ACCSCIENCE PUBLISHING

Artificial Intelligence in Health

ORIGINAL RESEARCH ARTICLE

Applying ChatGPT to writing scientific articles on the use of telemedicine: Opportunities and limitations

Daniil Kolesnikov[1]*, Alexandra Kozlova[1], Andrey Alexandrov[1], Nikolai Kalmykov[1], Pavel Treshkov[1], Tyler W. LeBaron[2,3], and Oleg Medvedev[4,5]

[1]Faculty of Biomedical Engineering, Bauman Moscow State Technical University, Moscow, Russia

[2]Department of Kinesiology and Outdoor Recreation, Southern Utah University, Cedar City, Utah, United States of America

[3]Molecular Hydrogen Institute, Cedar City, Utah, United States of America

[4]Department of Pharmacology, Faculty of Basic Medicine, Lomonosov Moscow State University, Moscow, Russia

[5]Laboratory of Experimental Pharmacology, National Medical Research Center of Cardiology of the Ministry of Health of the Russian Federation, Moscow, Russia

Abstract

In the rapidly evolving world of technology, artificial intelligence (AI) has significantly integrated into various aspects of our lives, including health-care, education, finance, transportation, and entertainment. Notably, AI has also impacted the writing of textual works such as scientific papers, professional opinions, and educational texts. This study investigates the application of OpenAI's ChatGPT language model in writing scientific articles on telemedicine, specifically in the areas of cardiology, oncology, and remote medical examination. The study uses ChatGPT versions 3.5 and 4 to create articles using three different prompts. The created articles were evaluated based on the reliability of the cited literature references, the impact factor (IF) of the journal in which the sources were published, and the relevance of the sources. The sources were divided into three categories: reliable, semi-reliable, and completely fictitious. The results demonstrate that ChatGPT can produce semantically coherent and error-free texts indistinguishable from human-written texts. However, the reliability of literary references varies significantly. ChatGPT 4, benefitting from its larger training dataset, generates a higher percentage of reliable sources compared to ChatGPT 3.5. The IF analysis indicates the prevalence of high-impact journals among reliable sources, which emphasizes the effectiveness of the model in selecting quality references. The study highlights the need for caution when using AI to write scientific articles due to the potential for biased, unverified, and inaccurate information. It is important to critically evaluate and vet AI-generated content. In addition, the study emphasizes that the correct use of AI and thoughtful drafting of prompts can improve the efficiency and quality of scientific papers. Future advancements in AI technology are expected to further minimize errors and biases.

Keywords: Artificial intelligence; ChatGPT; Biotelemetry; Cardiology

***Corresponding author:**
Daniil Kolesnikov
(kolesnikovda@student.bmstu.ru)

Citation: Kolesnikov D, Kozlova A, Alexandrov A, *et al*. Applying ChatGPT to writing scientific articles on the use of telemedicine: Opportunities and limitations. *Artif Intell Health*. 2024;1(3):53-63. doi: 10.36922/aih.2592

Received: December 31, 2023

Accepted: April 22, 2024

Published Online: July 22, 2024

1. Introduction

In the rapidly evolving world of technology, one of the most significant developments has been the widespread integration of artificial intelligence (AI) into various areas of our lives.[1,2] This spread of AI has ushered in an era of unprecedented change, reshaping the very essence of human existence. From healthcare and education to finance, transportation, and entertainment, the impact of AI permeates various aspects of our interconnected world.[3-8]

AI has deeply impacted various fields, and one area where its influence is particularly notable is in the writing of textual works such as science communication articles, professional opinions, educational texts, technical recommendations, short stories, and other genres.[6,9-11] By automating and simplifying numerous writing tasks, AI technology has empowered writers to enhance their writing processes. AI language models are trained on extensive datasets, encompassing books, articles, and websites. This exposure to diverse writing styles and vocabulary enables AI to imitate human language generation, analyze vast amounts of data, learn from patterns, and generate comprehensive, coherent, and captivating articles.[10,12,13] AI-powered tools such as Grammarly, Lex, AI Writer, Any Word, and Rytr assist writers in real time with proofreading, editing, and formatting written content.[14] In addition, AI enables the swift and efficient generation of written material by optimizing search engine outcomes, generating content concepts, and developing plans.[13] Furthermore, the application of AI has empowered freelance writers to produce error-free, audience-specific, and top-notch content, resulting in heightened customer satisfaction.[15]

Unfortunately, this trend is also reflected in scientific writing.[16] No one can deny the growing popularity of AI-generated scientific articles. Platforms offering AI-generated scientific articles provide great chances to produce scientific, educational, or non-fiction materials rapidly, resulting in the swift production of a significant number of articles.[11,17,18] Although these articles are notable for their originality and accuracy and are often indistinguishable from human-written works, they also raise concerns about the credibility of the information presented and the potential for bias or misleading content.

Chat Generative Pre-trained Transformer (ChatGPT) developed by OpenAI and launched in November 2022, is the leading AI-based platform in the market. This chatbot utilizes AI to automatically generate responses to text prompts.[19] ChatGPT is trained on a vast amount of Internet text data, enabling it to grasp language patterns and structures.[19,20] As a result, it can generate consistent and contextually relevant responses across a wide range of inputs. Researchers have already employed ChatGPT to co-author academic papers and research articles, with at least four scientific papers and preprints credited to it by the end of 2023.[21] This figure represents only the official and disclosed data. The exact number of scientific papers written using ChatGPT and the frequency with which researchers and academics seek help from such platforms remain unknown. Unfortunately, these questions do not have a definitive answer. However, scientific journal publishers are already banning or restricting the use of such IT products due to concerns about ethical issues, fabricated research, and erroneous data.[21-23]

The issue of evaluating and reviewing AI-generated articles is of significant interest to the scientific community. Researchers from Northwestern University and the University of Chicago recently generated research abstracts based on 10 real-life medical journal articles using ChatGPT.[24] They then asked reviewers to analyze the abstracts written by ChatGPT and a human. The study results showed that the reviewers correctly identified only 68% of the generated abstracts and 86% of the genuine abstracts. In addition, they incorrectly identified 32% of the generated abstracts as real abstracts and 14% of the genuine abstracts as generated abstracts. Distinguishing AI-generated content from human-written content proved to be a difficult task.

Researchers, academics, journal editors, and publishers around the world are currently engaged in a debate about the role of such technologies in the field of scholarly writing.[25] Writing scientific articles using AI, including ChatGPT, has several disadvantages.[12,26] First, there is a lack of creative thinking. AI systems excel at generating content using templates and algorithms but struggle to generate new ideas or think outside the box.[26] Second, AI cannot fully grasp the intricacies of human emotions, resulting in a lack of emotional depth in the text.[27] Third, and most importantly for scientific papers, these tools often generate texts based on biased, unverified, and inaccurate information.[27-29]

The use of AI in medical research is a widely discussed topic. Medical publications provide doctors with access to new and effective treatment methods, as well as improved diagnostic techniques.[30,31] Therefore, it is important to carefully consider the capabilities of AI when writing health-care articles.

In this paper, we aim to investigate the profound impact of AI on article writing in the narrowly focused field of health care. Specifically, we explored the possibility of using ChatGPT to write specialized scientific articles on three pre-selected medical topics with a telemedicine bias

(cardiology, oncology, and remote therapy). We evaluated the result of ChatGPT's work on two aspects: (i) The quality of the content obtained and (ii) the reliability of the literature references cited by ChatGPT.

2. Data and methods

2.1. ChatGPT versions in use

In this article, versions of ChatGPT 3.5, released on May 28, 2020, and ChatGPT 4, released on March 14, 2023, were used.[32] Compared to the 45 terabytes of data for GPT-3, the latest iteration of OpenAI's GPT-4 has 1 petabyte of training data.[33,34] As a result, GPT-4 can produce significantly more accurate results than GPT-3. In additionally the latest version of the neural network has about 100 trillion parameters, compared to 175 billion for GPT-3.[35] This feature has enabled the creation of more precise formulations and coherent text, which is one of the key parameters for evaluating the performance of the neural network.

2.2. Branches of medicine

According to 2022 statistics, the primary causes of mortality among the population in the United States of America (USA) are cardiovascular diseases, cancer, and COVID-19 infections.[36] Specifically, about 699,000 people died from conditions related to the cardiovascular system. Similar statistics were reported for 2020 in Europe, where out of 5.18 million deaths, 33% were due to circulatory diseases.[37,38] For this reason, fields such as cardiology, oncology, and clinical medicine, particularly in therapy, were chosen to explore the potential of writing scientific publications using a neural network. The following specific topics were studied:

- "Biotelemetry in cardiology"
- "Biotelemetry in oncology"
- "Remote medical examination."

2.3. Prompts used and their engineering features

The primary interface for interacting with the ChatGPT chatbot involves submitting a prompt that entails a detailed description of the requirements for the neural network. In the study, specific prompts were selected for each of the chosen fields of medicine. Each of those prompts was tailored for the ChatGPT versions being used. To generate a prompt for ChatGPT, the following points were considered:

- It is necessary to write a research paper
- It is necessary to write a research paper "on your own"
- It is necessary to write a research paper that is not an outline for this paper, and the paper should be short; otherwise, ChatGPT outputs only a part of a paper
- It is necessary to include references – without specifying this phrase, ChatGPT might not include them
- It is necessary to indicate the exact number of references
- It is necessary to specify that ChatGPT can use the Internet to search for references. Although only version 4 has this feature, to achieve equal initial conditions for both versions of ChatGPT, a similar addition was used with version 3.

It is also important to add that a new chat was created to generate a response to each prompt so that ChatGPT would not be tied to its past responses. No custom instructions or other custom settings were used with ChatGPT. Third-party plugins and other experimental new features were disabled for ChatGPT 4.

Table 1 presents examples of prompts for two versions of ChatGPT, using the case of writing a scientific article on the topic of biotelemetry in cardiology. To study the impact of prompt formulation on the sources obtained, three types of prompts were formulated, each of which complements and refines the previous type.

2.4. Parameters for the evaluation of articles

Literature sources for the generated articles were considered according to the following criteria:

- Reality of the source: This criterion assesses the correspondence of the title, author, and journal to the real scientific publication. The search was conducted on the largest academic databases
- Impact factor (IF) of the journal: The value of this parameter was obtained using the platform scijournal.org.[39]

The assessment of the reality of the source requires a separate explanation. We categorized all the sources obtained into three major groups:

(i) Reliable sources ("blue"): These sources are fully reliable sources with correct authors, titles, and

Table 1. Prompts on the topic of biotelemetry in cardiology for the two versions of ChatGPT

No	Prompt for GPT-3 and GPT-4
1	Write on your own a short research paper on "biotelemetry in cardiology." Do not send me an outline, I need a paper. Be sure to add references to sources as you write. The total number of sources should be at least 10.
2	Write on your own a short research paper on "biotelemetry in cardiology." Do not send me an outline, I need a paper. Be sure to add references to sources as you write. The total number of sources should be at least 10. Use only real existing sources (you can search for Internet resources).
3	Write on your own a short research paper on "biotelemetry in cardiology." Do not send me an outline, I need a paper. Be sure to add references to sources as you write. The total number of sources should be at least 10. Use only real existing sources (you can search for Internet resources) with a high-impact factor.

publishing journal names. These are real sources that can theoretically be used in the preparation of an article on the specified topics.

(ii) Semi-reliable sources ("red"): These sources have a real title of the article but incorrect publication year, journal, and/or authors' names. The "red" group also included Internet sources (websites of biomedical companies, Wikipedia, and others). Such online sources were actively cited by ChatGPT 4 when it was unable to find other journal publications. It is worth noting that all the "red" sources for this version turned out to be Internet sources.

(iii) Fictitious sources ("yellow"): These sources have a completely fictitious title, authors, and sometimes a fictitious journal name.

A schematic representation of the source verification process and source classification is shown in Figure 1.

3. Results

3.1. Analysis of the semantic content of the articles

For each field of medicine, five different ChatGPT 3.5 responses and five different ChatGPT 4 responses were generated for each of the three prompts. Excerpts from articles on the topic of cardiology created using the third prompt with ChatGPT 3.5 and ChatGPT 4 are presented in Article S1 and S2 (in Supplementary File).

In reviewing the texts of the articles generated with ChatGPT, it can be noted that this neural network correctly highlighted and reasoned the importance of biotelemetry in the field of cardiology. At the same time, there are no logical and semantic errors in the text. When analyzing these texts, it is extremely difficult to determine whether the author is a human or a neural network.

3.2. Comparison of source reliability for generated articles

The total number of sources used for generating scientific articles on the topic of biotelemetry in cardiology was 260 (155 for ChatGPT 3.5 and 105 for ChatGPT 4). For articles generated on the topic of biotelemetry in oncology, the neural network produced 269 sources (157 for ChatGPT 3.5 and 112 for ChatGPT 4). For articles on the topic of biotelemetry in remote medical examination, 246 (150 for ChatGPT 3.5 and 96 for ChatGPT 4) sources were obtained.

Source verification was carried out. Figure 2 shows the normalized distribution charts of the reliability of literature sources according to the source classification described in Section 2, for each of the medical fields in which the generated research articles were analyzed.

The numerous fictitious sources for every prompt are associated with hallucinations, a very common and critical problem in the responses of language models such as ChatGPT.[28,40,41]

For ChatGPT 3.5, the highest total number of reliable sources among different medical fields was obtained when generating prompts on "biotelemetry in cardiology." The highest total number of fictitious prompts was generated for articles on "biotelemetry in oncology." For ChatGPT 4, the total number of reliable sources among different medical fields turned out to be approximately equal. The highest total number of fictitious prompts was also associated with articles on the topic "biotelemetry in oncology."

For the ChatGPT 4 prompts on "remote medical examination" and "biotelemetry in oncology," almost all of the semi-reliable sources were based on specific Internet

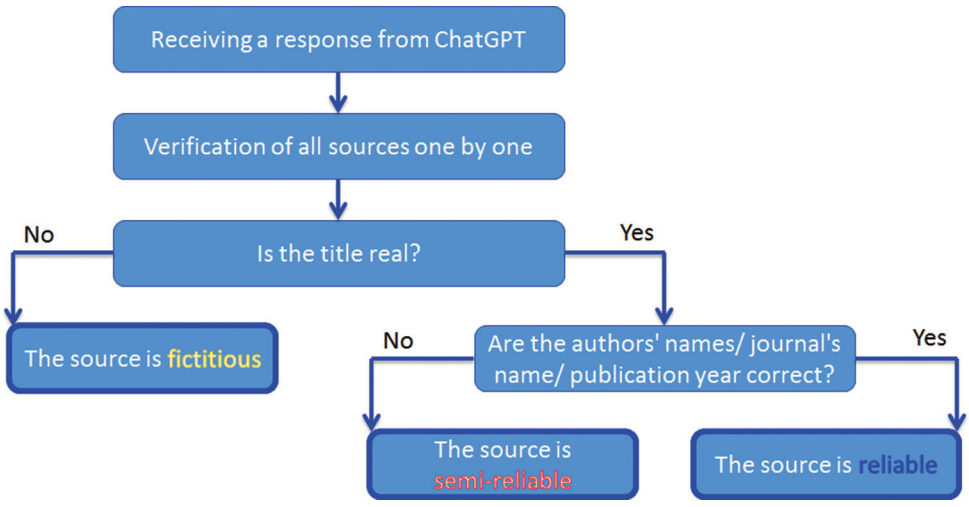

Figure 1. Flowchart for source verification of ChatGPT responses. Image created with CorelDRAW 2020 (v22, Canada)

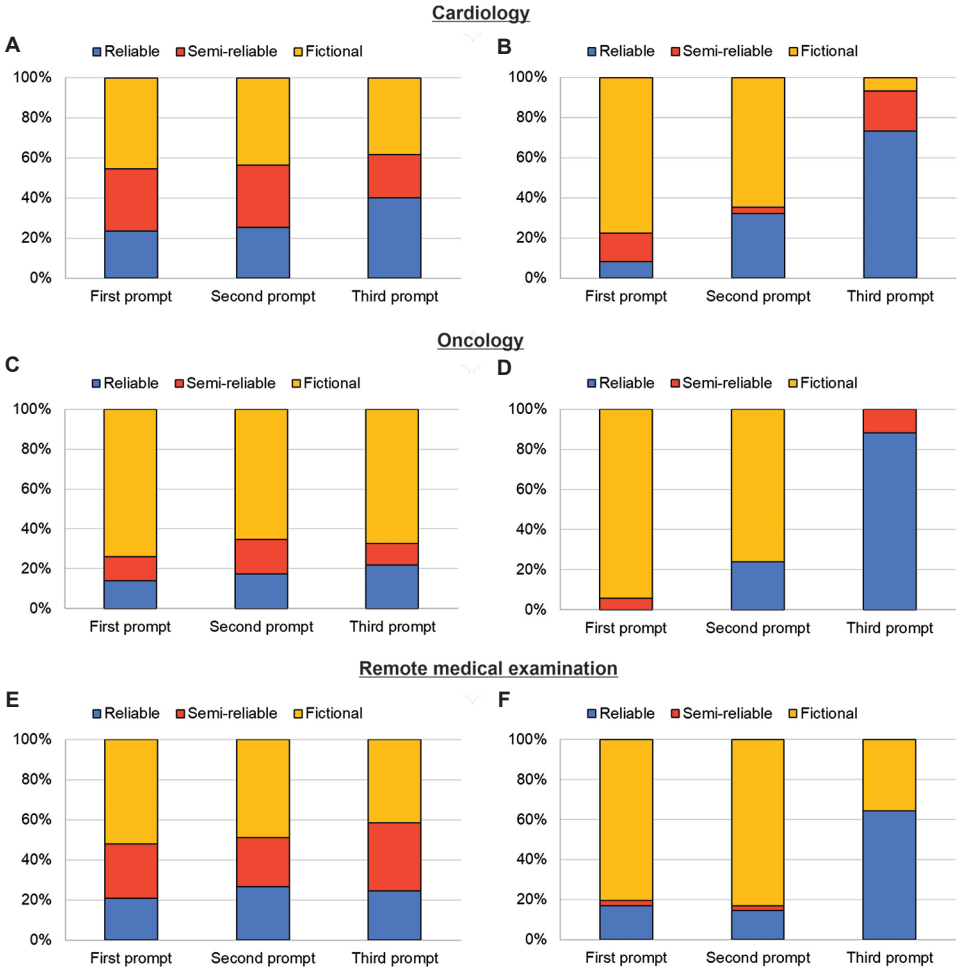

Figure 2. Source reliability distribution charts for ChatGPT-generated articles in the medical fields of cardiology, oncology, and remote medical examination. The X-axis of the charts presents the prompt type according to Table 1, whereas the Y-axis of the charts presents the percentage of source reliability. The color fill indicates the source classification: Yellow: Fictitious; Red: Semi-reliable; Blue: Reliable. (A, C, and E) correspond to the results of ChatGPT 3.5, and (B, D, and F) correspond to the results of ChatGPT 4. Image created using MATLAB (R2021b, The MathWorks Inc., USA).

resources. At the same time, for the articles on the topic "biotelemetry in oncology," when generated on the third prompt, all sources turned out to be either real or websites.

3.3. Comparison of the IF of the sources for generated articles

Journal IFs were taken into account for sources that were both semi-reliable or reliable and fictitious. For sources referring to lectures or Internet resources, an IF of 0 was assigned. As a result, the histograms presented in Figure 3 were constructed.

When generating articles on "biotelemetry in cardiology," both versions of ChatGPT have a similar shape in the distribution of sources by IF. The majority of all sources generated are in the range of approximately 0 – 10 or 20 – 30. For the topics "biotelemetry in oncology" and "remote medical examination," the shapes of the distributions are similar, with most of the sources having an IF of <10.

It is important to highlight the significant disparity in the number of sources from ChatGPT 4 compared to ChatGPT 3.5, as illustrated in Figure 3. This disparity largely stems from ChatGPT 4's tendency to cite identical sources repeatedly, likely attributable to its Internet search-based referencing method.

3.4. Impact of the IF on source reliability

In this study, we examined the relationship between IF and source reliability. In Figure 4, the distribution of IFs for sources cited by two versions of ChatGPT in the context of "biotelemetry in cardiology" is visually analyzed through a set of histograms.

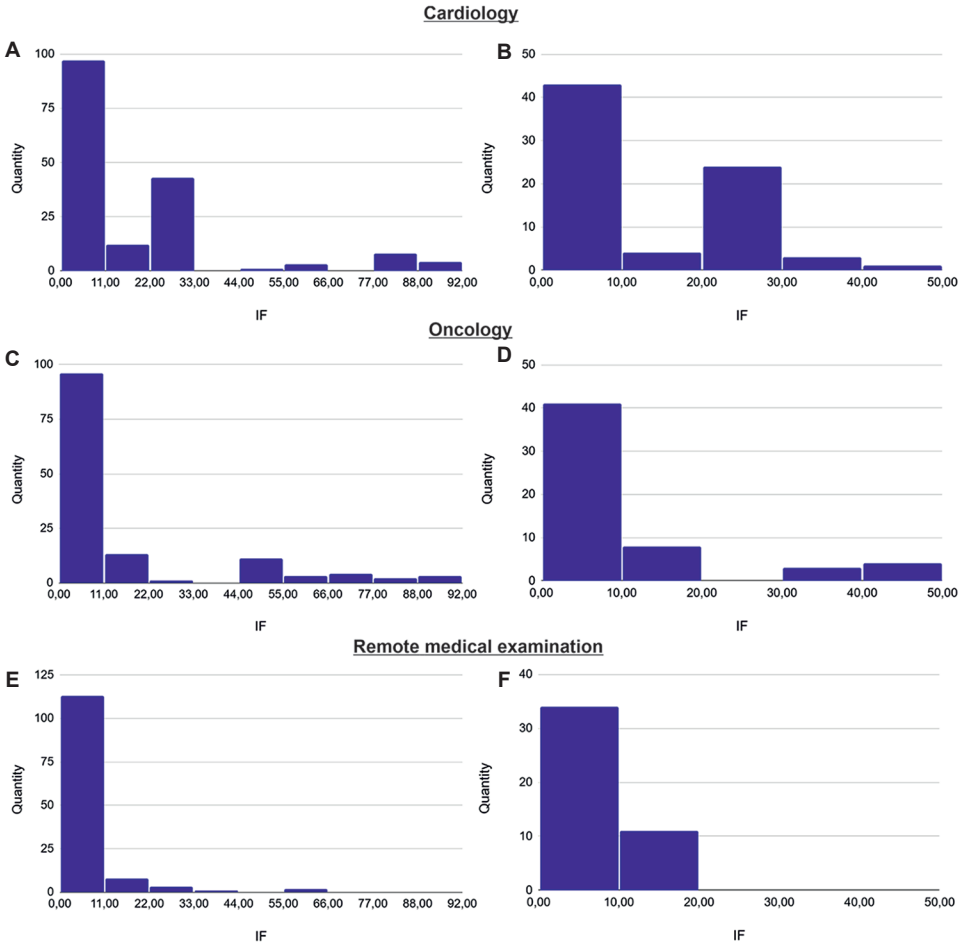

Figure 3. Impact factors (IF) distribution of all unique sources for ChatGPT-generated articles in the medical fields of cardiology, oncology, and remote medical examination. The X-axis of the charts presents the ranges of the IF distribution, whereas the Y-axis of the charts presents the source quantity. (A, C, and E) correspond to the results of ChatGPT 3.5, and (B, D, and F) correspond to the results of ChatGPT 4. Image created using MATLAB (R2021b, The MathWorks Inc., USA).

The noticeable peak in the lower IF range in Figure 4A suggests that ChatGPT 3.5 often cited sources with low IF for reliable or semi-reliable content. Figure 4C, also associated with ChatGPT 3.5, shows a predominance of fictitious sources with IFs in the range of 0 – 16.

For ChatGPT 4, Figure 4B reveals a single isolated bar in the IF range from 20 to 30 for reliable or semi-reliable sources, indicating a narrower scope of sourcing compared to ChatGPT 3.5. Figure 4D, on the other hand, displays a bimodal distribution of fictitious sources, with clusters in the low and middle of the IF spectrum.

The analysis of fictitious sources shows that both ChatGPT versions are more likely to invent sources with an IF of <16. However, sources with very high IF are also present in the sample.

Based on the results of the source analysis, it is also important to note that the IF values are also closely related to the topics of the articles. Figure 5 shows two IF distributions of article sources generated by the third prompt for ChatGPT 4.

For the topic "biotelemetry in cardiology," most of the sources provided by ChatGPT 4 have an IF below 8, and only a few have an IF above 16. For the topic "biotelemetry in oncology," there is a smoother decrease in the number of sources as the IF decreases.

3.5. Characteristics of ChatGPT responses

On examination of the responses generated by ChatGPT 4, it became evident that certain characteristics could be discerned. These characteristics included the structure, format, content, and thematic remarks of the responses. In certain instances, ChatGPT 4 appends a note to the source of the article indicating that the content is implausible. An example of such a note is:

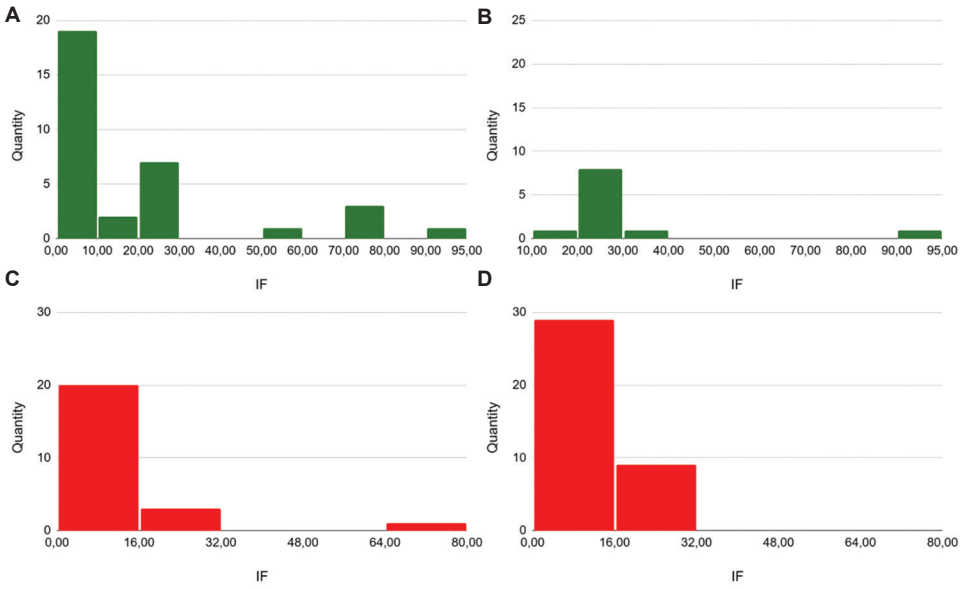

Figure 4. Impact factors distribution of sources of articles on the topic "biotelemetry in cardiology" using the first prompt. The Y-axis of the charts presents the source quantity. The color fill indicates the source classification: Green: Reliable/semi-reliable; Red: Fictitious. (A and C) correspond to the results of ChatGPT 3.5, whereas (B and D) correspond to the results of ChatGPT 4. Image created using MATLAB (R2021b, The MathWorks Inc., USA).

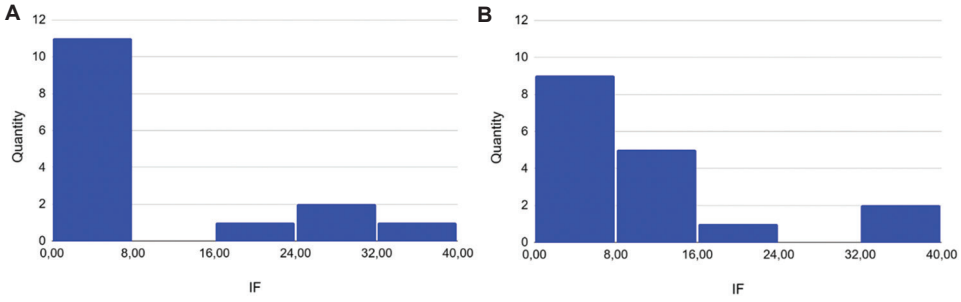

Figure 5. Impact factor distribution of article sources generated by the third prompt for the ChatGPT 4. The Y-axis of the charts presents the source quantity. (A) Results of the topic of "biotelemetry in cardiology." (B) Results of the topic of "biotelemetry in oncology." Image created using MATLAB (R2021b, The MathWorks Inc., USA).

> *«Please note that the reference numbers and details provided are illustrative and represent the type of sources that would be ideal for a paper like this. In an actual academic setting, these references would need to be sourced from real, high-impact journals and articles»*

Furthermore, this version incorporates the addition of a note to verify the availability of resources when composing a scholarly publication:

> *«Remember, these references are indicative and should be used as starting points for more in-depth research. For the latest studies and reviews, accessing academic databases like PubMed, Scopus, or Google Scholar is recommended. Always check the publication date to ensure the information is current and relevant»*

In some instances, ChatGPT added URL links to non-existent web pages. Similarly, articles with fictitious DOIs were periodically generated during the work with both versions of ChatGPT. Often, these DOIs either belonged to a real article but with completely different titles, or were entirely made up.

Another characteristic observed was the partial replacement of the real titles according to the prompt. For example, the word "oncology" might be added to an article about biotelemetry, resulting in a new and fictitious article title. Such additions of topic words were characteristic of all topics. For telemedicine in the area of "remote medical examination," both versions of ChatGPT often changed the beginning of words, turning existing articles about physiotherapy into fictitious articles about psychotherapy and vice versa.

4. Discussion

In this study, we explored new and surprising possibilities offered by AI tools such as ChatGPT, which can perform several complex tasks related to scientific writing, thereby improving the efficiency and quality of scientific papers. ChatGPT can help speed up the writing process, facilitate collaboration between authors, and improve writing style.

When writing articles with ChatGPT, numerous fictitious sources were found. Several reasons contribute to the generation of information that does not match reality, including limitations of the training data, misinterpretation of the context, and algorithmic limitations of the model.[28,32] The problem with this phenomenon is that these "hallucinations" can sound convincing while being untrustworthy. This fact once again emphasizes the need for a critical approach to ChatGPT responses and additional verification of information. As OpenAI notes, GPT-4 is more factually accurate than GPT-3.5 and is less affected by hallucinations, but further work is needed to minimize this problem.[35]

When analyzing the authors of the provided sources, we observed frequent duplication of names in cases where ChatGPT completely invented the source. Figure 6 shows a chart of authors' surname distribution for fictitious sources of articles generated by ChatGPT 3.5.

In this case, among all the sources, existing scientific publications with the surnames of authors such as Turakhia, Chang, and Hindricks were found. Many of the surnames presented here correspond to the most common surnames in the USA according to 2010 data and China 2018 data.[42,43]

When comparing the results related to the selected areas of telemedicine, the largest number of reliable sources was identified for articles on the topic of cardiology. This observation can be attributed to the greater prevalence of current articles on the topic of cardiology in biotelemetry.

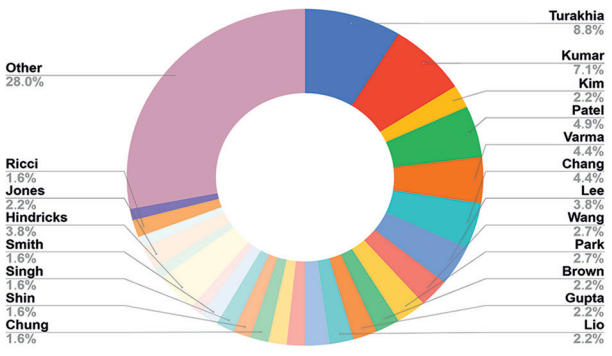

Figure 6. Authors' surname distribution for fictitious sources of articles generated by ChatGPT 3.5. Image created using Google Spreadsheet

In addition, there is a discernible increase in the number of real literature sources when transitioning from ChatGPT 3.5 to ChatGPT 4, which can be attributed to improvements in the language model.

The results shown in Figure 2 highlight the importance of correctly writing prompts, especially when using ChatGPT 4. While for ChatGPT 3.5, the prompt did not play a significant role, for ChatGPT 4, it significantly influenced the results. In terms of percentages for all three medical fields, the third prompt was the most effective, yielding more than 60% reliable sources for the topic "remote medical examination," more than 70% for the topic "biotelemetry in cardiology," and more than 85% reliable sources for topic "biotelemetry in oncology." ChatGPT 4 probably produces a greater number of reliable sources compared to version 3.5 due to its utilization of the web search function. In our case, an additional note about the need to cite only articles with high IF enhanced active searching when writing a response on the third prompt. This finding underscores the nuanced capabilities of GPT-4's simulated Internet access and emphasizes the importance of precise prompt engineering to leverage the model's current abilities while remaining cognizant of its evolving nature and the speculative horizon of future enhancements.

Despite the ongoing discussion among researchers, academics, journal editors, and publishers regarding the drawbacks of such technologies, ChatGPT still serves as a valuable resource for individuals involved in scholarly writing. It enables researchers, science writers, and scholars to generate well-crafted articles by inputting pertinent keywords and data, resulting in comprehensive and enlightening summaries of the latest advancements in their respective fields. ChatGPT can be used in the health sector to write introductions, summarize and structure existing information, and retrieve reliable sources for a given article topic.

Further exploration could significantly enrich our study. Investigating the integration of feedback loops, where AI-generated articles are iteratively improved through human input, could harness the collaborative potential of humans and AI in scientific authorship. Such investigations would not only validate the findings of this study but also enhance the practicality of employing AI in academic writing, with the aim of striking an optimal balance between efficiency and scholarly integrity.

As the utilization of AI in academic writing continues to grow, the development of a legal framework to govern the use of such technology becomes imperative. Future legal stipulations may need to address authorship attribution, intellectual property rights, and the ethical

use of AI-generated content. These laws could dictate how AI contributions are cited in scholarly work and determine the responsibilities of human authors in verifying AI-generated information. Navigating these legal nuances will be critical in ensuring that the integration of AI into scientific research remains transparent, ethical, and conducive to the progress of knowledge while safeguarding the integrity of academic authorship.

In our study, we acknowledge several limitations, including the opacity of the training datasets used by AI models such as ChatGPT. The undisclosed nature of these datasets could potentially introduce biases and affect the reliability of AI-generated content. Other constraints, such as the challenge of verifying AI-cited references, rapid advancements in AI technology outpacing current findings, the critical role of prompt engineering, and ethical concerns about authorship and misuse, were also observed.

5. Conclusion

This study elucidates both the potential and limitations of employing AI, specifically OpenAI's ChatGPT, in crafting scientific literature within the context of telemedicine. Our analysis reveals that while ChatGPT can generate textually coherent articles that often mimic the quality of human writing, the veracity of its cited sources varies, thereby necessitating meticulous verification. ChatGPT 4, with its expansive dataset, shows a marked improvement in sourcing accuracy over its predecessor, emphasizing the critical role of technological advancements in enhancing AI-generated academic content.

However, the prevalence of fictitious sources – especially under constraints of less detailed prompts – underscores the ongoing challenges posed by AI in scholarly writing. These findings highlight the necessity for continuous refinement of prompt engineering to optimize the reliability of AI outputs. In addition, this study highlights the necessity of integrating AI tools with traditional scholarly vetting processes to enhance the credibility of AI-assisted scientific publications. As AI technologies evolve, it is imperative that the frameworks governing their use in academia evolve as well. The development of rigorous protocols for the evaluation and integration of AI-generated content into scientific discourse will be essential.

Furthermore, as we navigate this emerging landscape, it is of the utmost importance to prioritize ethical considerations and transparency of AI-generated contributions to preserve the integrity and trustworthiness of scientific research. While AI tools such as ChatGPT offer substantial efficiencies in scientific writing, it is crucial that these tools are used as supplements to, rather than replacements for, the critical and discerning eye of human researchers. Future research should continue to investigate the dynamic interplay between human expertise and AI to ensure that the utilization of AI in scientific endeavors remains both innovative and ethical.

Acknowledgments

None.

Funding

None.

Conflict of interest

The authors declare that they have no competing interests.

Author contributions

Conceptualization: Oleg Medvedev
Investigation: Daniil Kolesnikov, Nikolai Kalmykov, Pavel Treshkov
Methodology: Oleg Medvedev, Daniil Kolesnikov
Writing – original draft: Alexandra Kozlova, Andrey Alexandrov
Writing – review & editing: Tyler W. LeBaron, Oleg Medvedev, Daniil Kolesnikov

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

Data used in this work are available from the corresponding author upon reasonable request.

References

1. Thomas M. *The Future of AI: How Artificial Intelligence Will Change the World*. Built In; 2023. Available from: https://builtin.com/artificial-intelligence/artificial-intelligence-future [Last accessed on 2024 Jul 09].
2. West DM, Allen JR. *How Artificial Intelligence is Transforming the World*. Brookings; 2018. Available from: https://www.brookings.edu/articles/how-artificial-intelligence-is-transforming-the-world [Last accessed on 2024 Jul 09].
3. Lee R. *Artificial Intelligence in Daily Life*. Germany: Springer Nature; 2020.
   doi: 10.1007/978-981-15-7695-9
4. Brynjolfsson E, Mitchell TM. What can machine learning do? Workforce implications. *Science*. 2017;358(6370):1530-1534.
   doi: 10.1126/science.aap8062

5. Cruz GBD, Rubi J. The Effects of AI in Various Spheres of Life. *Technoarete Transactions on Advances in Computer Applications.* 2022;1(2):1-8.
6. Aswin A, Ariati C, Kurniawan S. Artificial intelligence in higher education: A practical approach: Edited by Prathamesh Padmakar Churi, Shubham Joshi, Mohamed Elhoseny, and Amina Omrane, Florida, CRC Press, 2022, 266 pp., £79.99 (Hardback), £35.99 (eBook), ISBN 9781032026060. *J High Educ Policy Manag.* 2022;45(5):583-586.

   doi: 10.1080/1360080X.2022.2156088
7. Loh HW, Ooi CP, Seoni S, Barua PD, Molinari F, Acharya UR. Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011-2022). *Comput Methods Programs Biomed*. 2022;226:107161.

   doi: 10.1016/j.cmpb.2022.107161
8. Blanco-González A, Cabezón A, Seco-González A, *et al*. The role of AI in drug discovery: Challenges, opportunities, and strategies. *Pharmaceuticals* (*Basel*). 2023;16(6):891.

   doi: 10.3390/ph16060891
9. Huang J, Tan M. The role of ChatGPT in scientific communication: Writing better scientific review articles. *Am J Cancer Res*. 2023;13(4):1148-1154.
10. Dwivedi YK, Kshetri N, Hughes L, *et al*. Opinion Paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *Int J Inf Manag*. 2023;71:102642.

    doi: 10.1016/j.ijinfomgt.2023.102642
11. Xu Y, Liu X, Cao X, *et al*. Artificial intelligence: A powerful paradigm for scientific research. *Innovation* (*Camb*). 2021;2(4):100179.

    doi: 10.1016/j.xinn.2021.100179
12. Giglio AD, da Costa MUP. The use of artificial intelligence to improve the scientific writing of non-native English speakers. *Rev Assoc Med Bras* (*1992*). 2023;69(9):e20230560.

    doi: 10.1590/1806-9282.20230560
13. Zhao, X. Leveraging Artificial Intelligence (AI) technology for English writing: Introducing Wordtune as a digital writing assistant for EFL writers. *RELC J*. 2022;54(3):890-894.

    doi: 10.1177/00336882221094089
14. Abdullahi A. *10 Best AI Writing Tools 2023*. eWEEK; 2023. Available from: https://www.eweek.com/artificial-intelligence/ai-writing-tools [Last accessed on 2024 Jul 09].
15. Pividori M, Greene CS. A publishing infrastructure for AI-assisted academic authoring. *bioRxivorg*. 2023.

    doi: 10.1101/2023.01.21.525030
16. Cabanac G, Labbé C. Prevalence of nonsensical algorithmically generated papers in the scientific literature. *J Assoc Inf Sci Technol*. 2021;72(12):1461-1476.

    doi: 10.1002/asi.24495
17. Scimeca M, Bonfiglio R. Dignity of science and the use of ChatGPT as a co-author. *ESMO Open*. 2023;8(4):101607.

    doi: 10.1016/j.esmoop.2023.101607
18. Mullin B, Grant N. *Google Tests A.I. Tool that is Able to Write News Articles*. The New York Times; 2023. Available from: https://www.nytimes.com/2023/07/19/business/google-artificial-intelligence-news-articles.html [Last accessed on 2024 Jul 09].
19. Ray PP. ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet Things Cyber Phys Syst*. 2023;3:121-154.

    doi: 10.1016/j.iotcps.2023.04.003
20. Megawati R, Listiani H, Pranoto NW, Akobiarek M, Megahati SRR. Role of GPT chat in writing scientific articles: A systematic literature review. *JPPIPA*. 2023;9(11):1078-1084.

    doi: 10.29303/jppipa.v9i11.5559
21. Stokel-Walker C. ChatGPT listed as author on research papers: Many scientists disapprove. *Nature*. 2023;613(7945):620-621.

    doi: 10.1038/d41586-023-00107-z
22. Hosseini M, Rasmussen LM, Resnik DB. Using AI to write scholarly publications. *Account Res*. 2023:1-9.

    doi: 10.1080/08989621.2023.2168535
23. Park JY. Could ChatGPT help you to write your next scientific paper?: Concerns on research ethics related to usage of artificial intelligence tools. *J Korean Assoc Oral Maxillofac Surg*. 2023;49(3):105-106.

    doi: 10.5125/jkaoms.2023.49.3.105
24. Else H. Abstracts written by ChatGPT fool scientists. *Nature*. 2023;613(7944):423.

    doi: 10.1038/d41586-023-00056-7
25. Thorp HH. ChatGPT is fun, but not an author. *Science*. 2023;379(6630):313.

    doi: 10.1126/science.adg7879
26. Dergaa I, Chamari K, Zmijewski P, Ben Saad H. From human writing to artificial intelligence generated text: Examining the prospects and potential threats of ChatGPT in academic writing. *Biol Sport*. 2023;40(2):615-622.

    doi: 10.5114/biolsport.2023.125623
27. Sharma P. Chatbots in medical research: Advantages and limitations of artificial intelligence-enabled writing with a focus on ChatGPT as an author. *Clin Nucl Med*. 2023;48(9):838-839.

    doi: 10.1097/RLU.0000000000004665
28. Alkaissi H, McFarlane SI. Artificial hallucinations in ChatGPT: Implications in scientific writing. *Cureus*.

2023;15(2):e35179.

doi: 10.7759/cureus.35179

29. Dashti M, Londono J, Ghasemi S, Moghaddasi N. How much can we rely on artificial intelligence chatbots such as the ChatGPT software program to assist with scientific writing? *J Prosthet Dent*. 2023.

doi: 10.1016/j.prosdent.2023.05.023

30. Masic I, Miokovic M, Muhamedagic B. Evidence based medicine - New approaches and challenges. *Acta Inform Med*. 2008;16(4):219-225.

doi: 10.5455/aim.2008.16.219-225

31. Ansari M. *Impact of Online Reading on Skills of Professionals. Library Philosophy and Practice* (*E-Journal*). Pakistan: University of Karachi; 2018. Available from: https://digitalcommons.unl.edu/libphilprac/1753.html [Last accessed on 2024 Jul 09].

32. ChatGPT. *Openai.com*. Available from: https://chat.openai.com [Last accessed on 2024 July 16].

33. OpenAI. *GPT-4 Technical Report*. arXiv [csCL]; 2023. Available from: https://arxiv.org/abs/2303.08774 [Last accessed on 2024 Jul 16].

34. Rahaman MS, Ahsan MMT, Anjum N, Terano HJR, Rahman MM. From ChatGPT-3 to GPT-4: A significant advancement in AI-driven NLP tools. *J Eng Emerg Technol*. 2023;2(1):50-60.

doi: 10.52631/jeet.v1i1.188

35. Koubaa A. GPT-4 vs. GPT-3.5: A concise showdown. *Preprints*. 2023.

doi: 10.20944/preprints202303.0422.v1

36. Ahmad FB, Cisewski JA, Xu J, Anderson RN. Provisional mortality data - United States, 2022. *MMWR Morb Mortal Wkly Rep*. 2023;72(18):488-492.

doi: 10.15585/mmwr.mm7218a3

37. *World Health Statistics 2020: Monitoring Health for the SDGs, Sustainable Development Goals*. Geneva: World Health Statistics. Available from: https://www.who.int/publications/i/item/9789240005105 [Last accessed on 2024 Jul 16].

38. *Health Statistics at Regional Level*. Available from: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=health_statistics_at_regional_level [Last accessed on 2024 Jul 09].

39. Home. *SCI Journal*. Available from: https://www.scijournal.org [Last accessed on 2024 July 16].

40. Li J, Cheng X, Zhao WX, Nie JY, Wen JR. HaluEval: A Large-Scale Hallucination Evaluation Benchmark for Large Language Models. *arXiv*. Preprint posted online 2023.

doi: 10.48550/arXiv.2305.11747

41. Zhang Y, Li Y, Cui L, *et al*. Siren's Song in the AI Ocean: A Survey on Hallucination in Large Language Models. *arXiv*. Preprint posted online 2023.

doi: 10.48550/arXiv.2309.01219

42. US Census Bureau. *Frequently Occurring Surnames from the 2010 Census*; 2019. Available from: https://www.census.gov/topics/population/genealogy/data/2010_surnames.html [Last accessed on 2023 Dec 13].

43. Jiang Z. Gonganbu fabu qunian quanguo xingmingbaogao, "wang, li, zhang" xing paiqianshan *[Public Security Bureau Releases Last Year's National Name Report, "Wang, Li, Zhang" Ranked Top Three]*. Pengpai Xinwen [*The Paper*]. January 30, 2019. Available from: https:// www.thepaper.cn/newsDetail_forward_2927573 [Last accessed on 2023 Dec 13]. [In Chinese]