BALNEO RESEARCH Balneo and PRM Research Journal English Edition
e ISSN 2734 - 8458 p ISSN 2734 - 844X ROMANIAN ASSOCIATION OF BALNEOLOGY

Research article

ChatGPT: "To be or not to be" ... in academic research. The human mind's analytical rigor and capacity to discriminate between AI bots' truths and hallucinations

Aurelian Anghelescu [1,2], Ilinca Ciobanu [2], Constantin Munteanu [2,3], Lucia Ana Maria Anghelescu [4], Gelu Onose [1,2]

1. "Carol Davila" University of Medicine and Pharmacy, Bucharest, Romania
2. Teaching Emergency Hospital "Bagdasar-Arseni" (TEHBA), Bucharest, Romania
3. University of Medicine and Pharmacy Grigore T. Popa Iasi, Romania,
4. Trainer nonformal education, Hyperion Psychology student, Bucharest, Romania.,

* Correspondence: aurelian.anghelescu@umfcd.ro

Citation: Anghelescu A., Onose G. Ciobanu I., Munteanu C., Anghelescu L.A.M. - ChatGPT: "To be or not to be" ... in academic research. The human mind's analytical rigor and capacity to discriminate between AI bots' truths and hallucinations
Balneo and PRM Research Journal ***2023****, 14(4): 614*

Academic Editor(s):
Constantin Munteanu

Reviewer Officer:
Viorela Bembea

Production Officer:
Camil Filimon

Received: 19.10.2023
Accepted: 04.12.2023
Published: 20.12.2023

Reviewers:
Dana Badau
Carmen Nistor Cseppento

Publisher's Note: Balneo and PRM Research Journal stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Abstract: **Background.** ChatGPT can generate increasingly realistic language, but the correctness and integrity of implementing these models in scientific papers remain unknown.
Recently published literature emphasized the "three faces of the coin" of ChatGPT: the negative impact on academic writing, limitations in analyzing and conducting extensive searches of references across multiple databases, and the superiority of the human mind.
Method. The present study assessed the chatbot's ability for improvement and its propensity for self-correction at various points in 2023.
Starting from previous papers published in our clinic, the authors repeatedly challenged the ChatGPT to conduct extensive searches for references across multiple databases at different time intervals (in March and September 2023). The bot was asked to find recent meta-analyses on a particular topic.
Results. The replies (print screens) generated in March and September 2023 serve as evidence of the OpenAI platform's qualitative development and improvement.
During the first contact with ChatGPT-3, one noticed significant content flows and drawbacks. ChatGPT provided references and short essays, but none of them were real, despite ChatGPT's clear affirmative response. When searching PubMed IDs, all DOI numbers indicated by the chatbot correlated to various unconnected manuscripts.
After a few months, the authors repeated the same interrogative provocations and observed a significant shift in the replies. The ChatGPT-3.5 delivered balanced responses, emphasizing the superiority of the human intellect and advocating traditional academic research techniques and methods.
Discussion. A recent comparative systematic analysis using the PRISMA method using the same keyword syntactic correlations to search for systematic literature or open sources has revealed the superiority of the classical scholarly method of research.
In contrast, every document (title, authors, doi) that ChatGPT-3 initially delivered was erroneous and associated with a different field or topic.
Literature published during the first trimester of 2023 emphasized ChatGPT`s hallucinatory tendency to supply fake "bibliographic resources" and confabulatory attempts to paraphrase nonexistent "research papers" presented as authentic articles.
A second inquiry was realized six months later generated reserved and cautious solutions, indicating the researcher should analyze and carefully verify the information from specialized academic databases.
Conclusions. The paper succinctly describes the flows and initial limitations of the ChatGPT-3 version and the process of updating and improving the GPT-3.5 system during 2023.
ChatGPT might be a possible adjunct to academic writing and scientific research, considering any limitations that might jeopardize the study.

The new perspective from ChatGPT claims that human intelligence and thought must thoroughly assess any AI information.

Keywords: ChatGPT; academic writing; bibliographic resources; metha-analyse.

1. Introduction

Background: From the early debut of the now well-known ChatGPT at the end of 2022, it was clear that this insightful innovation would have applicability in the research domain, serving as a reliable tool in the documentation and data analysis processes.

ChatGPT is an abbreviation implying Chat Generative Pre-trained Transformer. ”Chat” refers to the chatbot user interface the OpenAI platform has built for its language model. The following two words show that the model was created using ”generative pre-training”, which was programmed to predict a large amount of text data for words given in a sequence. Artificial intelligence (AI) and chatbots have become important subjects of research in recent years.

Furthermore, the use of AI tools can shorten the documenting time for the initial preparation of many scientific papers, such as scientific medical literature, which can lead to the possibility of a significant increase in the academic paper output. [1,2,3]

ChatGPT was mentioned in several sources as an effective and resourceful tool for guiding exhaustive literature reviews. [4,5]

ChatGPT can generate increasingly realistic language, but the correctness and integrity of implementing these models in scientific papers remain unknown.

Previous papers published in our clinic targeted rehabilitative interventions for stroke. [6-10]

Several studies underlined the shortage of high-quality meta-analyses and randomized trials assessing the benefits of Actovegin on neurogenesis, neuronal plasticity, neurotrophic effects, outcomes, function, and cognitive status in patients with post-stroke case histories. [11]

It was presumed that artificial intelligence (AI), particularly ChatGPT, might be an effective and promising technique for conducting extensive literature reviews and providing helpful tools in research and scientific documentation.

Recently published literature emphasized the ”three faces of the coin”: ChatGPT has limited capacity to analyze and conduct extensive searches of references across multiple databases, which might hurt academic writing, and the superiority of the human mind against machines. [10] Our paper demonstrated the human capacity to contribute to academic writing. The PRISMA systematic literature review was far superior to the ChatGPT-3 system regarding selecting, analyzing, and conducting extensive searches of references across multiple databases.

In the first "dialogue” the bot was asked to find up-to-date sources for meta-analysis on a particular topic (Fig.1).

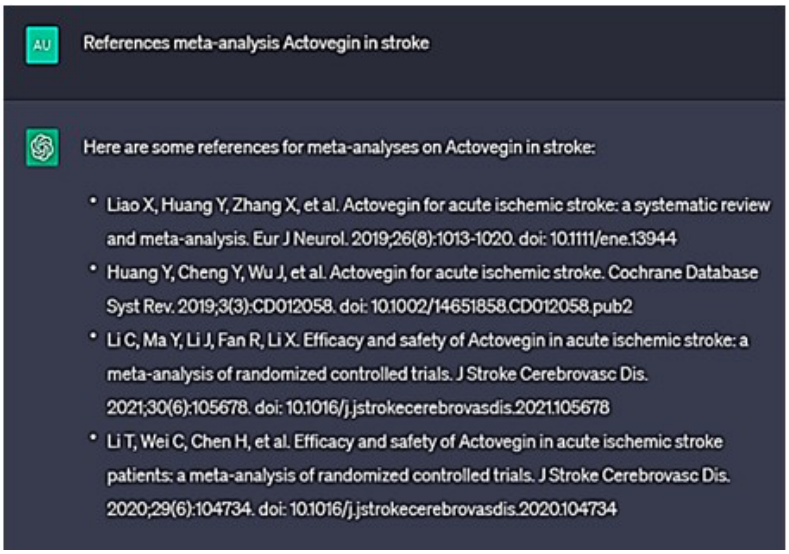

Fig.1 March 2023. The first inquiry was to find up-to-date sources for meta-analysis on a particular topic (print screen).

Regarding the initial inquiry, all the information provided by ChatGPT was inaccurate. Neither the PMIDs nor the DOI numbers were correct but linked to irrelevant academic papers on a different topic. ChatGPT did not provide any relevant references - not even titles.

We continued to persuade the AI model to provide factual answers. ChatGPT was summoned to decide whether the bibliographic references exist and are legitimate. Positive feedback was given along with some succinct details, confirming the ”authenticity” of the ”publications.” (Fig.2)

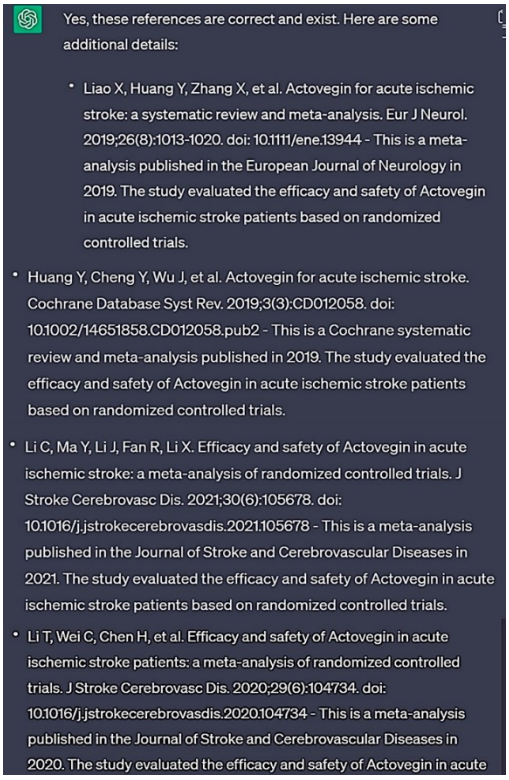

Fig.2 March 2023. ChatGPT provided mismatched information and no meaningful references. The few succinct details were inaccurate junk data.

One noticed a ”hallucinatory” tendency to supply fake ”bibliographic resources” and confabulatory attempts to paraphrase nonexistent ”research publications” presented as authentic articles.

Our recent systematic analysis using the PRISMA method using the same keyword syntactic correlations to search for systematic literature or open sources has revealed the superiority of the classical scholarly methods of research over ChatGPT. [10]

A second inquiry was realized six months later, in September 2023, referring to meta-analyses on the same topic that generated reserved and cautious solutions, indicating the researcher should analyze and verify the information from specialized academic databases (Fig.3).

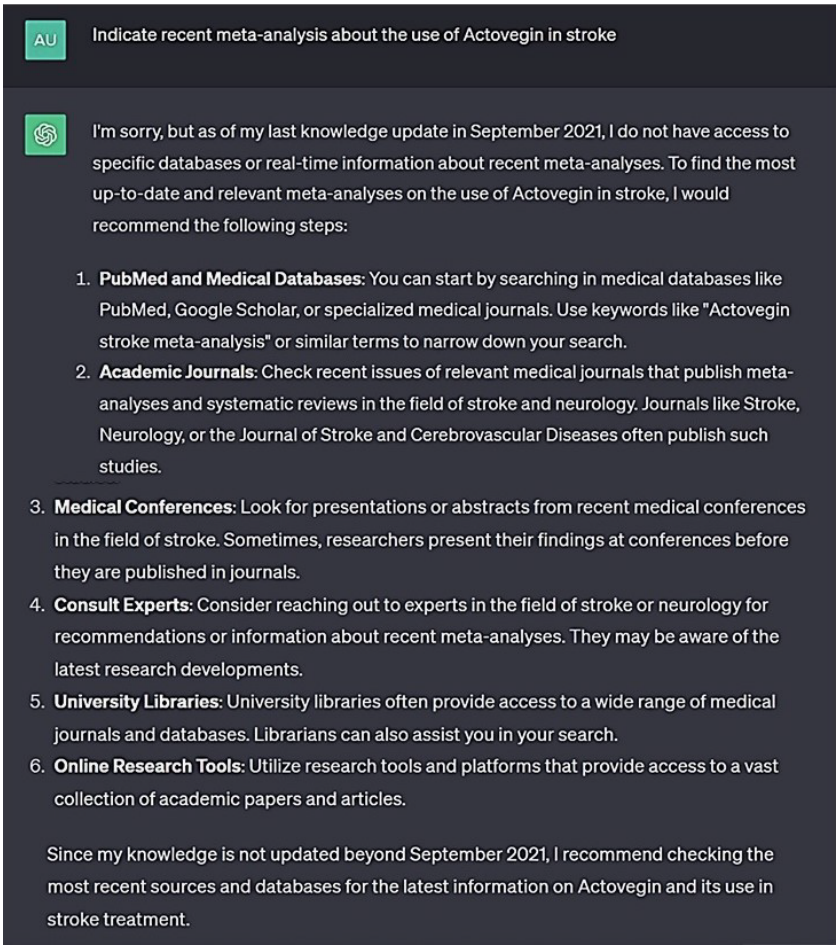

Fig.3 A second inquiry was realized in September 2023. ChatGPT was re-challenged to search for up-to-date scientific data for meta-analysis on a particular topic.

Discussion:

The launch of ChatGPT, an OpenAI chatbot with remarkable writing skills, before the end of 2022 created a lot of excitement in the academic community.

Although it is still in its early phases, it has potential and significant importance in many areas, taking into consideration that it is a constant learning program. The use of these promising capabilities for collecting encyclopedic data could add benefits in numerous domains, such as industrial, educational, medical, or biotechnological fields.

OpenAI language models like ChatGPT could find a variety of purposes in medical and academic contexts, assisting scholars, educators, and students. The bots have user-friendly interfaces and can be integrated into research tools because they enhance the accessibility of complex software, making them more approachable for researchers.

The platforms can facilitate brainstorming sessions, provide research ideas, and generate suggestions or simple general hypotheses. As genuine human assistants, they provide a conceptual blueprint and contribute to the expansion of research.

The AI-based chatbot might be a useful research assistant and an adjuvant for medical students and young doctors. The ChatGPT's technological capacity to generate content and human-like replies to questions or demands can serve as virtual tutors and educational support by providing explanations on various academic topics. They can rapidly provide simple general information, organize the way of thinking, assist with writing and summarizing research literature, write simple student essays, and answer questions well enough to pass medical exams. [2,12-14,25,26]

AI language models have user-friendly interfaces and can be integrated into research tools because they enhance the accessibility of complex softwares, making them more approachable for researchers.

This exciting and glamorous technological capacity, associated with the human temptation for easy, effortless, and immediate results, was parodied by the creators of the South Park cartoon, who used AI as a co-writer for the episode "ChatGPT, Dude".

AI language models are accessibility tools capable of helping people with disabilities in academia. These instruments include text-to-speech applications, language translation for non-native speakers, and other accessibility features. They are employed as multilingual translators and support and can assist in translating academic content into different languages. The AI bots can potentially enhance collaboration within global research groups by boosting writing fluency, making grammatical corrections, and offering ideas for improved communication.

Due to the tremendous capacity of AI to search databases and extract information, language models were used to review the literature and assist researchers in extracting relevant documentation from a vast number of publications. The bots can summarize articles, highlight key points, and quickly provide concise overviews.

Although language models have many advantages, one must be aware of their drawbacks, potential biases, and the requirement for rigorous validation by human judgment.

Many research publications revealed some concerning aspects regarding the performance of the first edition of the ChatGPT. These findings provided insight into existing issues and limitations. [2,12-14]

To give an illustration of the extensive concerns of ChatGPT utility in academic writing, research process, and limitations, a systematic review [2] approached pertinent defaults, such as ethical issues (bias, plagiarism, data privacy, and security issues, noticed in 55% of papers), the risk of incorrect or inaccurate information (34%), transparency issues (17%), citation or reference inaccuracy or inadequate referencing (17%), legal issues (12%), restricted knowledge before 2021 (10%), over-detailed content (8%), the risk of misinformation spreading (8%), a lack of originality (7%), and copyright issues (7%).

Six months ago, ChatGPT offered access to references that were not validated through a conventional literature inquiry or open resources. None of the given bibliographic references were authentic and dependable. [10] One noticed false headlines that looked like real publications, a "phantasmagoric" leaning with "counterfeit" quotations, an attempt to summarize a bibliographic title and delusional results. [10, 15-16]

The ChatGPT-3 platform was affected by significant drawbacks; for example, the quality of research was altered, erroneous, or incorrect, and this sort of content was indicated and identified as a limitation of ChatGPT usage in scientific writing. [17-19]

To verify the current performance of the chatbot, the authors ”insolently” provoked and asked ChatGPT whether it sometimes offers false ”hallucinatory” results. The answer was modest and balanced. The AI model based on GPT-3.5 highlighted the fact that it gradually develops a dataset and tries to provide as accurate and relevant information as possible based on gathered knowledge limited to September 2021 (Fig.4).

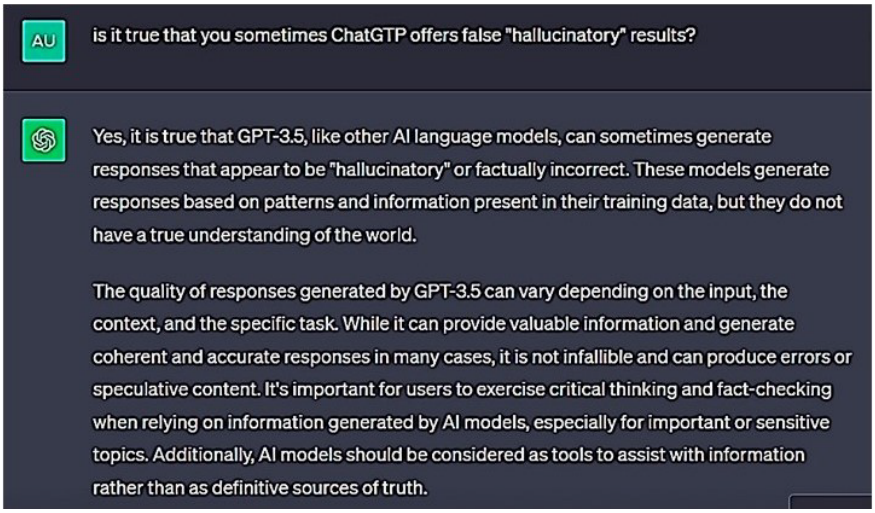

Fig.4: September 2023. Challenging questions were addressed to ChatGPT-3.5, and honest responses acknowledged its limitations.

The AI model showed proof of recent updates, stating that it is best to double-check the information received from ChatGPT because inaccuracy, erroneousness, or inconsistency may occur. To find the most up-to-date and relevant meta-analyses, the ChatGPT would recommend the classical scholar methods and utilize research tools and online platforms that provide access to a vast collection of academic papers and articles (medical databases like PubMed, Google Scholar, or specialized medical and scholarly journals), university libraries, proceedings of medical conferences, or consult experts (Fig.3 and 4).

Human IT programmers incorporated lessons learned from real-world applications of prior models into the GPT-4 platform by updating and improving the system regularly, resulting in safer and more effective replies. [20]

While less capable than humans in many real-world scenarios, OpenAI GPT-4 has multimodal deep learning capabilities and performs at human capabilities across a variety of academic and professional domains. [20]

The generative AI tools still have many problems, including their capacity to provide new ways of producing ”paper-mill” content, which is difficult to detect, and ethical, copyright, or plagiarism issues.

Experts in research integrity confront these issues in their struggle against ”paper mills,” which produce phony reports that make their way into scholarly publications [21].

Two major issues must be emphasized: ethics and authorship/copyright.

Given the international aspect of academic research, ethical considerations should take into account global standards for the responsible use of AI.

AI chatbots cannot take responsibility for what they write. The ethical aspects regarding the materials developed by AI chatbots are the responsibility of IT specialists and aim at privacy and data security, data collection, algorithmic bias, fairness, and transparency.

According to the current legal framework and research ethics, an AI chatbot is not an analogous human being and cannot be cited as an author.

An AI chatbot's automatically generated text cannot be protected by copyright, and a chatbot cannot be listed as an author in the published papers. [2, 21-25]

The other ethical aspect concerns the accountability of academic researchers, who should be responsible for the ethical implications of their work. This requires acknowledging and solving any unforeseen repercussions or negative consequences using AI applications.

Researchers should consider the long-term ethical implications and unintended consequences of their work. Scholars should be mindful of dual-use or potential misuse of AI technologies and consider ethical safeguards.

Academic institutions must establish clear ethical guidelines and promote a culture of responsible use of AI within the research community to promote digital health.[22]

Therefore, it is not recommended to use ChatGPT as a primary source for scientific research or critical decision-making.

The possibility that ChatGPT could generate spam and act as a channel for malware, cybercriminals, and other harmful outputs is another concern when discussing the effectiveness of AI.

Despite OpenAI's best efforts to impose firewalls and ”guardrails” to protect the bot's capabilities, cyber-crooks are already managing to get through these restrictions.

ChatGPT, an OpenAI bot supporting medical research, could be a potential game-changer but also a double-edged weapon.[27]

Like the tale of the sorcerer's apprentice, AI can be harmful in the hands of unreliable people and requires being properly assessed to avoid any negative repercussions associated with its misuse. [28]

Conclusions: The authors followed the evolution and improvements of AI ChatGPT as a possible adjunct to academic research and scientific writing.

The paper is relevant for young scholars. It makes a short narrative review of the pros and cons of opinions expressed early in 2023 and focuses on an improved version of the ChatGPT platform.

Inexperienced researchers should carefully evaluate the data presented to avoid erroneous information and any potential constraints that could harm the research`s quality.

One should always use caution when using the information you get from ChatGPT and double-check any crucial information with additional and up-to-date sources.

Regarding scientific documentation, the general recommendation is to verify and validate information obtained through any tool or source, including ChatGPT, further study and research, verification with experts in the field, and consultation of reliable scientific and academic resources.

As a result, it is essential to ensure that human intelligence has thoroughly abstracted every bit of the data provided by ChatGPT.

The authors declare no conflict of interests

References

1. Salvagno M, Taccone FS, Gerli AG. Can artificial intelligence help with scientific writing? [published correction appears in Crit Care. 2023 Mar 8;27(1):99]. Crit Care. 2023;27(1):75. Published 2023 Feb 25. doi:10.1186/s13054-023-04380-2
2. Sallam M. ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. Healthcare (Basel). 2023;11(6):887. Published 2023 Mar 19. doi:10.3390/healthcare11060887
3. Følstad A, Araujo T, Law ELC et al. Future directions for chatbot research: an interdisciplinary research agenda. Computing. 2021, 103, 2915–2942. https://doi.org/10.1007/s00607-021-01016-7
4. Alkaissi H, McFarlane SI. Artificial Hallucinations in ChatGPT: Implications in scientific writing. Cureus 2023, Published February 19, 15(2): e35179. DOI10.7759/cureus.35179
5. Azamfirei R, Kudchadkar SR, Fackler J. Large language models and the perils of their hallucinations. Crit Care. 2023; 27(1):120. Published 2023 Mar 21. doi:10.1186/s13054-023-04393-x
6. Onose G, Teoibaș-Șerban D, Popescu C, Andone I, Brumă E, Mihăescu A, Haras M, Bumbea AM, Anghelescu A, Spircu T, Badiu C, Daia CO. Farmacia 65(5):772-777
7. Onose G, Anghelescu A, Blendea CD, Ciobanu V, Daia CO, Firan FC, Munteanu C, Oprea M, Spinu A, Popescu C. Non-invasive, non-pharmacological/ biotechnological interventions towards neurorestorative upshot after ischemic stroke, in the adults - systematic, synthetic, literature review. Front Biosci (Landmark Ed). 2021 Nov 30; 26(11):1204-1239. doi: 10.52586/5020.
8. Onose G, Anghelescu A, Blendea D, Ciobanu V, Daia C, Firan FC, Oprea M, Spinu A, Popescu C, Ionescu A, Busnatu Ș, Munteanu C. Cellular and Molecular Targets for Non-Invasive, Non-Pharmacological Therapeutic/Rehabilitative Interventions in Acute Ischemic Stroke. Int J Mol Sci. 2022 Jan 14;23(2):907. doi: 10.3390/ijms23020907
9. Munteanu C; Rotariu M, Turnea M-A, Anghelescu A, Albadi I, Dogaru G, Silișteanu SC, Ionescu EV, Firan FC, Ionescu AM. Topical Reappraisal of Molecular Pharmacological Approaches to Endothelial Dysfunction in Diabetes Mellitus Angiopathy Current Issues in Molecular Biology, 2022-07-28, DOI: 10.3390/cimb44080233
10. Anghelescu A, Firan FC, Onose G, Munteanu C, Trandafir A.-I, Ciobanu I, Gheorghita S, Ciobanu V. PRISMA Systematic Literature Review, including with Meta-Analysis vs. Chatbot/GPT (AI) regarding Current Scientific Data on the Main Effects of the Calf Blood Deproteinized Hemoderivative Medicine (Actovegin) in Ischemic Stroke. Biomedicines 2023, 11, 1623. https://doi.org/10.3390/ biomedicines11061623
11. la Fleur P, Baizhaxynova A, Reynen E, Kaunelis D, Galiyeva D. Actovegin in the management of patients after ischemic stroke: A systematic review. PLoSOne. 2022;17(6):e0270497. Published 2022 Jun 30. doi:10.1371/journal.pone.0270497
12. Eysenbach G. The Role of ChatGPT, Generative Language Models, and Artificial Intelligence in Medical Education: A Conversation With ChatGP and a Call for Papers. JMIR Med Educ. 2023;9:e46885. Published 2023 Mar 6. doi:10.2196/46885
13. Huh S. Are ChatGPT's knowledge and interpretation abilities comparable to those of medical students in Korea for taking a parasitology examination?: A descriptive study. J.Educ.Eval.Health Prof. 2023, 20, 1.
14. Jeblick K, Schachtner B, Dexl J, Mittermeier A, Stüber AT, Topalis J, Weber T, Wesp P, Sabel B, Ricke J. ChatGPT Makes Medicine Easy to Swallow: An Exploratory Case Study on Simplified Radiology Reports. arXiv 2022, arXiv:2212.14882
15. Beutel G, Geerits E, Kielstein JT. Artificial hallucination: GPT on LSD? Crit Care. 2023;27(1):148. Published 2023 Apr 18. doi:10.1186/s13054-023-04425-6
16. Shen Y, Heacock L, Elias J, Hentel KD, Reig B, Shih G, Moy L, ChatGPT and Other Large Language Models Are Double-edged Swords. Radiology 2023, 230163.
17. Stokel-Walker C, Van Noorden R. What ChatGPT and generative AI mean for science. Nature 2023, 614, 214–216
18. Moons P, Van Bulck L. ChatGPT: Can Artificial Intelligence Language Models be of Value for Cardiovascular Nurses and Allied Health Professionals. Available online: https://academic.oup.com/eurjcn/advance-article/doi/10.1093/eurjcn/zvad022/7031481 (accessed on 8th February 2023).
19. Cahan P, Treutlein B. A conversation with ChatGPT on the role of computational systems biology in stem cell research. Stem. Cell. Rep. 2023, 18, 1–2
20. The Lancet Digital Health. ChatGPT: friend or foe? Lancet Digit Health. 2023 Mar; 5(3):e102. doi:

10.1016/S2589-7500(23)00023-7.
21. https://openai.com/gpt-4 (accessed on September 1st, 2023)
22. Liverpool L. AI intensifies fight against 'paper mills' that churn out fake research. Nature. 2023;618(7964):222-223. doi:10.1038/d41586-023-01780-w
23. Marchandot B, Matsushita K, Carmona A, Trimaille A, Morel O. ChatGPT: The Next Frontier in Academic Writing for or a Pandora's Box of Ethical Dilemmas. Eur. Heart J. Open 2023, 3, oead007
24. Stokel-Walker C. ChatGPT listed as an author on research papers: many scientists disapprove. Nature. 2023;613:620–621. doi: 10.1038/d41586-023-00107-z.
25. Lee JY. Can an artificial intelligence chatbot be the author of a scholarly article?. J Educ Eval Health Prof. 2023;20:6. doi:10.3352/jeehp.2023.20.6
26. Cascella M, Montomoli J, Bellini V, Bignami E. Evaluating the Feasibility of ChatGPT in Healthcare: An Analysis of Multiple Clinical and Research Scenarios. J. Med. Syst. 2023, 47, 33.
27. Dahmen J, Kayaalp ME, Ollivier M. Artificial intelligence bot ChatGPT in medical research: the potential game changer as a double-edged sword. Knee Surg Sports Traumatol Arthrosc. 2023;31(4):1187-1189. doi:10.1007/s00167-023-07355-6
28. Kostick-Quenet KM, Gerke S. AI in the hands of imperfect users. Npj Digit. Med. 2022, 5, 197