



OPEN ACCESS

EDITED BY

Wencai Liu,
Shanghai Jiao Tong University, China

REVIEWED BY

Harish Kempegowda,
Boston University, United States
Hartanto Hartanto,
Universitas Widya Dharma, Indonesia
Nicola Manocchio,
University of Rome Tor Vergata, Italy

*CORRESPONDENCE

Litai Ma
✉ ma.litai@163.com

RECEIVED 07 November 2024

ACCEPTED 12 February 2025

PUBLISHED 25 February 2025

CITATION

Jia X, Ma L, Yang Y, Deng Y, Shen C, Zhang K
and Li Y (2025) ChatGPT4o's theranostic
performance in the management of
thoracolumbar spine fractures.
Front. Surg. 12:1524396.
doi: 10.3389/fsurg.2025.1524396

COPYRIGHT

© 2025 Jia, Ma, Yang, Deng, Shen, Zhang and
Li. This is an open-access article distributed
under the terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other forums is
permitted, provided the original author(s) and
the copyright owner(s) are credited and that
the original publication in this journal is cited,
in accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

ChatGPT4o's theranostic performance in the management of thoracolumbar spine fractures

Xuehai Jia¹, Litai Ma^{1*}, Yi Yang¹, Yi Deng¹, Changyong Shen¹,
Kerui Zhang¹ and Ya Li²

¹Department of Orthopedics, Orthopedic Research Institute, West China Hospital, Sichuan University, Chengdu, Sichuan, China, ²The First Affiliated Hospital of Shihezi University, Shihezi, Xinjiang Uyghur, China

KEYWORDS

ChatGPT4o, thoracolumbar spine fractures, theranostic performance, clinical practice, AI in medicine

Introduction

ChatGPT, developed by OpenAI (<https://chat.openai.com>), is a publicly accessible tool that utilizes advanced machine learning algorithms to process and analyze extensive data, generating responses to user inquiries. On May 13, 2024, OpenAI launched the ChatGPT4o model, which, according to information on the OpenAI website, represents the latest, fastest, and most advanced version. This model supports a context length of up to 128k tokens (equivalent to the length of a long novel) and offers multimodal capabilities, including text and image inputs, as well as text, image, and audio outputs (<https://help.openai.com>). While numerous studies have explored ChatGPT's potential applications and challenges in the biomedical field (1, 2), limited research has been conducted on the specific capabilities of ChatGPT4o in the medical domain. A REVIEW article (3) published in *Frontiers in Surgery* mentions that ChatGPT lacks sufficient expertise and background understanding in specialized fields. However, the application of ChatGPT4o may have the potential to change this situation. To validate this model, we investigate the theranostic performance of ChatGPT4o in managing thoracolumbar spine fractures to assess its potential effectiveness and applications in clinical practice.

Method

For our evaluation, we formulated 38 clinical questions based on the diagnostic, treatment, and management guidelines for thoracolumbar fractures established by the Congress of Neurological Surgeons (CNS) (4–14) and the Chinese Medical Association (CMA) (15). We input all 38 questions into ChatGPT-4o (OpenAI, accessed November 3, 2024) without providing additional context or guidelines. Each question was posed once, and the initial generated response was recorded. To minimize variability, no iterative refinement of prompts was performed. The responses were anonymized and compiled in [Supplementary Material S1](#). Each response was subsequently reviewed by three independent spine surgery experts, who evaluated the responses according to both the established guidelines and their own clinical experience. Each expert used a five-point Likert scale to rate the responses: (1) indicating completely incorrect; (2) more incorrect than correct; (3) an equal mix of correct and incorrect; (4) more correct than

TABLE 1 Five-point Likert scores for responses from inquires posed to chat-GPT4o.

Questions	Five-point Likert scores ^a			
	Expert 1	Expert 2	Expert 3	Median
1. Which patients need to consider combined thoracolumbar spinal cord injury?	5	5	4	5
2. How to immobilize, transport, and transfer patients with suspected thoracolumbar spinal cord injury?	5	5	5	5
3. How to assess the degree of neurological injury in patients with acute thoracolumbar spinal cord injury?	5	5	4	5
4. How to perform radiological assessment for patients with suspected acute thoracolumbar spinal cord injury?	4	4	4	4
5. How to assess the morphology of injury in patients with acute thoracolumbar spinal cord injury?	4	4	4	4
6. Is the use of high-dose corticosteroids and gangliosides recommended for the treatment of spinal cord injury?	5	5	5	5
7. What are the indications for the treatment selection in patients with acute thoracolumbar spinal cord injury?	5	5	5	5
8. What are the recommended conservative treatment methods for thoracolumbar spinal cord injury?	5	5	5	5
9. What is the timing for surgery in patients with acute thoracolumbar spinal cord injury?	4	5	4	4
10. How to choose the surgical approach for patients with acute thoracolumbar spinal cord injury?	5	5	4	5
11. When is laminectomy necessary for patients with acute thoracolumbar spinal cord injury?	5	5	5	5
12. How to select the fixation segment for posterior surgery in patients with acute thoracolumbar spinal cord injury?	5	5	5	5
13. Is it necessary to fix the injured vertebrae posteriorly in patients with acute thoracolumbar spinal cord injury?	4	5	4	4
14. What are the indications for percutaneous internal fixation in patients with acute thoracolumbar spinal cord injury?	5	5	5	5
15. Is it necessary to perform bone graft fusion for patients with acute thoracolumbar spinal cord injury who undergo surgery?	5	4	5	5
16. Is simple pedicle-based bone grafting effective in patients with acute thoracolumbar spinal cord injury?	5	5	5	5
17. How to manage the urinary system in patients with acute thoracolumbar spinal cord injury?	5	5	5	5
18. How to prevent and treat deep vein thrombosis in patients with acute thoracolumbar spinal cord injury?	5	5	5	5
19. Is it necessary to prevent and treat pressure sores in patients with acute thoracolumbar spinal cord injury?	5	5	5	5
20. How to manage neurogenic bowel in patients with acute thoracolumbar spinal cord injury?	5	5	5	5
21. Does early surgical intervention improve outcomes for patients with thoracic and lumbar fractures?	4	5	4	4
22. Does the choice of surgical approach (anterior, posterior, or combined anterior-posterior) improve clinical outcomes in patients with thoracic and lumbar fractures?	2	2	2	2
23. Are there radiographic findings in patients with traumatic thoracolumbar fractures that can predict the need for surgical intervention?	5	5	5	5
24. Are there radiographic findings in patients with traumatic thoracolumbar fractures that can assist in predicting clinical outcomes?	3	4	3	3
25. Does routine screening for DVT prevent PE (or VTE-associated morbidity and mortality) in patients with thoracic and lumbar fractures?	5	5	5	5
26. For patients with thoracic and lumbar fractures, is one regimen of VTE prophylaxis superior to others with respect to prevention of PE (or VTE-associated morbidity and mortality)?	5	5	5	5
27. Is there a specific treatment regimen for documented VTE that provides fewer complications than other treatments in patients with thoracic and lumbar fractures?	4	4	4	4
28. Does the administration of a specific pharmacologic agent (e.g., methylprednisolone) improve clinical outcomes in patients with thoracic and lumbar fractures and spinal cord injury?	5	5	5	5
29. Does the surgical treatment of burst fractures of the thoracic and lumbar spine improve clinical outcomes compared to nonoperative treatment?	4	5	5	5
30. Does the surgical treatment of nonburst fractures of the thoracic and lumbar spine improve clinical outcomes compared to nonoperative treatment?	4	5	4	4
31. Does the addition of arthrodesis to instrumented fixation improve outcomes in patients with thoracic and lumbar burst fractures?	5	5	5	5
32. How does the use of minimally invasive techniques (including percutaneous instrumentation) affect outcomes in patients undergoing surgery for thoracic and lumbar fractures compared to conventional open techniques?	5	5	5	5
33. Does the use of external bracing improve outcomes in the nonoperative treatment of neurologically intact patients with thoracic and lumbar burst fractures?	5	5	5	5
34. Which neurological assessment tools have demonstrated internal reliability and validity in the management of patients with thoracic and lumbar fractures (i.e., do these instruments provide consistent information between different care providers)?	4	5	4	4
35. Are there any clinical findings (e.g., presenting neurological grade/function) in patients with thoracic and lumbar fractures that can assist in predicting clinical outcomes?	5	5	5	5
36. Does the active maintenance of arterial blood pressure after injury affect clinical outcomes in patients with thoracic and lumbar fractures?	5	5	5	5
37. Are there classification systems for fractures of the thoracolumbar spine that have been shown to be internally valid and reliable (i.e., do these instruments provide consistent information between different care providers)?	5	4	5	5
38. In treating patients with thoracolumbar fractures, does employing a formally tested classification system for treatment decision-making affect clinical outcomes?	5	5	5	5

DVT, deep vein thrombosis; PE, pulmonary embolism.

^aFive-point Likert score system: 1 means completely incorrect; 2 means more incorrect than correct; 3 means equally incorrect and correct; 4 means more correct than incorrect; 5 means completely correct.

incorrect; and (5) completely correct. The median score from the three experts was used as the final rating to minimize bias.

Result

When ChatGPT4o was presented with “yes or no” questions, it typically responded with comprehensive diagnostic criteria and therapeutic principles rather than a simple “yes” or “no.” According to our results (Table 1), 0 responses (0%) received a score of 1, 1 response (2.63%) received a score of 2, 1 response (2.63%) scored a 3, 8 responses (21.05%) scored a 4, and 28 responses (73.68%) scored a 5. Approximately 94.7% of the responses were largely or entirely accurate.

Discussion

When asked, “Does the choice of surgical approach (anterior, posterior, or combined anterior-posterior) improve clinical outcomes in patients with thoracic and lumbar fractures?”, ChatGPT4o provided an affirmative answer along with detailed explanations. However, according to CNS guidelines, for patients with burst fractures of the thoracolumbar spine, surgeons may use an anterior, posterior, or combined approach, as the choice of approach does not significantly affect clinical or neurological outcomes, a Grade B recommendation. Although ChatGPT4o provided a detailed explanation of the indications for each approach, the experts noted that while the response was generally accurate, the final conclusion was not entirely consistent with guideline recommendations. Furthermore, while ChatGPT4o appears capable of conducting targeted searches on open websites, its “independent reasoning” abilities require further refinement.

In summary, ChatGPT4o demonstrates promising performance in diagnosing and treating thoracolumbar trauma. Its ability to search open websites and provide detailed responses could be a useful reference for clinical practitioners. However, ChatGPT4o does not consistently provide fully accurate answers, particularly with “yes or no” questions. Its dependence on specific sources for data retrieval may introduce biases that limit its broader application in the field of spine surgery. ChatGPT requires substantial medical data for further training to enhance model performance. Moreover, given the specific ethical considerations in medicine, ChatGPT4o’s use in clinical settings must ensure patient safety, data privacy, ethical standards, and adherence to relevant “AI regulations”. Although ChatGPT4o’s responses may improve clinical efficiency, it should only serve as a clinical assistant, with spine surgeons validating the accuracy of its information.

This study has several methodological limitations: firstly, the lack of comparative analyses with established AI systems (e.g., Google Med-PaLM, IBM Watson) or traditional decision-support tools hinders definitive performance benchmarking; secondly, simulated testing environments may overestimate system efficacy, as diagnostic performance degradation in real-world clinical

settings requires urgent empirical validation; finally, the rapid evolution of AI technology necessitates dynamically updated training databases and ethical evaluation frameworks. To address these gaps, subsequent research will incorporate the Partial Credit Model (PCM) and Item Response Theory (IRT) through latent trait modeling, systematically quantifying AI response difficulty levels, refining multidimensional scoring criteria, and strengthening clinical applicability assessments to establish a psychometrically-based evaluation framework. This methodological advancement will enhance the granular understanding of AI’s role in complex medical decision-making (e.g., surgical approach selection, prognostic stratification). Future research priorities include: (1) comparative effectiveness studies across AI systems, (2) real-world clinical validation of performance, and (3) development of specialty-specific human-AI collaboration guidelines to systematically improve the clinical utility of intelligent assistive tools in spinal surgery.

Author contributions

XJ: Investigation, Methodology, Writing – original draft. LM: Supervision, Writing – review & editing. YY: Data curation, Validation, Writing – original draft. YD: Data curation, Investigation, Writing – original draft. CS: Methodology, Validation, Writing – original draft. KZ: Data curation, Investigation, Methodology, Writing – original draft. YL: Data curation, Investigation, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that Generative AI was used in the creation of this manuscript. Provided answers to guideline-related questions regarding thoracolumbar spine fractures.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those

of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Tian S, Jin Q, Yeganova L, Lai P-T, Zhu Q, Chen X, et al. Opportunities and challenges for ChatGPT and large language models in biomedicine and health. *Brief Bioinform.* (2023) 25:bbad493. doi: 10.1093/bib/bbad493
2. Zhang J, Sun K, Jagadeesh A, Falakafaki P, Kayayan E, Tao G, et al. The potential and pitfalls of using a large language model such as ChatGPT, GPT-4, or LLaMA as a clinical assistant. *J Am Med Inform Assoc.* (2024) 31:1884–91. doi: 10.1093/jamia/ocae184
3. Giorgino R, Alessandri-Bonetti M, Luca A, Migliorini F, Rossi N, Peretti GM, et al. ChatGPT in orthopedics: a narrative review exploring the potential of artificial intelligence in orthopedic practice. *Front Surg.* (2023) 10:1284015. doi: 10.3389/fsurg.2023.1284015
4. Dailey AT, Arnold PM, Anderson PA, Chi JH, Dhall SS, Eichholz KM, et al. Congress of neurological surgeons systematic review and evidence-based guidelines on the evaluation and treatment of patients with thoracolumbar spine trauma: classification of injury. *Neurosurgery.* (2019) 84:E24–7. doi: 10.1093/neuros/nyy372
5. Dhall SS, Dailey AT, Anderson PA, Arnold PM, Chi JH, Eichholz KM, et al. Congress of neurological surgeons systematic review and evidence-based guidelines on the evaluation and treatment of patients with thoracolumbar spine trauma: hemodynamic management. *Neurosurgery.* (2019) 84:E43–5. doi: 10.1093/neuros/nyy368
6. Harrop JS, Chi JH, Anderson PA, Arnold PM, Dailey AT, Dhall SS, et al. Congress of neurological surgeons systematic review and evidence-based guidelines on the evaluation and treatment of patients with thoracolumbar spine trauma: neurological assessment. *Neurosurgery.* (2019) 84:E32–5. doi: 10.1093/neuros/nyy370
7. Hoh DJ, Qureshi S, Anderson PA, Arnold PM, John HC, Dailey AT, et al. Congress of neurological surgeons systematic review and evidence-based guidelines on the evaluation and treatment of patients with thoracolumbar spine trauma: nonoperative care. *Neurosurgery.* (2019) 84:E46–9. doi: 10.1093/neuros/nyy369
8. Chi JH, Eichholz KM, Anderson PA, Arnold PM, Dailey AT, Dhall SS, et al. Congress of neurological surgeons systematic review and evidence-based guidelines on the evaluation and treatment of patients with thoracolumbar spine trauma: novel surgical strategies. *Neurosurgery.* (2019) 84:E59–62. doi: 10.1093/neuros/nyy364
9. Rabb CH, Hoh DJ, Anderson PA, Arnold PM, Chi JH, Dailey AT, et al. Congress of neurological surgeons systematic review and evidence-based guidelines on the evaluation and treatment of patients with thoracolumbar spine trauma: operative versus nonoperative treatment. *Neurosurgery.* (2019) 84:E50–2. doi: 10.1093/neuros/nyy361
10. Arnold PM, Anderson PA, Chi JH, Dailey AT, Dhall SS, Eichholz KM, et al. Congress of neurological surgeons systematic review and evidence-based guidelines on the evaluation and treatment of patients with thoracolumbar spine trauma: pharmacological treatment. *Neurosurgery.* (2019) 84:E36–8. doi: 10.1093/neuros/nyy371
11. Raksin PB, Harrop JS, Anderson PA, Arnold PM, Chi JH, Dailey AT, et al. Congress of neurological surgeons systematic review and evidence-based guidelines on the evaluation and treatment of patients with thoracolumbar spine trauma: prophylaxis and treatment of thromboembolic events. *Neurosurgery.* (2019) 84:E39–42. doi: 10.1093/neuros/nyy367
12. Qureshi S, Dhall SS, Anderson PA, Arnold PM, Chi JH, Dailey AT, et al. Congress of neurological surgeons systematic review and evidence-based guidelines on the evaluation and treatment of patients with thoracolumbar spine trauma: radiological evaluation. *Neurosurgery.* (2019) 84:E28–31. doi: 10.1093/neuros/nyy373
13. Anderson PA, Raksin PB, Arnold PM, Chi JH, Dailey AT, Dhall SS, et al. Congress of neurological surgeons systematic review and evidence-based guidelines on the evaluation and treatment of patients with thoracolumbar spine trauma: surgical approaches. *Neurosurgery.* (2019) 84:E56–8. doi: 10.1093/neuros/nyy363
14. Eichholz KM, Rabb CH, Anderson PA, Arnold PM, Chi JH, Dailey AT, et al. Congress of neurological surgeons systematic review and evidence-based guidelines on the evaluation and treatment of patients with thoracolumbar spine trauma: timing of surgical intervention. *Neurosurgery.* (2019) 84:E53–5. doi: 10.1093/neuros/nyy362
15. Chinese Medical Doctor Association Orthopedics Branch, Editorial Board of the Evidence-based Clinical Practice Guidelines for Acute Thoracolumbar Spinal Cord Injury in Adults by the Chinese Medical Doctor Association Orthopedics Branch. Evidence-based clinical practice guidelines for orthopedics by the Chinese medical doctor association orthopedics branch: evidence-based clinical practice guidelines for acute thoracolumbar spinal cord injury in adults. *Chin J Surg.* (2019) 57(3):161–5. doi: 10.3760/cma.j.issn.0529-5815.2019.03.001

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fsurg.2025.1524396/full#supplementary-material>