Surg 2023 Oct;124(5):101456 [doi: 10.1016/j.jormas.2023.101456] [Medline: 36966950]
- Buvat I, Weber W. Nuclear medicine from a novel perspective: Buvat and Weber talk with OpenAI's ChatGPT. J Nucl Med 2023 Apr 23;64(4):505-507 [doi: <u>10.2967/jnumed.123.265636</u>] [Medline: <u>36958855</u>]

RenderX

- Johnson D, Goodman R, Patrinely J, Stone C, Zimmerman E, Donald R, et al. Assessing the accuracy and reliability of AI-generated medical responses: an evaluation of the Chat-GPT model. Res Sq 2023 Feb 28;28:A [FREE Full text] [doi: 10.21203/rs.3.rs-2566942/v1] [Medline: 36909565]
- 20. Yeo YH, Samaan JS, Ng WH, Ting P, Trivedi H, Vipani A, et al. Assessing the performance of ChatGPT in answering questions regarding cirrhosis and hepatocellular carcinoma. Clin Mol Hepatol 2023 Jul 22;29(3):721-732 [FREE Full text] [doi: 10.3350/cmh.2023.0089] [Medline: 36946005]
- 21. Rillig MC, Ågerstrand M, Bi M, Gould KA, Sauerland U. Risks and benefits of large language models for the environment. Environ Sci Technol 2023 Mar 07;57(9):3464-3466 [doi: <u>10.1021/acs.est.3c01106</u>] [Medline: <u>36821477</u>]
- 22. Gabrielson AT, Odisho AY, Canes D. Harnessing generative artificial intelligence to improve efficiency among urologists: welcome ChatGPT. J Urol 2023 May;209(5):827-829 [doi: 10.1097/ju.0000000003383]
- 23. Moons P, Van Bulck L. ChatGPT: can artificial intelligence language models be of value for cardiovascular nurses and allied health professionals. Eur J Cardiovasc Nurs 2023 Oct 19;22(7):e55-e59 [doi: <u>10.1093/eurjcn/zvad022</u>] [Medline: <u>36752788</u>]
- 24. Hosseini M, Horbach SPJM. Fighting reviewer fatigue or amplifying bias? Considerations and recommendations for use of ChatGPT and other large language models in scholarly peer review. Res Integr Peer Rev 2023 May 18;8(1):4 [FREE Full text] [doi: 10.1186/s41073-023-00133-5] [Medline: 37198671]
- 25. Rao A, Kim J, Kamineni M, Pang M, Lie W, Succi MD. Evaluating ChatGPT as an adjunct for radiologic decision-making. medRxiv. Preprint posted online Feb 7, 2023 [FREE Full text] [doi: 10.1101/2023.02.02.23285399] [Medline: 36798292]
- 26. Ali SR, Dobbs TD, Hutchings HA, Whitaker IS. Using ChatGPT to write patient clinic letters. Lancet Digit Health 2023 Apr;5(4):e179-e181 [FREE Full text] [doi: 10.1016/S2589-7500(23)00048-1] [Medline: 36894409]
- Johnson S, King A, Warner E, Aneja S, Kann B, Bylund C. Using ChatGPT to evaluate cancer myths and misconceptions: artificial intelligence and cancer information. JNCI Cancer Spectr 2023 Mar 01;7(2):a [FREE Full text] [doi: 10.1093/jncics/pkad015] [Medline: 36929393]
- 28. Schorrlepp M, Patzer K. ChatGPT in der hausarztpraxis: die künstliche intelligenz im check. MMW Fortschr Med 2023 Mar 24;165(6):12-16 [doi: 10.1007/s15006-023-2473-3] [Medline: <u>36961685</u>]
- 29. Salvagno M, Taccone FS, Gerli AG. Can artificial intelligence help for scientific writing? Crit Care 2023 Feb 25;27(1):75 [FREE Full text] [doi: 10.1186/s13054-023-04380-2] [Medline: 36841840]
- 30. Zheng H, Zhan H. ChatGPT in scientific writing: a cautionary tale. Am J Med 2023 Aug;136(8):725-726.e6 [doi: 10.1016/j.amjmed.2023.02.011] [Medline: 36906169]
- 31. Baumgartner C. The potential impact of ChatGPT in clinical and translational medicine. Clin Transl Med 2023 Mar;13(3):e1206 [FREE Full text] [doi: 10.1002/ctm2.1206] [Medline: 36854881]
- 32. Goodman RS, Patrinely JR, Osterman T, Wheless L, Johnson DB. On the cusp: Considering the impact of artificial intelligence language models in healthcare. Med 2023 Mar 10;4(3):139-140 [doi: 10.1016/j.medj.2023.02.008] [Medline: 36905924]
- Hallsworth JE, Udaondo Z, Pedrós-Alió C, Höfer J, Benison KC, Lloyd KG, et al. Scientific novelty beyond the experiment. Microb Biotechnol 2023 Jun 14;16(6):1131-1173 [FREE Full text] [doi: 10.1111/1751-7915.14222] [Medline: 36786388]
- 34. Dunn C, Hunter J, Steffes W, Whitney Z, Foss M, Mammino J, et al. Artificial intelligence-derived dermatology case reports are indistinguishable from those written by humans: A single-blinded observer study. J Am Acad Dermatol 2023 Aug;89(2):388-390 [doi: 10.1016/j.jaad.2023.04.005] [Medline: 37054810]
- Andersen MZ, Fonnes S, Rosenberg J. Time from submission to publication varied widely for biomedical journals: a systematic review. Curr Med Res Opin 2021 Jun 06;37(6):985-993 [doi: <u>10.1080/03007995.2021.1905622</u>] [Medline: <u>33735591</u>]
- 36. Asaad M, Rajesh A, Banuelos J, Vyas KS, Tran NV. Time from submission to publication in plastic surgery journals: The story of accepted manuscripts. J Plast Reconstr Aesthet Surg 2020 Feb;73(2):383-390 [doi: 10.1016/j.bjps.2019.09.029]
- Sebo P, Fournier J, Maisonneuve H. Is statistician involvement as co-author associated with reduced time to publication of quantitative research in general medical journals? A bibliometric study. Fam Pract 2019 Jul 31;36(4):431-436 [doi: 10.1093/fampra/cmy115] [Medline: 30476030]
- 38. Joanna W. What's a good impact factor (ranking in 27 categories). SCI Journal. URL: <u>https://www.scijournal.org/articles/good-impact-factor</u> [accessed 2023-04-22]
- 39. SWOT analysis. Wikipedia. URL: https://en.wikipedia.org/wiki/SWOT_analysis [accessed 2023-04-23]
- 40. Dess DG, Lumpkin GT, Eisner AB, McNamara G, editors. The limitations of SWOT analysis. In: Strategic Management: Text and Cases. 6th edition. New York, NY: McGraw-Hill/Irwin; 2012:82
- 41. Hill T, Westbrook R. SWOT analysis: It's time for a product recall. Long Range Plann 1997 Feb;30(1):46-52 [doi: 10.1016/s0024-6301(96)00095-7]
- 42. Koch AJ. SWOT does not need to be recalled: It needs to be enhanced. www.westga.edu. URL: <u>https://www.westga.edu/</u> <u>~bquest/2000/swot1.html</u> [accessed 2023-05-01]
- 43. Chermack TJ, Kasshanna BK. The use and misuse of SWOT analysis and implications for HRD professionals. Hum Resour Dev Int 2007 Dec;10(4):383-399 [doi: 10.1080/13678860701718760]



- 44. Stuart-Harris C. Epidemiology of influenza in man. Br Med Bull 1979 Jan;35(1):3-8 [doi: 10.1093/oxfordjournals.bmb.a071538] [Medline: 367483]
- Huang J, Yeung AM, Kerr D, Klonoff DC. Using ChatGPT to predict the future of diabetes technology. J Diabetes Sci Technol 2023 May 17;17(3):853-854 [doi: <u>10.1177/19322968231161095</u>] [Medline: <u>36799231</u>]
- 46. Garg S. Top 30 ChatGPT alternatives that will blow your mind in 2023 (free and paid). Writesonic Blog. URL: <u>https://writesonic.com/blog/chatgpt-alternatives/#31-best-chatgpt-alternatives-for-your-to-choose-from</u>, [accessed 2023-04-23]
- 47. GPTZero alternatives. Sourceforge. URL: <u>https://sourceforge.net/software/product/GPTZero/alternatives</u> [accessed 2023-04-23]
- 48. Gao C, Howard F, Markov N, Dyer E, Ramesh S, Luo Y, et al. Comparing scientific abstracts generated by ChatGPT to original abstracts using an artificial intelligence output detector, plagiarism detector, and blinded human reviewers. bioRxiv. Preprint posted online Dec 27, 2022 [FREE Full text] [doi: 10.1101/2022.12.23.521610]
- Moltu C, Stefansen J, Svisdahl M, Veseth M. Negotiating the coresearcher mandate service users' experiences of doing collaborative research on mental health. Disabil Rehabil 2012;34(19):1608-1616 [doi: <u>10.3109/09638288.2012.656792</u>] [Medline: <u>22489612</u>]
- 50. Teixeira da Silva JA. Is ChatGPT a valid author? Nurse Educ Pract 2023 Mar;68:103600 [doi: 10.1016/j.nepr.2023.103600]
- Siegerink B, Pet LA, Rosendaal FR, Schoones JW. ChatGPT as an author of academic papers is wrong and highlights the concepts of accountability and contributorship. Nurse Educ Pract 2023 Mar;68:103599 [doi: <u>10.1016/j.nepr.2023.103599</u>] [Medline: <u>36898252</u>]
- 52. Zielinski C, Winker M, Aggarwal R. Chatbots, ChatGPT, and scholarly manuscripts. WAME recommendations on ChatGPT and chatbots in relation to scholarly publications. World Association of Medical Editors (WAME). URL: <u>https://wame.org/page3.php?id=106</u> [accessed 2023-04-23]
- 53. Defining the role of authors and contributors. International Committee of Medical Journal Editors (ICMJE). URL: <u>https://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html#two</u> [accessed 2023-04-27]
- 54. The use of AI and AI-assisted writing technologies in scientific writing. Elsevier. URL: <u>https://www.elsevier.com/about/</u> policies/publishing-ethics/the-use-of-ai-and-ai-assisted-writing-technologies-in-scientific-writing [accessed 2023-04-22]

Abbreviations

AI: artificial intelligence
AUROC: area under the receiver operating characteristic
BMJ: British Medical Journal
ICMJE: International Committee of Medical Journal Editors
LLM: large language model
MeSH: Medical Subject Headings
ORCID: Open Researcher and Contributor ID
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
SWOT: strengths, weaknesses, opportunities, and threats
WAIS: Wechsler Adult Intelligence Scale
WAME: World Association of Medical Editors

Edited by A Mavragani; submitted 26.05.23; peer-reviewed by DM Hirkani, J Glanville, M Succi; comments to author 05.09.23; revised version received 26.09.23; accepted 27.09.23; published 16.11.23

<u>Please cite as:</u> Gödde D, Nöhl S, Wolf C, Rupert Y, Rimkus L, Ehlers J, Breuckmann F, Sellmann T A SWOT (Strengths, Weaknesses, Opportunities, and Threats) Analysis of ChatGPT in the Medical Literature: Concise Review J Med Internet Res 2023;25:e49368 URL: <u>https://www.jmir.org/2023/1/e49368</u> doi: <u>10.2196/49368</u> PMID: <u>37865883</u>

©Daniel Gödde, Sophia Nöhl, Carina Wolf, Yannick Rupert, Lukas Rimkus, Jan Ehlers, Frank Breuckmann, Timur Sellmann. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 16.11.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the



Gödde et al

Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.