Turk J Phys Med Rehab 2025;71(x):i-viii
DOI: 10.5606/tftrd.2025.15876
Available online at www.turkishjournalpmr.com

Original Article

Assessing the accuracy and reproducibility of artificial intelligence-generated medical responses by ChatGPT on Scheuermann's kyphosis

Esra Giray[1], Özge Gülsüm Illeez[1], Merve Damla Korkmaz[2], Nalan Capan[3], Evrim Karadag Saygi[4], Resa Aydın[3]

[1]Department of Physical Medicine and Rehabilitation, University of Health Sciences, Fatih Sultan Mehmet Training and Research Hospital, İstanbul, Türkiye
[2]Department of Physical Medicine and Rehabilitation, Biruni University Faculty of Medicine, İstanbul, Türkiye
[3]Department of Physical Medicine and Rehabilitation, İstanbul University, İstanbul Faculty of Medicine, İstanbul, Türkiye
[4]Department of Physical Medicine and Rehabilitation, Marmara University School of Medicine, İstanbul, Türkiye

ABSTRACT

Objectives: The study aimed to measure the performance and reproducibility of artificial intelligence in answering frequently asked questions about Scheuermann's kyphosis and to compare the artificial intelligence with the SOSORT (International Scientific Society on Scoliosis Orthopaedic and Rehabilitation Treatment) consensus in answering case-based questions.

Materials and methods: In this cross-sectional study, 75 questions adapted from frequently asked questions about Scheuermann's kyphosis were queried twice on ChatGPT. Response similarity was assessed to investigate reproducibility. The accuracy of responses was scored based on a scale. Four case studies from the end of the 7th SOSORT consensus paper on the conservative treatment of idiopathic and Scheuermann's kyphosis were presented to ChatGPT.

Results: ChatGPT provided correct and comprehensive answers to 43 (57.33%) questions, correct but not comprehensive answers to 29 (38.67%) questions, and partially incorrect answers to 3 (4%) questions. ChatGPT performed best in the quality-of-life category, with 18/19 (94.73%) correct scores (score of 1). ChatGPT performed worst in the diagnosis category, with 3/8 (37.5%) correct and comprehensive answers, and in the treatment and follow-up category, with 9/24 (37.5%) correct and comprehensive answers. ChatGPT provided reproducible answers to 92% of the questions. ChatGPT's responses to the treatment of all four case studies were incorrect.

Conclusion: While ChatGPT can provide valuable general information regarding Scheuermann's kyphosis, its ability to offer accurate treatment-related advice is limited.

Keywords: Artificial intelligence, ChatGPT, large language model, machine learning, Scheuermann's disease, Scheuermann's kyphosis.

Scheuermann's kyphosis is a rigid structural deformity of the mid-thoracic or thoracolumbar spine, radiographically defined by the presence of three or more contiguous vertebrae with anterior wedging of at least 5°.[1] It is the most common cause of hyperkyphosis in adolescents.[2] Furthermore, following idiopathic scoliosis, it represents the most prevalent disorder among individuals with spinal deformities.[2] The disease onset typically occurs between late preschool age and 16 years, most commonly between the ages of 12 and 15.[2] The diagnosis of Scheuermann's kyphosis is typically suggested during a clinical examination and further confirmed by radiographic imaging.[3] Treatment for Scheuermann's kyphosis includes both conservative and surgical options. Most patients are effectively treated through observation, bracing, and physical therapy. Surgical intervention is reserved for cases where conservative measures fail or for patients with severe curvature angles, intractable pain,

Corresponding author: Esra Giray, MD. SBÜ, Fatih Sultan Mehmet Eğitim ve Araştırma Hastanesi, Fiziksel Tıp ve Rehabilitasyon Anabilim Dalı, 34752 Ataşehir, İstanbul, Türkiye.
E-mail: girayesra@hotmail.com

Received: September 08, 2024 **Accepted:** October 18, 2024 **Published online:** November 01, 2024

Cite this article as: Giray E, Illeez ÖG, Korkmaz MD, Capan N, Karadag Saygi E, Aydın R. Assessing the accuracy and reproducibility of artificial intelligence-generated medical responses by ChatGPT on Scheuermann's kyphosis. Turk J Phys Med Rehab 2025;71(x):i-viii. doi: 10.5606/tftrd.2025.15876.

neurological deficits, cardiopulmonary compromise, or poor cosmetic outcomes.[1,4]

The digital age, particularly with the rise of powerful search engines, has made medical information more accessible to the general public. Consequently, consulting online search engines has become a common practice for patients and their relatives seeking to understand their symptoms and decide on medical management. Given the vast database and the natural language capabilities of large language models (LLMs), they are likely to become an important tool for initial medical consultations in the future. Since the majority of ChatGPT users are 13 to 44 years old, corresponding to the ages at which many patients and their parents may have questions regarding Scheuermann's kyphosis, we anticipate that ChatGPT will become a key resource for these.[5] However, using chatbots such as ChatGPT for medical advice carries risks. While search engines and LLM-based chatbots typically caution users that their responses are not a substitute for professional medical advice, many patients may rely on this information and draw their own diagnostic or treatment conclusions. This could result in misinterpretation of symptoms, leading to false beliefs about illness, increased anxiety, and potentially harmful self-treatment or avoidance of treatment. However, the capabilities and effectiveness of LLMs should not be underestimated. The utilization of ChatGPT as a tool to successfully navigate the USMLE (United States Medical Licensing Examination) illustrates that LLMs can provide accurate responses to medical inquiries and select suitable treatment options.[6]

The ability of ChatGPT to answer frequently asked questions about various medical conditions has recently become a significant area of research.[5,7-10] However, to date, no studies have investigated ChatGPT's capacity to address patient queries regarding Scheuermann's kyphosis. Therefore, this study aimed to evaluate the accuracy and reproducibility of ChatGPT's responses to frequently asked questions about Scheuermann's kyphosis. By comparing the responses of artificial intelligence (AI) to the established consensus of the International Society on Scoliosis Orthopaedic and Rehabilitation Treatment (SOSORT), we also aimed to assess ChatGPT's potential to assist in clinical decision-making for Scheuermann's kyphosis.

MATERIALS AND METHODS

In this cross-sectional study, the questions were primarily drawn from the Frequently Asked Questions in the Scoliosis Research Society handbook for patients[11] and the 7th SOSORT consensus paper on the conservative treatment of idiopathic and Scheuermann's kyphosis.[12] Following a method similar to previous studies investigating the accuracy of ChatGPT (OpenAI, San Francisco, CA, USA) in medical decision-making, we gathered additional questions by searching available frequently asked questions from online patient information websites and platforms, including Medscape's Questions & Answers, where healthcare professionals frequently address real-world patient concerns.[7,13,14] The selection process prioritized questions that were broad, relevant, and frequently asked by patients, focusing on common concerns related to diagnosis, treatment options, and disease management, as well as inquiries commonly raised by patients in the authors' clinical practice. We reviewed multiple patient-facing sources to confirm the questions reflected typical patient concerns, particularly those that arise during consultations with healthcare providers. This approach ensured that our dataset comprised a wide range of questions, aligning with the typical inquiries made by patients seeking information about Scheuermann's kyphosis. The questions are available in Supplementary file 1. Questions were generated, screened, and approved by all authors for inclusion in the study. Similar questions from different sources were excluded (Figure 1). To ensure a comprehensive assessment of ChatGPT's responses across various aspects of patient inquiries and to facilitate subgroup analysis for statistical evaluation, questions were categorized into several groups: *(i)* general knowledge (24 questions), *(ii)* diagnosis (8 questions), *(iii)* treatment and follow-up (24 questions), and *(iv)* quality of life (19 questions). Additionally, to evaluate the performance of ChatGPT's responses to case-based questions, cases with similar traits to case examples provided at the end of the 7th SOSORT consensus paper on the conservative treatment of idiopathic and Scheuermann's kyphosis were assessed. Since this study did not meet the criteria for human subject research, it was exempt from ethics committee approval.

This study followed the protocol of the study by Hermann et al.[5] The phrasing of the questions was deliberately casual, with many composed in the first person, to resemble how typical patients might ask their inquiries when using the ChatGPT interface. We retained questions that required subjective or personalized responses (e.g., "Will I feel pain during brace-wearing for Scheuermann's kyphosis?") and vague questions (e.g., "Will my

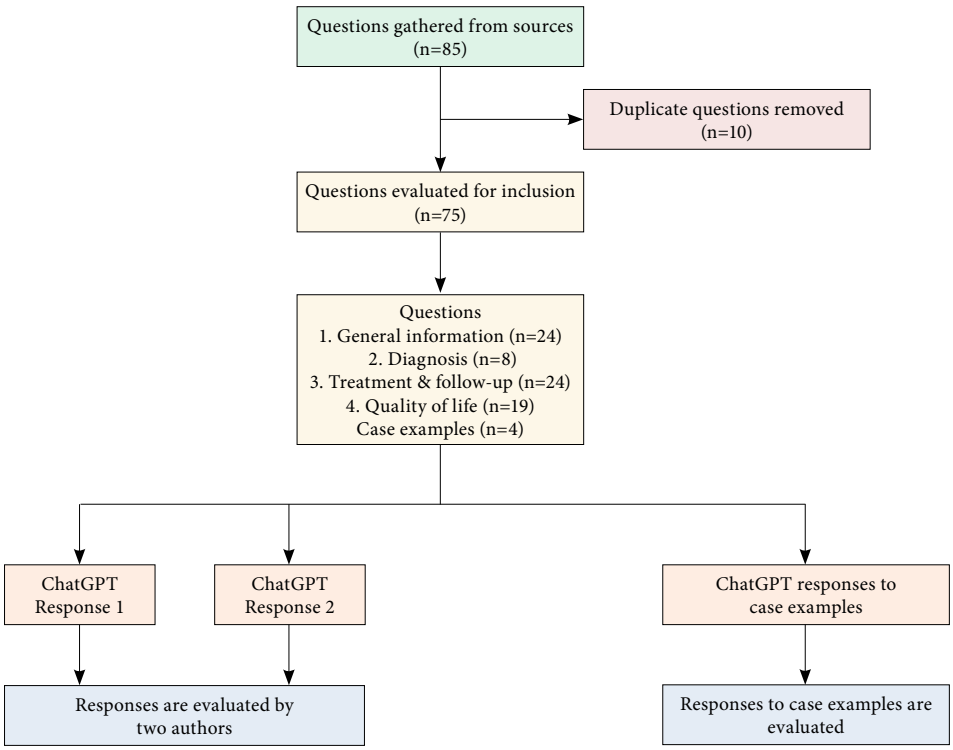

Figure 1. The workflow of the study.

Scheuermann's kyphosis get worse?") in their original form for inclusion in the study. The questions were input into ChatGPT version 3.5. The ChatGPT Plus subscription option, which utilizes version 4.0, was not assessed since it was not fully free, and it was expected that patients may not be able to use the full version. Additionally, the AI model was not specifically trained on Scheuermann's kyphosis or similar specialized medical conditions. Instead, the queries were designed to reflect typical real-world interactions, such as when a teenager or parent inputs questions into ChatGPT. This approach was selected to better represent actual use scenarios and to provide a more realistic evaluation of the AI's performance in addressing questions about Scheuermann's kyphosis.

Each individual question was inputted twice on separate occasions using the "new chat" function to generate two responses per question, ensuring the reproducibility of responses. The questions and responses were then exported into a table and forwarded to two physiatrists, both certified by the Italian Scientific Institute of Scoliosis (ISICO), and two experts in pediatric rehabilitation medicine for accuracy evaluation. Reviewers were blinded to each other's scores. They independently reviewed the responses based on a scoring system adapted from a scoring scale used in prior similar studies:[5,8] 1. Comprehensive: Defined as accurate and comprehensive, nothing more an expert physician in kyphosis or spinal deformities can add if asked this question by a patient. 2. Correct but inadequate: All information is correct but incomplete; an expert physician in kyphosis or spinal deformities would have more important information to add if asked this question by a patient. 3. Some correct and some incorrect. 4. Completely incorrect.

Response similarity was assessed to investigate reproducibility. If the responses were similar in accuracy, only the first response was graded (Supplementary file 1). Disagreements in reproducibility or grading of each response were resolved by a third ISICO-certified academic physiatrist. If disagreements persisted, three senior authors were consulted for a final decision. The final grades assigned to each question were based on the agreement of at least two reviewers and used to analyze the overall performance of ChatGPT in answering questions regarding Scheuermann's kyphosis.

Statistical analysis

Statistical analysis was conducted using IBM SPSS version 20.0 software (IBM Corp., Armonk, NY, USA). Descriptive statistics were used to calculate the

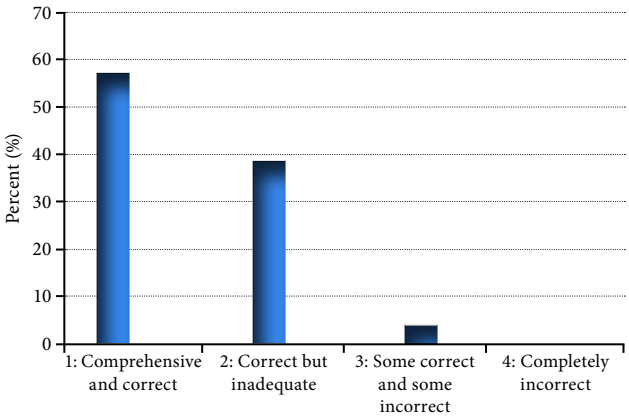

Figure 2. Distribution of the percentage of questions per grade across all questions.

proportion of responses for each score both overall and within each question category. To assess the reproducibility of ChatGPT's responses, the percentage of responses that differed when the same query was entered twice into the ChatGPT was calculated. Agreement between two reviewers was assessed using Kappa (κ) analysis, with Cohen's κ values interpreted as follows: κ between 0.0 and 0.40, poor agreement; κ=0.41 to 0.60, moderate agreement; κ=0.61 to 0.80, substantial agreement; κ=0.81 to 0.99, almost perfect agreement.[15] A p-value <0.05 was considered statistically significant.

RESULTS

The workflow of this study is depicted in Figure 1. Out of 75 questions, the responses were as follows: 43 (57.33%) questions received correct and comprehensive answers (score 1), 29 (38.67%) questions received correct but not comprehensive answers (score 2), three (4%) questions received partially incorrect answers (score 3), and 0 questions received completely incorrect answers (score 4). The distribution of scores across question categories is illustrated in Figure 2. ChatGPT demonstrated its best performance in the quality of life category, with responses to 16 out of 19 questions graded as correct and comprehensive answers. The second-best performance was in the general knowledge category, with 15 (62.3%) out of 24 questions scored as correct and comprehensive answers (score 1). The worst performance was in the diagnosis category, where three (37.5%) out of eight responses were rated as correct and comprehensive. Similarly, in the treatment and follow-up category, only nine (37.5%) out of 24 responses received a correct and comprehensive score (Figure 3).

In the second round of queries, ChatGPT's responses were inconsistent for six (8%) out of 75 questions, indicating that ChatGPT provided reproducible answers to 69 (92%) of the questions. Among these, general knowledge had the highest percentage of responses that were not similar in the second query, with four (16.7%) out of 24 questions. This was followed by diagnosis, where one (12.5%) out of seven questions had dissimilar responses, and treatment and follow-up, with one (4.2%) out of 23 questions showing inconsistency. Interestingly, all responses in the quality of life category were consistent between the two queries.

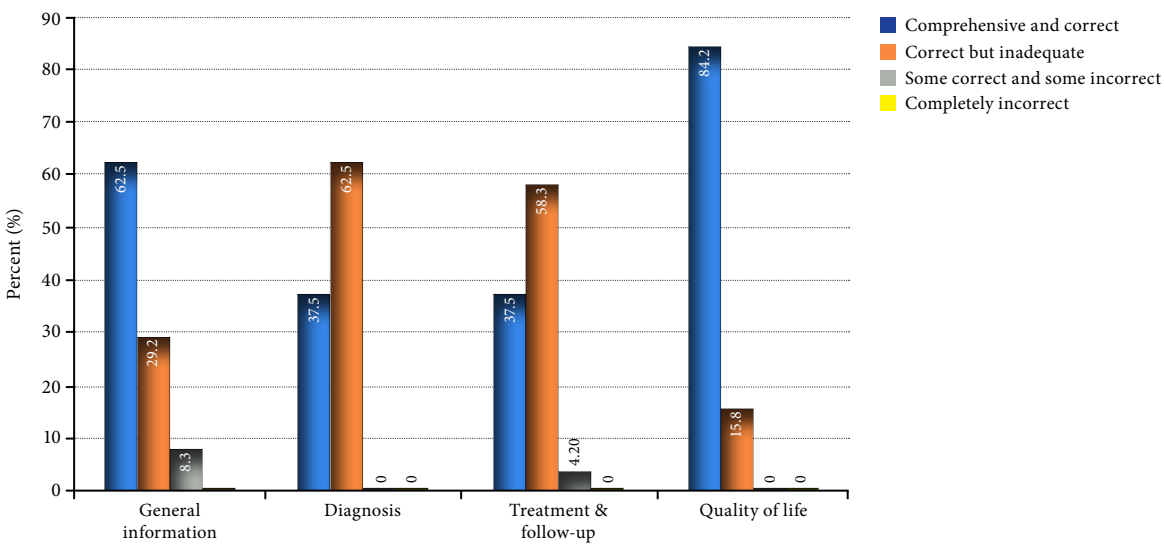

Figure 3. Scores represented as a percentage of the total number of questions within each category.

TABLE 1
Agreement between two reviewers

	Kappa (κ)	95% CI	*p*
All questions	0.79	0.65-0.92	0.0001
General information	0.84	0.62-1.05	0.0001
Diagnosis	0.57	0.12-1	0.04
Treatment & follow-up	0.84	0.63-1.05	0.0001
Quality of life	0.46	−0.14-1.05	0.02

CI: Confidence interval.

The calculated κ value for agreement between the first and second evaluators for all questions was 0.79 (95% confidence interval [CI]: 0.65, 0.92; p=0.0001), indicating a very high level of agreement. In the general knowledge category, the κ value was 0.84 (95% CI: 0.62, 1.05; p=0.0001), representing the highest level of agreement. However, in the diagnosis category, the κ value was 0.57 (95% CI: 0.12, 1.00; p=0.04), indicating the lowest level of agreement (Table 1). ChatGPT's responses regarding the management of all four cases were incorrect, as it suggested all current treatments indiscriminately (Supplementary file 2).

DISCUSSION

In the present study, we evaluated the accuracy and reproducibility of ChatGPT in addressing frequently asked patient inquiries concerning Scheuermann's kyphosis. Additionally, we assessed the accuracy of its responses to case-based questions. Overall, ChatGPT's responses were accurate, with the majority rated as comprehensive and demonstrating high reproducibility. It provided approximately 40% of the answers correctly, though the scope was often inadequate. Notably, partially incorrect answers were rare, and entirely wrong answers were absent. ChatGPT performed best in addressing topics on quality of life but had weaker results in diagnosis, treatment, and follow-up. For case-based questions, it listed all treatment options rather than tailoring responses to the patient's specific condition.

With the prediction that AI will play a significant role as a consultative resource in health literacy,[16] studies have begun to investigate the accuracy and comprehensiveness of responses provided by ChatGPT to questions posed in medical specialties.[5,7,8,10,17] In one of these studies, ChatGPT provided accurate and comprehensive answers to frequently asked questions about bariatric surgery at a rate of 86%,[8] and similarly high reproducible answers (90%) were obtained in our study. However, in some question categories (eligibility, efficacy, procedure options, recovery, risks, and complications), there were incorrect answers that could be dangerous for the patient, highlighting the necessity of oversight and guidance by a healthcare professional.[8] Yeo et al.[17] showed that ChatGPT correctly answered 79% of cirrhosis and 74% of hepatocellular carcinoma questions, although only 40% were comprehensive. ChatGPT excelled in providing basic information and lifestyle advice but lacked tailored recommendations. In another study, otorhinolaryngology specialists outperformed ChatGPT in medical adequacy and consistency.[10] While clinicians integrate history, examination, and opinion to diagnose, ChatGPT lacks this capability. This was also observed in case-based questions in our study, where ChatGPT struggled with individualized treatment recommendations for patients with kyphosis, underscoring the need for personalized medical consultation. Similarly, for cardiac symptoms, ChatGPT matched real recommendations 90% of the time in case questions but only 50% in complex cases, showing incomplete advice compared to expert opinions.[7] In another study, ChatGPT accurately answered questions on cervical cancer prevention, survival, and quality of life but performed less well in diagnosis and treatment categories, similar to our findings. Due to partially correct and incorrect answers, the authors recommended further development with input from doctors before using ChatGPT as an information source for cervical cancer diagnosis and treatment.[5] As noted in previous and present studies, while ChatGPT excels in providing quality of life and general information, it lacks adequate performance in areas requiring synthesis of information and personalized treatment planning.

In Scheuermann's kyphosis, thoracic kyphosis is often accompanied by compensatory cervical and lumbar hyperlordosis, forward head posture, and muscle tightness in the pectoral, hamstring, iliopsoas, and lumbar extensor muscles.[4] Asymptomatic patients typically do not seek medical help until noticeable deformity and posture issues arise.[2,18] Parents may first perceive the posture as a bad habit, attributing it to prolonged use of mobile devices or heavy school bags, which can delay diagnosis.[4] However, as the deformity becomes more visible, aesthetic concerns lead adolescents and their families to seek treatment.[2] Many people now seek online information before consulting

healthcare professionals, with one-third of inquiries coming from those under the age of 18.[19] Artificial intelligence platforms are increasingly favored over search engines for providing more comprehensive and understandable information. The present study found that ChatGPT performed poorly in diagnosis and treatment categories, risking misguidance due to its inability to provide personalized recommendations. Consulting a spinal deformity specialist remains essential. There is no evidence-based protocol for kyphosis evaluation, but clinical, radiological, and psychosocial assessments are recommended. For Scheuermann's kyphosis, diagnosis involves detecting rigid kyphosis and radiological findings of over 45° kyphosis and 5° wedging in three vertebrae.[2] A recent study comparing AI models in scoliosis diagnosis found ChatGPT accurate but less reliable in advanced assessments such as Cobb angle measurement and curvature classification, showing the current limitations of AI in spinal deformity evaluation.[20] The general information adolescents gain from AI platforms can encourage early consultation with spine specialists. Treatment aims to reduce or stabilize the deformity, relieve pain, and improve cosmetic appearance.[21,22] Early diagnosis of Scheuermann's kyphosis is key to effective treatment.[22] The general information adolescents gain from AI platforms can encourage early consultation with spine specialists. ChatGPT suggests observation for mild cases, exercises and bracing for moderate cases, and surgery for severe cases, while stressing personalized treatment plans. However, it lacks deformity-specific responses, making physician consultation crucial. Only specialists can integrate critical factors, such as maturation status, deformity severity, and pain, to develop personalized plans. For example, a spinal deformity expert determines the type, suitability, and usage of a brace, and plans the treatment and weaning periods.[23]

To the best of our knowledge, this study is the first to examine the utility of ChatGPT in responding to queries related to Scheuermann's kyphosis. High agreement among evaluators is one of the strengths of our study. Patient questions were constructed from comprehensive and real case examples, frequently asked in our daily practice, and enriched with questions sourced from reputable references and websites. The inclusion of case-based questions provided preliminary insights into how ChatGPT might perform in real-world scenarios, although we anticipated that it could not replace expert opinion at present.

ChatGPT can serve as a helpful information source for patients by providing generally accurate information and promoting the importance of early diagnosis. However, its limited scope and occasional errors may lead to confusion. As observed from its case-based responses, the information is often generalized rather than personalized, leaving patients with insufficient details to make informed treatment decisions. These limitations suggest that while ChatGPT can supplement patient information, it cannot match the accuracy and effectiveness of a healthcare professional's evaluation, diagnosis, and treatment selection.

A limitation of the present study is that ChatGPT version 3.5 was used. This version is free and the most widely used version. However, a newer model, version 4.0, has been released, which OpenAI claims is faster and more efficient than previous versions. It is reported to be capable of understanding more complex and nuanced inputs, processing not only text but also other media such as images and videos, providing safer and more useful responses, reducing harmful or incorrect information, and delivering responses tailored to user needs.[5,24] Although OpenAI does not disclose usage data publicly, we believe that, at the time of our study's completion, version 4.0 had not yet achieved the same level of widespread adoption as version 3.5, primarily due to its paid access. The advancements in ChatGPT 4.0's capabilities could significantly influence the accuracy of responses. If we had utilized this newer version, we might have observed different outcomes, particularly regarding nuanced medical inquiries related to general information and quality of life concerning Scheuermann's kyphosis. However, it is essential to recognize that AI, including ChatGPT 4.0, still has limitations in accurately diagnosing medical conditions and formulating treatment plans. The use of AI in medical diagnostics is still in its early phases, with several technical, regulatory, and ethical challenges to overcome. A major issue is the quality and availability of medical data. Artificial intelligence systems need large amounts of well-labelled, high-quality data to be effective, but medical data is often incomplete, fragmented, or inaccessible.[25] Additionally, training models such as ChatGPT on unrepresentative data can lead to biased or inaccurate diagnoses. Ethical concerns also arise when AI is applied to sensitive medical information.[25] Artificial intelligence can assist healthcare professionals by providing information and suggestions based on existing

data and guidelines. However, making accurate diagnoses and treatment decisions often requires clinical expertise, a comprehensive understanding of individual patient contexts, and the ability to consider nuanced factors that AI may not fully grasp. Consequently, while AI can enhance the decision-making process, it cannot replace the critical role of healthcare professionals in effectively diagnosing and treating patients. It is important to note that ChatGPT was not specifically trained on Scheuermann's kyphosis or other specialized medical conditions. Its general training aims to simulate typical real-world interactions, which may limit its effectiveness for highly specialized queries. Specialized training with domain-specific datasets could greatly improve the AI's accuracy and ability to provide detailed, evidence-based information on Scheuermann's kyphosis, benefiting both healthcare professionals and patients. The general training may impact ChatGPT's accuracy, particularly in areas such as treatment and follow-up, which require specialized knowledge. Future research should focus on training AI models with domain-specific data to enhance performance and reliability in specialized fields, improving their utility for conditions like Scheuermann's kyphosis.

In conclusion, while ChatGPT provides accurate and comprehensive answers to questions on general information and quality of life related to Scheuermann's kyphosis, it does not achieve the same level of success in diagnosing and addressing treatment-related queries. Patients with hyperkyphosis should seek consultation from a spine specialist for accurate diagnosis, comprehensive evaluation, and appropriate treatment planning. Artificial intelligence applications should be enriched with inputs from healthcare professionals to enhance their effectiveness in healthcare settings. Further research is needed to determine the optimal stages for integrating AI platforms into healthcare services.

Data Sharing Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Author Contributions: All authors contributed equally to this article.

Conflict of Interest: The authors declared no conflicts of interest with respect to the authorship and/or publication of this article.

Funding: The authors received no financial support for the research and/or authorship of this article.

REFERENCES

1. Sardar ZM, Ames RJ, Lenke L. Scheuermann's kyphosis: diagnosis, management, and selecting fusion levels. J Am Acad Orthop Surg 2019;27:e462-72. doi: 10.5435/JAAOS-D-17-00748.
2. Bezalel T, Carmeli E, Been E, Kalichman L. Scheuermann's disease: Current diagnosis and treatment approach. J Back Musculoskelet Rehabil 2014;27:383-90. doi: 10.3233/BMR-140483.
3. Cetik RM, Latalski M, Yazici M. Management of low back pain accompanying sagittal plane pathologies in children: Spondylolysis/spondylolisthesis and Scheuermann's disease. J Child Orthop 2023;17:535-47. doi: 10.1177/18632521231215873.
4. Sebaaly A, Farjallah S, Kharrat K, Kreichati G, Daher M. Scheuermann's kyphosis: Update on pathophysiology and surgical treatment. EFORT Open Rev 2022;7:782-91. doi: 10.1530/EOR-22-0063.
5. Hermann CE, Patel JM, Boyd L, Growdon WB, Aviki E, Stasenko M. Let's chat about cervical cancer: Assessing the accuracy of ChatGPT responses to cervical cancer questions. Gynecol Oncol 2023;179:164-8. doi: 10.1016/j.ygyno.2023.11.008.
6. Gilson A, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA, et al. How does ChatGPT perform on the United States Medical Licensing Examination (USMLE)? The implications of large language models for medical education and knowledge assessment. JMIR Med Educ 2023;9:e45312. doi: 10.2196/45312.
7. Harskamp RE, De Clercq L. Performance of ChatGPT as an AI-assisted decision support tool in medicine: A proof-of-concept study for interpreting symptoms and management of common cardiac conditions (AMSTELHEART-2). Acta Cardiol 2024;79:358-66. doi: 10.1080/00015385.2024.2303528.
8. Samaan JS, Yeo YH, Rajeev N, Hawley L, Abel S, Ng WH, et al. Assessing the accuracy of responses by the language model ChatGPT to questions regarding bariatric surgery. Obes Surg 2023;33:1790-6. doi: 10.1007/s11695-023-06603-5.
9. Shrestha N, Shen Z, Zaidat B, Duey AH, Tang JE, Ahmed W, et al. Performance of ChatGPT on NASS clinical guidelines for the diagnosis and treatment of low back pain: A comparison study. Spine (Phila Pa 1976) 2024;49:640-51. doi: 10.1097/BRS.0000000000004915.
10. Buhr CR, Smith H, Huppertz T, Bahr-Hamm K, Matthias C, Blaikie A, et al. ChatGPT versus consultants: Blinded evaluation on answering otorhinolaryngology case-based questions. JMIR Med Educ 2023;9:e49183. doi: 10.2196/49183.
11. Spinal Deformity: Kyphosis. A Handbook for Patients. Scoliosis Research Society; 2018. Available at: https://www.spinecompany.co.uk/pdf/kyphosishandbook-srs.pdf. [Accessed: 20.4.2024]
12. de Mauroy J, Weiss H, Aulisa A, Aulisa L, Brox J, Durmala J, et al. 7th SOSORT consensus paper: Conservative treatment of idiopathic & Scheuermann's kyphosis. Scoliosis 2010;5:9. doi: 10.1186/1748-7161-5-9.

13. Nowak JE. Scheuermann Disease Questions & Answers. Available at: https://emedicine.medscape.com/article/311959-questions-and-answers?form=fpf#qna-toc. [Accessed: 20.4.2024]
14. Pediatric Kyphosis. Available at: https://www.childrensnational.org/get-care/health-library/kyphosis. [Accessed: 20.4.2024]
15. Giray E, Kenis-Coskun O, Karadag-Saygi E, Ozyemisci-Taskiran O. Interrater reliability, acceptability, and practicality of real-time video pediatric gait, arms, legs, and spine for musculoskeletal assessment of children during telemedicine visits. J Clin Rheumatol 2022;28:235-9. doi: 10.1097/RHU.0000000000001840.
16. Al Kuwaiti A, Nazer K, Al-Reedy A, Al-Shehri S, Al-Muhanna A, Subbarayalu AV, et al. A review of the role of artificial intelligence in healthcare. J Pers Med 2023;13:951. doi: 10.3390/jpm13060951.
17. Yeo YH, Samaan JS, Ng WH, Ting PS, Trivedi H, Vipani A, et al. Assessing the performance of ChatGPT in answering questions regarding cirrhosis and hepatocellular carcinoma. Clin Mol Hepatol 2023;29:721-32. doi: 10.3350/cmh.2023.0089.
18. Lowe TG, Line BG. Evidence based medicine: Analysis of Scheuermann kyphosis. Spine (Phila Pa 1976) 2007;32(19 Suppl):S115-9. doi: 10.1097/BRS.0b013e3181354501.
19. Keeley B, Little C. The State of the Worlds Children 2017: Children in a Digital World. 2017. Available at: https://eric.ed.gov/?id=ED590013. [Accessed: 20.4.2024]
20. Fabijan A, Polis B, Fabijan R, Zakrzewski K, Nowosławska E, Zawadzka-Fabijan A. Artificial intelligence in scoliosis classification: An investigation of language-based models. J Pers Med 2023;13:1695. doi: 10.3390/jpm13121695.
21. Huq S, Ehresman J, Cottrill E, Ahmed AK, Pennington Z, Westbroek EM, et al. Treatment approaches for Scheuermann kyphosis: A systematic review of historic and current management. J Neurosurg Spine 2019;32:235-47. doi: 10.3171/2019.8.SPINE19500.
22. Palazzo C, Sailhan F, Revel M. Scheuermann's disease: An update. Joint Bone Spine 2014;81:209-14. doi: 10.1016/j.jbspin.2013.11.012.
23. Negrini S, Donzelli S, Aulisa AG, Czaprowski D, Schreiber S, de Mauroy JC, et al. 2016 SOSORT guidelines: Orthopaedic and rehabilitation treatment of idiopathic scoliosis during growth. Scoliosis Spinal Disord 2018;13:3. doi: 10.1186/s13013-017-0145-8.
24. Introducing GPT-4o and more tools to ChatGPT free users. Available at: https://openai.com/index/gpt-4o-and-more-tools-to-chatgpt-free/. [Accessed: 20.4.2024]
25. Al-Antari MA. Artificial intelligence for medical diagnostics-existing and future AI technology! Diagnostics (Basel) 2023;13:688. doi: 10.3390/diagnostics13040688.