

# Editorial

## Evaluating Generative AI-Based Models in Healthcare

**Mousa Al-Akhras** <sup>1</sup>✉

Since their release, generative AI models are increasingly used in many domains. Healthcare is one of the major domains where the incorporation of generative AI models in general and with focus on ChatGPT as the primary leading tool could be particularly transformative to the sector.

Recognizing the urgency and significance of this research area, the Jordan Medical Journal (JMJ) has launched this special issue titled "Evaluating Generative AI-Based Models in Healthcare". This special issue aims to address existing challenges and highlights novel perspectives in this rapidly evolving field to advance the collective understanding of generative AI's role in healthcare. This special issue highlights key prioritized areas in the assessment of ChatGPT utilization in healthcare.

The leading article entitled "Envisioning the Future of ChatGPT in Healthcare: Insights and Recommendations from a Systematic Identification of Influential Research and a Call for Papers" by Sallam et al. [1] paved the way for the call for papers of the remaining articles. The authors conducted a systematic search on Scopus, Web of Science, and Google Scholar to identify the top ChatGPT-related published records in healthcare with top articles denoted to be influential based on the citation count. This study [1] was based on a previous bibliometric analysis that identified the top ten ChatGPT-related healthcare published records over the period of a year (November 2022–November 2023) [2].

In the current study [1] the authors identified a total of 22 different articles discussing the use of ChatGPT in the healthcare sector. The authors defined future research areas regarding the utilization of ChatGPT in healthcare education, practice, and research based on the insights and recommendations from the top influential published records. Several important themes were identified by authors based on their search, including improving healthcare education, improving the efficiency of clinical processes, addressing ethical concerns, supporting research tasks, mitigating model's bias due to the possible bias in the training data, improving patient education and engagement, and developing standardized assessment protocols for ChatGPT utility in healthcare.

In the second article [3] entitled "Ensuring Security and Privacy in Healthcare Systems: A Review Exploring Challenges, Solutions, Future Trends" by Bala et al. the authors discussed the importance of ensuring security and offering privacy in healthcare systems in the context of AI integration in such systems. The authors discussed the belief that healthcare systems should emphasize providing medical attention to patients and pay less attention to the security aspects of the system. This was due to the historical misconception that healthcare systems would not be targeted by attackers.

The authors identified several factors contributing to the increasing vulnerabilities of the healthcare systems: (1) the increased use of interconnected technologies with their inherited vulnerabilities, (2) continuous monitoring of patients' well-being beyond the clinical setting leading to increased use of extra medical devices heightening the vulnerability of data breaches, (3) The increased use of mobile consumer electronics, such as cell phones making more challenging to safeguard health data from the risks associated with multipurpose gadgets, (4) Many healthcare organizations continue to employ outdated systems in various sectors, such as Windows XP, which has not received recent updates.

The authors discussed: (1) The detailed architecture of typical healthcare system components, (2) existing cyberattacks in the healthcare sector proposing solutions to combat these attacks. (3) the research community's challenges in ensuring security and privacy in healthcare systems. (4) The importance of AI in protecting medical data, protecting privacy, and monitoring patients. (5) The Open challenges and future research directions for addressing security and privacy issues in health systems.

---

<sup>1</sup> King Abdullah II School of Information Technology, The University of Jordan.

✉Corresponding author: [mousa.akhras@ju.edu.jo](mailto:mousa.akhras@ju.edu.jo)

The authors also emphasized that the integration of generative AI in healthcare introduces a range of cybersecurity challenges and ethical considerations. If not properly secured AI systems can be exploited in cyberattacks [4] or be manipulated to generate false data or misleading medical predictions, which could have severe negative consequences for patient care and privacy [5]. The proliferation of AI-driven technologies increases the surface area for potential breaches, as every point in an AI system—from data input to model training and output—can be a vulnerability [6]. Data bias and reliability were another source of concern raised by the authors. AI systems are only as good as the data they are trained on. If training data is flawed or biased, this can lead to inaccurate medical predictions and treatments, exacerbating health disparities [7].

In the third article [8] entitled “Evaluation of Artificial Intelligence-Based Chatbot Responses to Common Dermatological Queries” the authors aimed to investigate the utility of conversational AI models in addressing diagnostic challenges and treatment recommendations for common dermatological ailments.

A dataset of 22 cases vignettes of dermatological conditions was compiled, each case accompanied by three specific queries. These cases were presented to four distinct conversational AI models - ChatGPT 3.5, Google Gemini, Microsoft Copilot (GPT 4), and Perplexity.ai and responses were saved and evaluated by two expert dermatologists (>5 years experience post-graduation) independently using a 5-point Likert scale ranging from highly accurate (= 5) to inaccurate (= 1).

This study [8] pointed to the efforts of utilizing generative AI potential in dermatological healthcare education [9, 10]. The results indicate that the 4 used generative AI systems may produce different accuracy results to the same question due to their variations in underlying architecture, domain-specific fine-tuning, and intended application [11, 12, 13]. The authors pointed to several possible reasons, such as the specific algorithm, training data, and linguistic nuances employed [14, 15]. The study [8] findings underscore the limitations of the used generative models in accurate disease diagnosis. The authors recommended that the programs may be used as supplementary resources rather than primary diagnostic tools.

In the fourth article [16] entitled “Enhancing Security and Privacy in Healthcare with Generative Artificial Intelligence-Based Detection and Mitigation of Data Poisoning Attacks Software” the authors proposed stopping poisoning attacks by mixing federated learning, homomorphic encryption, and auto encoder-based anomaly identification without changing the original data.

In the fifth article [17] entitled “Evaluating ChatGPT's Role in Assessing Turkey's Medical Tourism Trends,” the authors evaluated the performance of ChatGPT as an expert in the field of medical tourism and to assess the current and future states of medical tourism in Turkey. The authors developed a questionnaire based on a study entitled “Bridging the gap of bibliometric analysis: the evolution, current state, and future directions of tourism research using ChatGPT” [18]. Questions were evaluated by 8 experts of different relevant backgrounds. Thirteen questions were asked to ChatGPT 4-o and questions were statistically analyzed. The reporting was conducted using the CLEAR tool and the METRICS checklist, which are prepared for artificial intelligence studies [19, 20].

The authors [17] reported that the general answers to ChatGPT 4-o were accurate, informative, and helpful in providing a good overview of medical tourism in Turkey. However, the responses lacked details, did not provide evidence-based information, and did not always address the nuances and cultural/social impacts of medical tourism. ChatGPT 4-o views the future of Turkey's medical tourism as shiny.

In the sixth article [21] entitled “Cross-Linguistic Evaluation of Generative AI Models for Diabetes and Endocrine Queries” the authors aimed to evaluate the performance of two genAI models, ChatGPT and Microsoft Copilot, in addressing endocrine and metabolic disorders queries in English and Arabic. Comparing responses generated in English and Arabic aimed to identify potential disparities in accuracy, relevance, and completeness.

During the evaluation, the METRICS checklist [19, 20] was used to evaluate the study. ChatGPT-4o outperformed Microsoft Copilot in all CLEAR components, but notable language-based disparities were evident. Addressing these limitations is crucial to ensure equitable access to endocrine care for non-English-speaking patients.

In the seventh (last) article [22] entitled “The Future of Pediatric Care: AI and ML as Catalysts for Change in Genetic Syndrome Management” The authors thoroughly investigate the potential impact of Artificial Intelligence (AI) and Machine Learning (ML) in pediatric healthcare and education, with a focus on genetic syndromes. Exploring several disorders highlights the significance of AI and ML in improving diagnostic accuracy, tailoring treatment approaches, and transforming educational methods. This review emphasizes the effect of incorporating AI and ML into pediatric practices for genomic analysis and adaptive education. These technologies help enhance comprehension of genetic disorders, allowing for personalized interventions customized to each child's specific

requirements. The review suggests that combining AI and ML significantly enhances diagnostic accuracy, treatment effectiveness, and educational results, thereby establishing higher benchmarks for pediatric care.

In conclusion, the JMJ special issue published 7 articles that cover several aspects of the use of generative AI in healthcare education, practice, and research. The leading article, based on a systematic search process, paved the way to highlight key prioritized areas of generative AI utilization in healthcare.

The second and fourth articles discussed several challenges related to securing healthcare systems and ensuring the privacy of patient's data. The third and sixth articles discussed the accuracy of prompting generative AI models in Dermatological Queries, and Diabetes and Endocrine Queries, respectively.

Medical tourism was covered in the fifth article while the use of generative AI in paediatric care was the subject of the seventh article.

JMJ releases this special issue to shed light on several important aspects of the use of generative AI models in the healthcare sector. More research in future articles can definitely define additional research directions that could achieve some goals that were unthinkable before.

## References:

1. Sallam, M., Al-Farajat, A. ., & Egger, J. (2024). Envisioning the Future of ChatGPT in Healthcare: Insights and Recommendations from a Systematic Identification of Influential Research and a Call for Papers. *Jordan Medical Journal*, 58(3). <https://doi.org/10.35516/jmj.v58i1.2285>
2. Sallam M. Bibliometric Top Ten Healthcare Related ChatGPT Publications in Scopus, Web of Science, and Google Scholar in the First ChatGPT Anniversary. *JMIR Preprints* 2023.
3. Bala, I. ., Pindoo, I. ., Mijwil, M. M. ., Abotaleb, M. ., & Yundong, W. . (2024). Ensuring Security and Privacy in Healthcare Systems: A Review Exploring Challenges, Solutions, Future Trends, and the Practical Applications of Artificial Intelligence. *Jordan Medical Journal*, 58(3). Retrieved from <https://journals.ju.edu.jo/index.php/JMJ/article/view/2527>
4. Yamin MM, Ullah M, Ullah H, Katt B. Weaponized AI for cyber attacks. *Journal of Information Security and Applications* 2021;57:102722...
5. Chen Y, Esmailzadeh P. Generative AI in Medical Practice: In-Depth Exploration of Privacy and Security Challenges. *J Med Internet Res* 2024;26:e53008.
6. Kaur R, Gabrijelčič D, Klobučar T. Artificial intelligence for cybersecurity: Literature review and future research directions. *Information Fusion* 2023;97:101804.
7. Alowais SA, Alghamdi SS, Alsuhebany N, Alqahtani T, Alshaya AI, Almohareb SN, et al. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. *BMC Medical Education* 2023;23(1):689.
8. Podder I., Pipil N., Dhabal A., Mondal S., Pienyii V. J. and H. Mondal. Evaluation of Artificial Intelligence-Based Chatbot Responses to Common Dermatological Queries. *Jordan Medical Journal* 2024, Supplement 1.
9. Dave M, Patel N. Artificial intelligence in healthcare and education. *Br Dent J.* 2023;234(10):761-764. doi: 10.1038/s41415-023-5845-2.
10. Tudor Car L, Dhinakaran DA, Kyaw BM, Kowatsch T, Joty S, Theng YL, Atun R. Conversational Agents in Health Care: Scoping Review and Conceptual Analysis. *J Med Internet Res.* 2020;22(8):e17158. doi: 10.2196/17158.
11. Agarwal M, Sharma P, Goswami A. Analysing the Applicability of ChatGPT, Gemini, and Copilot to Generate Reasoning-Based Multiple-Choice Questions in Medical Physiology. *Cureus.* 2023;15(6):e40977. doi: 10.7759/cureus.40977.
12. Meyer, J.G., Urbanowicz, R.J., Martin, P.C.N. et al. ChatGPT and large language models in academia: opportunities and challenges. *BioData Mining* 2023;16:20. doi: 10.1186/s13040-023-00339-9
13. Kumari A, Kumari A, Singh A, Singh SK, Juhi A, Dhanvijay AD, et al. Large Language Models in Hematology Case Solving: A Comparative Study of ChatGPT-3.5, Google Gemini, and Microsoft Copilot. *Cureus* 2023;15(8): e43861. doi:10.7759/cureus.43861.

14. Rahsepar AA, Tavakoli N, Kim GHJ, Hassani C, Abtin F, Bedayat A. How AI Responds to Common Lung Cancer Questions: ChatGPT vs Google Gemini. *Radiology*. 2023;307(5):e230922. doi: 10.1148/radiol.230922.
15. Fan X, Chao D, Zhang Z, Wang D, Li X, Tian F. Utilization of Self-Diagnosis Health Chatbots in Real-World Settings: Case Study. *J Med Internet Res*. 2021;23(1):e19928. doi: 10.2196/19928.
16. Mohialden Y. M., Salman S. A., Mijwil M. M, Hussien N. M., Aljanabi M., Abotaleb M., Dhoska K. and P. Mishra. Enhancing Security and Privacy in Healthcare with Generative Artificial Intelligence-Based Detection and Mitigation of Data Poisoning Attacks Software. *Jordan Medical Journal* 2024, Supplement 1.
17. AKPUR A. and K. ENES. Evaluating ChatGPT's Role in Assessing Turkey's Medical Tourism Trends. *Jordan Medical Journal* 2024, Supplement 1.
18. Shin, H. and J. Kang. (2023). Bridging the gap of bibliometric analysis: The evolution, current state, and future directions of tourism research using ChatGPT. *Journal of Hospitality and Tourism Management*, 57, 40-47.
19. Sallam, M., Barakat, M., & Sallam, M. (2023). Pilot Testing of a Tool to Standardize the Assessment of the Quality of Health Information Generated by Artificial Intelligence-Based Models. *Cureus*, 15(11). doi:10.7759/cureus.49373
20. Sallam, M., Barakat, M., & Sallam, M. (2024). A Preliminary Checklist (METRICS) to Standardize the Design and Reporting of Studies on Generative Artificial Intelligence-Based Models in Health Care Education and Practice: Development Study Involving a Literature Review. *Interactive journal of medical research*, 13. doi:10.2196/54704
21. Abbasi H., Al-Qudheeb M., Kheyami Z. A., Khalil R., Khamees N., Hijjawi O., Sallam M. and M. Barakat. Cross-Linguistic Evaluation of Generative AI Models for Diabetes and Endocrine Queries. *Jordan Medical Journal* 2024, Supplement 1.
22. Ghunaim, L., Ali Agha, A, Al-Samydai, A., & Aburjai, T. (2024). The Future of Pediatric Care: AI and ML as Catalysts for Change in Genetic Syndrome Management. *Jordan Medical Journal*, 58(4). <https://doi.org/10.35516/jmj.v58i4.2787>.