

## Medical ChatGPT – A systematic Meta-Review

Jan Egger<sup>1,2,3,4\*</sup>, Malik Sallam<sup>5,6</sup>, Gijs Luijten<sup>1,4</sup>, Christina Gsaxner<sup>1,4</sup>, Antonio Pepe<sup>4</sup>  
Jens Kleesiek<sup>1,3,7</sup>, Behrus Puladi<sup>8,9</sup>, Jianning Li<sup>1,3</sup>

<sup>1</sup>Institute for Artificial Intelligence in Medicine (IKIM), Essen University Hospital (AöR), Girardetstraße 2, 45131 Essen, Germany

<sup>2</sup>Center for Virtual and Extended Reality in Medicine (ZvRM), Essen University Hospital (AöR), Hufelandstraße 55, 45147 Essen, Germany

<sup>3</sup>Cancer Research Center Cologne Essen (CCCE), University Medicine Essen (AöR), Hufelandstraße 55, 45147 Essen, Germany

<sup>4</sup>Institute of Computer Graphics and Vision, Graz University of Technology, Inffeldgasse 16, 8010 Graz, Austria

<sup>5</sup>Department of Pathology, Microbiology, and Forensic Medicine, School of Medicine, University of Jordan, Amman 11942, Jordan

<sup>6</sup>Department of Clinical Laboratories and Forensic Medicine, Jordan University Hospital, Amman 11942, Jordan

<sup>7</sup>Department of Physics, TU Dortmund University, Otto-Hahn-Straße 4, 44227 Dortmund, Germany

<sup>8</sup>Department of Oral and Maxillofacial Surgery, University Hospital RWTH Aachen, Pauwelsstraße 30, 52074 Aachen, Germany

<sup>9</sup>Institute of Medical Informatics, University Hospital RWTH Aachen, Pauwelsstraße 30, 52074 Aachen, Germany

\*Corresponding author: [jan.egger@uk-essen.de](mailto:jan.egger@uk-essen.de) (J.E.)

**Abstract**– Since its release at the end of 2022, ChatGPT has seen a tremendous rise in attention, not only from the general public, but also from medical researchers and healthcare professionals. ChatGPT definitely changed the way we can communicate now with computers. We still remember the limitations of (voice) assistants, like Alexa or Siri, that were “overwhelmed” by a follow-up question after asking about the weather, not to mention even more complex questions, which they could not handle at all. ChatGPT and other Large Language Models (LLMs) turned that in the meantime upside down. They allow fluent and continuous conversations on a human-like level with very complex sentences and diffused in the meantime into all kinds of applications and areas. One area that was not spared from this development, is the medical domain. An indicator for this is the medical search engine PubMed, which comprises currently more than 36 million citations for biomedical literature from MEDLINE, life science journals, and online books. As of March 2024, the search term “ChatGPT” already returns over 2,700 results. In general, it takes some time, until reviews, and especially systematic reviews appear for a “new” topic or discovery. However, not for ChatGPT, and the additional search restriction to “systematic review” for article type under PubMed, returns still 31 contributions, as of March 19 2024. After filtering out non-systematic reviews from the returned results, 19 publications are included. In this meta-review, we want to take a closer look at these contributions on a higher level and explore the current evidence of ChatGPT in the medical domain, because systematic reviews belong to the highest form of knowledge in science.

**Keywords**– ChatGPT, Medicine, Healthcare, Meta-Review, Systematic Reviews, PubMed, LLMs

## Introduction

Comprehensive systematic reviews that synthesize the findings of multiple studies can provide a more robust and reliable understanding of ChatGPT's performance, its limitations, and the challenges that still need to be addressed, which are particularly useful in fields where research is abundant and rapidly evolving, such as the adoption of ChatGPT in healthcare. They allow researchers to quickly grasp the key aspects of a topic without having to sift through numerous individual studies. The increasing popularity of ChatGPT among healthcare professionals and researchers alike leads to a surge in research focusing on the evaluation of ChatGPT's performance in different application scenarios, such as medical consultation [1], research [2], education [3], or different medical specialties, such as neurology [4, 5], pediatric [6, 7], cosmetic surgery [8, 9] and dermatology [10, 11]. Each of these fields presents unique challenges and opportunities for ChatGPT. Understanding how ChatGPT performs in these different contexts is crucial for its continued development and improvement.

Due to the large amount of publications produced in a relatively short time, it is demanding for researchers to stay up-to-date with the latest development of ChatGPT in their specific domain and understand its challenges and limitations. Systematic reviews provide a quick and informative overview on the use of ChatGPT in a particular scenario or specialty [12, 13, 14, 15, 16, 17, 18, 19], which keep researchers updated with the field.

A meta-review on ChatGPT in healthcare synthesizes the findings of multiple systematic reviews, and therefore takes a broader view of the field, going beyond the scope of systematic reviews that focus only on specific scenarios or specialties. It aims to provide a concise and high-level summary of the current state of ChatGPT in the general healthcare sector [20], and help researchers, healthcare professionals, and policymakers understand the big picture in order to make sound decisions and policies. Future research directions can also be suggested.

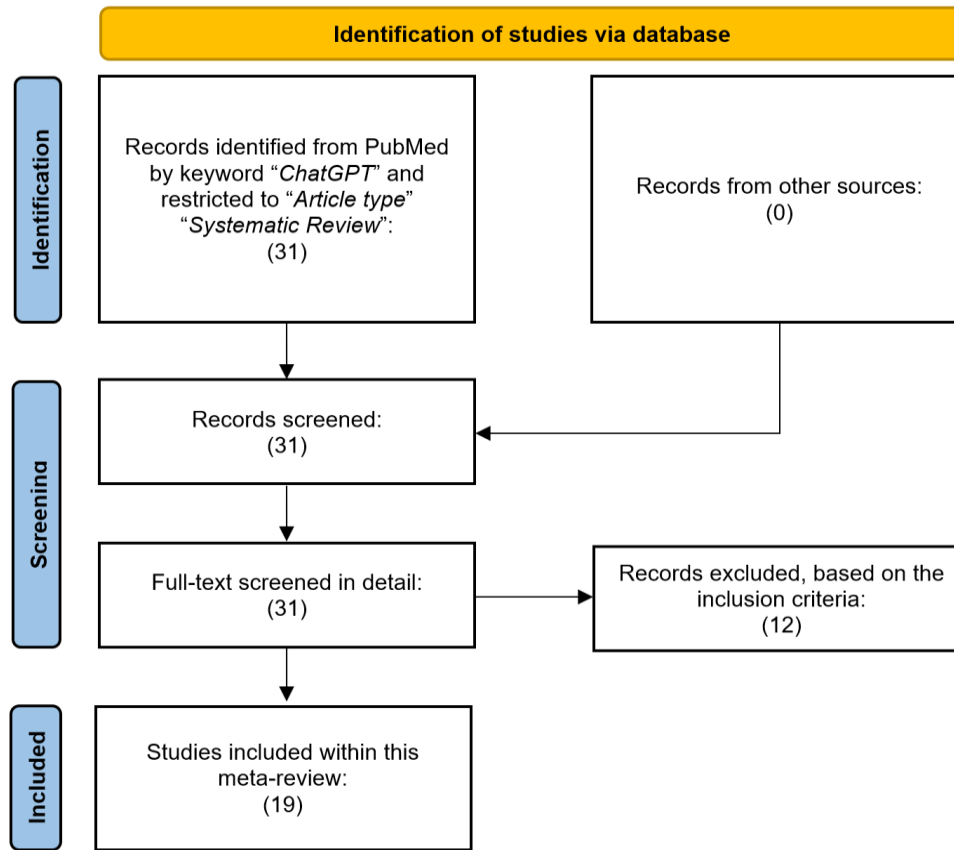
We position our work as a meta-review of systematic reviews of ChatGPT in healthcare, which look at systematic reviews across various application scenarios and medical specialties. The aim of the meta-review is to provide a concise yet comprehensive summary of the status quo of ChatGPT in the healthcare sector, assess the overall performance of ChatGPT, identify common trends and patterns, highlight key challenges and limitations, and point out areas where further research is needed. However, it is important to notice that, at this stage, the number of comprehensive systematic reviews is still small, and the majority of related publications remain to be high-level commentaries or small-scale evaluations [21], which highlights a gap in the literature and an opportunity for future research.

## Meta-Review of Systematic Reviews

### Methodology

As search strategy, we used the search string "*ChatGPT*" in PubMed and restricted the "*Article type*" to "*Systematic Review*", and ordered the resulting publications by date in ascending order: [https://pubmed.ncbi.nlm.nih.gov/?term="ChatGPT"&filter=pubt.systematicreview&sort=date&sort\\_order=asc](https://pubmed.ncbi.nlm.nih.gov/?term=)

Afterwards, the resulting publications were screened by two authors (J.E. and J.L.). The inclusion criteria were systematic reviews about ChatGPT in healthcare, also if they have been systematic scoping reviews, like [29]. The exclusion criteria were non-systematic reviews, publications that don't cover mainly a healthcare topic, comments, editorials and publications that were not written in English. The overall search strategy used in this systematic meta-review is shown in Diagram 2. and utilizes a slightly adapted version of the PRISMA Flow Diagram.



**Diagram 1.** The overall search strategy used in this systematic meta-review utilizing a slightly adapted version of the PRISMA Flow Diagram.

Table 1. summarizes the published systematic reviews about ChatGPT in the medical domain according to PubMed (status as of March 2024), which are arranged by epub (electronic publication) date in chronological order. The earliest and most representative publication is from Sallam M. [22], which was published already online on the 19th of March 2023, so exactly one year ago from the systematic search for this contribution. It covers a variety of publications that apply ChatGPT to the healthcare domain, whereby several of them are more on an abstract level, like comments or editorials, e.g. [24]. There were recently some critics that the publication does not fulfill all criteria for a systematic review [25] with a response from Sallam [26]. However, the outcome of the review showed strong evidence that the prospect of ChatGPT in healthcare is rather promising.

For some publications, the application of ChatGPT in a specific medical speciality or application scenario is discussed, including ischemic stroke [12], dentistry [13], pediatrics [14], neurosurgery [15, 18], patient care [16], medical education and examination [17, 18, 27] and scientific writing [19, 24]. In particular, Najafali D. et al. [19] discussed whether ChatGPT is able to independently carry out a systematic review on cosmetic surgery. It is concluded that the content generated by ChatGPT is not rigorously following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines for systematic reviews, and that in its current form, ChatGPT can be beneficial for exploring systematic review ideas. Chen T.J. et al. [24] conclude that ChatGPT and other artificial intelligence (AI)-based language tools could help non-native English speakers refine scientific writing in English. Two of the earliest systematic reviews on this topic, e.g., Sallam M. [22] and Tustumi T. [23], have discussed the general use of ChatGPT in healthcare, both of which have concluded that further improvements and evaluations are required for ChatGPT to be a reliable tool for healthcare professionals, while highlighting its promising prospect in this field. Furthermore, it is interesting to notice that researchers working on the topic are actively interacting with each other. Gupta R. [28], which is published as a 'letter to editor', commented on the publication from Najafali D. [19], adding that the ethical concerns revolving around using ChatGPT in scientific writing still remain to be addressed, which is not sufficiently covered by the original publication. Similar to Gupta R. [28], Moreno E. [25] is also a comment on a previous publication i.e., Sallam M. [22], which, while acknowledging the contribution of Sallam M. [22] to raising the awareness among healthcare professionals of the inevitable prevalence of AI tools, such as ChatGPT in medical practice, pointed out the flaws of the work concerning the conformity to the PRISMA guideline. A response to the comment was published by the original authors [26], which addressed the critics and, at the same time, acknowledged the interest of Moreno E. et al. in their work. These comments and responses not only reflect a healthy interaction among researchers but also show that applying ChatGPT or other AI tools in healthcare is an active area of research.

Generally speaking, it is useful and necessary to discuss how ChatGPT will impact healthcare in abstract terms, as in Sallam M. [22] and Tustumi T. [23]. However, practical implementation of ChatGPT in healthcare practice requires detailed quantitative evaluations and careful consideration of the unique needs of different healthcare specialties [21]. This comprehensive approach ensures that ChatGPT can provide accurate, reliable, and useful assistance. To elaborate:

**Abstract Discussion of ChatGPT's Impact on Healthcare:** The works of Sallam M. [22] and Tustumi T. [23] have initiated discussions on how ChatGPT could influence healthcare at a conceptual level, which involves exploring how ChatGPT can assist in diagnosing diseases, providing medical advice, automating administrative tasks, refining and accelerating scientific writing. These discussions are crucial as they provide a theoretical framework for understanding the potential benefits and challenges of integrating AI into healthcare.

**Need for Detailed Quantitative Evaluations:** While abstract discussions provide a theoretical understanding, they are often not enough for practical implementations. It's necessary to conduct detailed quantitative evaluations of ChatGPT in various medical specialties and application

scenarios. This could involve rigorous testing and validation of the AI model's performance in real-world clinical settings. For instance, how accurately can ChatGPT diagnose a specific condition compared to a human doctor? How does its performance vary across different specialties like cardiology, neurology, or psychiatry, and how reliable is the performance? What legal and ethical issues and concerns are exposed, which should be addressed from a legislative and regulative perspective? Answering these questions is the prerequisite for the practical deployment of an AI tool, as discussed in [21].

**Consideration of Uniqueness and Special Requirements in Different Healthcare Specialties:** Each medical specialty has its unique characteristics and requirements. For example, the type of language used, the complexity of cases, and the level of urgency can vary substantially. Therefore, for ChatGPT to be effectively implemented in healthcare practice, it must be tailored and optimized to meet these unique needs, which again requires individual evaluations as discussed above. This could be addressed by training the model on specialty-specific medical literature or fine-tuning it based on feedback from healthcare professionals in that specialty.

<b>Publications</b>	<b>Date (epub)</b>	<b>Num. Refs</b>	<b>Outcomes</b>
Sallam M. [22]	19 March 2023	96	The use of ChatGPT in healthcare education and research is inevitable. However, substantial research is needed to evaluate ChatGPT's performance and impact in practice.
Tustumi T. [23]	8 May 2023	25	Even though the use of ChatGPT in medicine is promising, its limitations must be carefully addressed. Responsible use of the tool must also be observed by healthcare professionals.
Anghelescu A. [12]	2 June 2023	43	ChatGPT is currently not fully capable of conducting critical systematic reviews. Inconsistency is observed by comparing the information extracted from a conventional system review and ChatGPT in terms of the treatment effects of Actovegin/AODEJIN on ischemic stroke.
Tiwari A. [13]	13 June 2023	21	In public health dentistry, ChatGPT cannot replace dentists, who not only make diagnoses but also analyze clinical findings, even though ChatGPT is a powerful tool for dentistry.
Balla Y. [14]	15 July 2023	42	The use of ChatGPT in pediatrics medicine has met with both challenges and opportunities.
Roman A. [15]	15 August 2023	32	ChatGPT has the potential to improve diagnosis, treatment and patient outcome in neurosurgeries, while a responsible and ethical use of the AI tool must be followed.
Levin G. [27]	21 August 2023	6	ChatGPT shows overall passing grade in medical examinations.
Garg R. K. [16]	11 September 2023	145	ChatGPT is helpful as an assistant in research and scientific writing, whereas issues concerning its accuracy, authorship and bias need to be further addressed.
Bečulić H. [18]	14 October 2023	49	More research is needed to understand the benefits and limitations of using AI tools such as ChatGPT in neurosurgery. Currently they should be used with caution.
Nikolas S. [36]	28 November 2023	41	ChatGPT depicts promising performance in various clinical tasks. Following studies should focus on expanding the application scope of ChatGPT and investigating the solutions for ethical issues.

Hiroj B. [35]	29 November 2023	51	ChatGPT can give appropriate answers to dentistry-related questions, which however need to be carefully evaluated before adoption.
Eleni M. [34]	10 December 2023	158	Smart digital tools like ChatGPT can be useful for mindfulness training and mitigate mental health issues.
Eyal K. [33]	25 December 2023	13	ChatGPT has shown potential in gastroenterology, especially in terms of helping with the interactions between patients and physicians. However, customization and ethical issues should be further addressed.
Hussain A. Y. [32]	4 January 2024	93	ChatGPT has shown transformative potential for a variety of medical applications. Responsible and ethical use of the tool must be observed.
Li J. [21]	15 January 2024	147	The development of ChatGPT in healthcare should aim for the respective medical speciality, where a unified evaluation system is still lacking.
Hugo C. T. [31]	19 January 2024	37	ChatGPT demonstrates potential in improving the efficiency and diagnosis accuracy of radiological workflows, as well as in assisting research. An interdisciplinary team involving AI experts, radiologists and policymakers are needed for the implementation of ChatGPT in radiological practice.
Wei Q. [30]	8 March 2024	86	The current scheme for evaluating ChatGPT's ability to answer medical inquiries are heterogeneous and lack conformity in reporting. Future studies are needed to objectively quantify ChatGPT's performance in medical question and answering.
Xiaojun X. [29]	15 March 2024	55	ChatGPT can potentially transform medical education. Special attention should be given to how ChatGPT affects the students and teachers.
Anusha S. [37]	13 March 2024	30	ChatGPT shows overall satisfactory performance in various medical examinations containing different questions types. Its performance drops as the difficulty of the question increases.

**Table 1.** List of published systematic reviews about ChatGPT in the medical domain according to PubMed (status as of March 19 2024); ordered by epub (electronic publication) date.

## Conclusion

Based on the systematic reviews about ChatGPT in the medical field in this contribution and the above discussions, we project that the number of systematic reviews will continue to increase, with more and more research on the evaluation of ChatGPT in a specific domain being published in the near future. However, it is important to note that, despite these reviews keeping track of the latest development of ChatGPT in healthcare, the quality and performance of the AI tool still depends on the underlying large language model, which needs to be continuously improved by researchers and practitioners in natural language processing (NLP). As healthcare professionals and researchers, our primary responsibility is to rigorously test the tool in our respective medical specialities in order to provide feedback on the actual capabilities of ChatGPT and expose its limitations and the ethical and legal concerns that arise. These in turn help NLP researchers to improve the language model and provide legislators the basis to formulate proper regulations. Summarized, we extracted the following three main challenges for ChatGPT in healthcare, but also medical NLPs in general, from our meta-review:

- **(1.) Dependence on the Underlying Language Model:** The quality and performance of ChatGPT are reliant on the underlying large language model, which is trained on vast amounts of text data and responsible for generating appropriate responses. Nevertheless, like any machine learning model, ChatGPT is prone to mistakes or failure given complex or ambiguous inputs. Therefore, continuous improvements by NLP researchers and practitioners are essential for fine-tuning the language model for a specific application. Furthermore, developing an artificial general intelligence (AGI) model with a wide spectrum of medical knowledge is a challenging but rewarding task.
- **(2.) Role of Healthcare Professionals and Researchers:** As users of ChatGPT in the medical field, healthcare professionals and researchers play an important role, who continuously provide feedback to NLP developers. Rigorous testing of the tool in their respective medical specialties is essential in providing valuable feedback on its actual capabilities in order to improve ChatGPT's performance. The expert-in-the-loop process involves not only identifying applications that ChatGPT is good at, e.g., providing medical information or assisting with patient communication, but also helps to expose the limitations of ChatGPT, such as hallucination or inability to provide accurate information for complex medical questions. The feedback from healthcare professionals not only help NLP researchers to improve the language model but also provide legislators with a basis to formulate proper regulations.
- **(3.) Ethical and Legal Concerns:** Using AI tools like ChatGPT in healthcare can also inevitably raise ethical and legal issues, such as patient privacy and data security. A clear guideline on when and how ChatGPT should be used in a healthcare setting is required. Healthcare professionals, policymakers and regulators should work jointly to address these concerns.



To conclude, the development and use of AI tools like ChatGPT in healthcare is a collaborative effort that should involve NLP researchers, healthcare professionals and legislators. Each group has a crucial role to play in ensuring that these tools are (in this order) safe, beneficial and effective for patient care.

## Acknowledgements

This work was supported by the REACT-EU project KITE (Plattform für KI-Translation Essen, <https://kite.ikim.nrw/>, EFRE-0801977) and FWF enFaced 2.0 (KLI 1044, <https://enfaced2.ikim.nrw/>). Behrus Puladi was funded by the Medical Faculty of the RWTH Aachen University in Germany as part of the Clinician Scientist Program. Furthermore, we acknowledge the Center for Virtual and Extended Reality in Medicine (ZvRM, <https://zvrm.ume.de/>) of the University Hospital in Essen, Germany.

## Disclaimer

For some parts of the paper, Microsoft's Bing Chat was used to generate hints using customized prompts.

## References

- [1] Howard, A., Hope, W. and Gerada, A., 2023. ChatGPT and antimicrobial advice: the end of the consulting infection doctor?. *The Lancet Infectious Diseases*, 23(4), pp.405-406.
- [2] Hosseini, M. and Horbach, S.P., 2023. Fighting reviewer fatigue or amplifying bias? Considerations and recommendations for use of ChatGPT and other Large Language Models in scholarly peer review. *Research Integrity and Peer Review*, 8(1), p.4.
- [3] O'Connor, S., 2022. Open artificial intelligence platforms in nursing education: Tools for academic progress or abuse?. *Nurse education in practice*, 66, pp.103537-103537.
- [4] Madadi, Y., Delsoz, M., Lao, P.A., Fong, J.W., Hollingsworth, T.J., Kahook, M.Y. and Yousefi, S., 2023. ChatGPT assisting diagnosis of neuro-ophthalmology diseases based on case reports. *medRxiv*.
- [5] Chen, T.C., Kaminski, E., Koduri, L., Singer, A., Singer, J., Couldwell, M., Delashaw, J., Dumont, A. and Wang, A., 2023. Chat GPT as a neuro-score calculator: analysis of a large language model's performance on various neurological exam grading scales. *World neurosurgery*, 179, pp.e342-e347.
- [6] Wei, Q., Wang, Y., Yao, Z., Cui, Y., Wei, B., Li, T. and Xu, X., 2023. Evaluation of ChatGPT's performance in providing treatment recommendations for pediatric diseases. *Pediatric Discovery*, 1(3), p.e42.
- [7] Corsello, A. and Santangelo, A., 2023. May Artificial Intelligence Influence Future Pediatric Research?—The Case of ChatGPT. *Children*, 10(4), p.757.
- [8] Humar, P., Asaad, M., Bengur, F.B. and Nguyen, V., 2023. ChatGPT is equivalent to first year plastic surgery residents: evaluation of ChatGPT on the plastic surgery in-service exam. *Aesthetic Surgery Journal*, p.sjad130.
- [9] Xie, Y., Seth, I., Hunter-Smith, D.J., Rozen, W.M., Ross, R. and Lee, M., 2023. Aesthetic surgery advice and counseling from artificial intelligence: a rhinoplasty consultation with ChatGPT. *Aesthetic Plastic Surgery*, pp.1-9.
- [10] Joly-Chevrier, M., Nguyen, A.X.L., Lesko-Krleza, M. and Lefrançois, P., 2023. Performance of ChatGPT on a practice dermatology board certification examination. *Journal of cutaneous medicine and surgery*, 27(4), pp.407-409.
- [11] Ferreira, A.L., Chu, B., Grant-Kels, J.M., Ogunleye, T. and Lipoff, J.B., 2023. Evaluation of ChatGPT dermatology responses to common patient queries. *JMIR dermatology*, 6, p.e49280.

- [12] Anghelescu A, Firan FC, Onose G, Munteanu C, Trandafir AI, 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 Jun 2;11(6):1623.
- [13] Tiwari A, Kumar A, Jain S, Dhull KS, Sajjanar A, Puthenkandathil R, Paiwal K, Singh R, Sajjanar A. Implications of ChatGPT in Public Health Dentistry: A Systematic Review. *Cureus*. 2023 Jun 13;15(6).
- [14] Balla Y, Tirunagari S, Windridge D. Pediatrics in Artificial Intelligence Era: A Systematic Review on Challenges, Opportunities, and Explainability. *Indian Pediatrics*. 2023 Jul;60(7):561-9.
- [15] Roman A, Al-Sharif L, Gharyani MA. The Expanding Role of ChatGPT (Chat-Generative Pre-Trained Transformer) in Neurosurgery: A Systematic Review of Literature and Conceptual Framework. *Cureus*. 2023 Aug 15;15(8).
- [16] Garg RK, Urs VL, Agarwal AA, Chaudhary SK, Paliwal V, Kar SK. Exploring the role of ChatGPT in patient care (diagnosis and treatment) and medical research: A systematic review. *Health Promotion Perspectives*. 2023;13(3):183.
- [17] Weidener L, Fischer M. Teaching AI Ethics in Medical Education: A Scoping Review of Current Literature and Practices. *Perspectives on Medical Education*. 2023;12(1):399.
- [18] Bečulić H, Begagić E, Skomorac R, Mašović A, Selimović E, Pojskić M. ChatGPT's contributions to the evolution of neurosurgical practice and education: a systematic review of benefits, concerns and limitations. *Med Glas (Zenica)*. 2024 Feb 1;21(1).
- [19] Najafali D, Camacho JM, Reiche E, Galbraith LG, Morrison SD, Dorafshar AH. Truth or lies? The pitfalls and limitations of ChatGPT in systematic review creation. *Aesthetic Surgery Journal*. 2023 Aug;43(8):NP654-5.
- [20] Egger, J., Gsaxner, C., Pepe, A., Pomykala, K.L., Jonske, F., Kurz, M., Li, J. and Kleesiek, J., 2022. Medical deep learning—A systematic meta-review. *Computer methods and programs in biomedicine*, 221, p.106874.
- [21] Li, J., Dada, A., Puladi, B., Kleesiek, J. and Egger, J., 2024. ChatGPT in healthcare: a taxonomy and systematic review. *Computer Methods and Programs in Biomedicine*, p.108013.
- [22] Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. In *Healthcare* 2023 Mar 19 (Vol. 11, No. 6, p. 887). MDPI.
- [23] Tustum F, Andreollo NA, Aguilar-Nascimento JE. Future of the language models in healthcare: the role of chatGPT. *ABCD. Arquivos Brasileiros de Cirurgia Digestiva (São Paulo)*. 2023 May 8;36:e1727.
- [24] Chen TJ. ChatGPT and other artificial intelligence applications speed up scientific writing. *Journal of the Chinese Medical Association*. 2023 Apr 1;86(4):351-3.
- [25] Moreno E, Alvarez-Lozada LA, Arrambide-Garza FJ, Quiroga-Garza A, Elizondo-Omaña RE. Comment on Sallam, M. ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. *Healthcare* 2023, 11, 887. *Healthcare*. 2023; 11(21):2819.
- [26] Sallam M. Reply to Moreno et al. Comment on “Sallam, M. ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. *Healthcare* 2023, 11, 887”. *Healthcare*. 2023; 11(22):2955.
- [27] Levin G, Horesh N, Brezinov Y, Meyer R. Performance of ChatGPT in medical examinations: A systematic review and a meta-analysis. *BJOG: An International Journal of Obstetrics & Gynaecology*. 2023 Aug 21.
- [28] Gupta R, Bagdady K, Mailey BA. Response to: Truth or lies? The pitfalls and limitations of ChatGPT in systematic review creation. *Aesthetic Surgery Journal*. 2023 Aug;43(8):NP656-7.
- [29] Xu X, Chen Y, Miao J. Opportunities, challenges, and future directions of large language models, including ChatGPT in medical education: a systematic scoping review. *J Educ Eval Health Prof*. 2024;21:6. doi:10.3352/jeehp.2024.21.6

- [30] Wei, Q., Yao, Z., Cui, Y., Wei, B., Jin, Z. and Xu, X., 2024. Evaluation of ChatGPT-generated medical responses: A systematic review and meta-analysis. *Journal of Biomedical Informatics*, p.104620.
- [31] Temperley, H.C., O'Sullivan, N.J., Mac Curtain, B.M., Corr, A., Meaney, J.F., Kelly, M.E. and Brennan, I., 2024. Current applications and future potential of ChatGPT in radiology: A systematic review. *Journal of Medical Imaging and Radiation Oncology*.
- [32] Younis, H.A., Eisa, T.A.E., Nasser, M., Sahib, T.M., Noor, A.A., Alyasiri, O.M., Salisu, S., Hayder, I.M. and Younis, H.A., 2024. A Systematic Review and Meta-Analysis of Artificial Intelligence Tools in Medicine and Healthcare: Applications, Considerations, Limitations, Motivation and Challenges. *Diagnostics*, *14*(1), p.109.
- [33] Klang, E., Sourosh, A., Nadkarni, G.N., Sharif, K. and Lahat, A., 2023. Evaluating the role of ChatGPT in gastroenterology: a comprehensive systematic review of applications, benefits, and limitations. *Therapeutic Advances in Gastroenterology*, *16*, p.17562848231218618.
- [34] Mitsea, E., Drigas, A. and Skianis, C., 2023. Digitally Assisted Mindfulness in Training Self-Regulation Skills for Sustainable Mental Health: A Systematic Review. *Behavioral Sciences*, *13*(12), p.1008.
- [35] Bagde, H., Dhopte, A., Alam, M.K. and Basri, R., 2023. A systematic review and meta-analysis on ChatGPT and its utilization in medical and dental research. *Heliyon*.
- [36] Schopow, N., Osterhoff, G. and Baur, D., 2023. Applications of the Natural Language Processing Tool ChatGPT in Clinical Practice: Comparative Study and Augmented Systematic Review. *JMIR Medical Informatics*, *11*, p.e48933.
- [37] Sumbal, A., Sumbal, R. and Amir, A., 2024. Can ChatGPT-3.5 Pass a Medical Exam? A Systematic Review of ChatGPT's Performance in Academic Testing. *Journal of Medical Education and Curricular Development*, *11*, p.23821205241238641.