JB&JS
OPEN ACCESS

AMERICAN ORTHOPAEDIC ASSOCIATION

AOA Critical Issues in Education

ChatGPT-4 Knows Its A B C D E but Cannot Cite Its Source

Diane Ghanem, MD, Alexander R. Zhu, BA, Whitney Kagabo, MD, Greg Osgood, MD, FAOA, and Babar Shafiq, MD, FAOA

Introduction: The artificial intelligence language model Chat Generative Pretrained Transformer (ChatGPT) has shown potential as a reliable and accessible educational resource in orthopaedic surgery. Yet, the accuracy of the references behind the provided information remains elusive, which poses a concern for maintaining the integrity of medical content. This study aims to examine the accuracy of the references provided by ChatGPT-4 concerning the Airway, Breathing, Circulation, Disability, Exposure (ABCDE) approach in trauma surgery.
Methods: Two independent reviewers critically assessed 30 ChatGPT-4–generated references supporting the well-established ABCDE approach to trauma protocol, grading them as 0 (nonexistent), 1 (inaccurate), or 2 (accurate). All discrepancies between the ChatGPT-4 and PubMed references were carefully reviewed and bolded. Cohen's Kappa coefficient was used to examine the agreement of the accuracy scores of the ChatGPT-4–generated references between reviewers. Descriptive statistics were used to summarize the mean reference accuracy scores. To compare the variance of the means across the 5 categories, one-way analysis of variance was used.
Results: ChatGPT-4 had an average reference accuracy score of 66.7%. Of the 30 references, only 43.3% were accurate and deemed "true" while 56.7% were categorized as "false" (43.3% inaccurate and 13.3% nonexistent). The accuracy was consistent across the 5 trauma protocol categories, with no significant statistical difference ($p = 0.437$).
Discussion: With 57% of references being inaccurate or nonexistent, ChatGPT-4 has fallen short in providing reliable and reproducible references—a concerning finding for the safety of using ChatGPT-4 for professional medical decision making without thorough verification. Only if used cautiously, with cross-referencing, can this language model act as an adjunct learning tool that can enhance comprehensiveness as well as knowledge rehearsal and manipulation.

Introduction

With the advent of artificial intelligence (AI), the landscape of medical information and decision making is undergoing a significant transformation. One of the most notable developments in this field is the publicly available Chat Generative Pretrained Transformer (ChatGPT), an advanced large language model launched by OpenAI. While not originally designed for medical applications, its sophisticated algorithm and extensive

The authors have no funding, financial relationships, or conflict of interest to disclose.

Disclosure: The **Disclosure of Potential Conflicts of Interest** forms are provided with the online version of the article (http://links.lww.com/JBJSOA/A667).

Copyright © 2024 The Authors. Published by The Journal of Bone and Joint Surgery, Incorporated. All rights reserved. This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

knowledge base have sparked interest in its potential utility in health care, particularly as a tool for supporting clinical decision making and improving medical education.[1,2]

In orthopaedic surgery education, Kung et al. and Ghanem et al. examined ChatGPT's performance on the Orthopaedic In-Training Examination (OITE)[3,4]. Both studies suggested that ChatGPT could be used as an adjunct to residents' education by providing evidence-based information and helping improve their understanding of OITE cases and general orthopaedic principles. In clinical decision making, ChatGPT can assist surgeons in making informed decisions about surgical procedures. It can provide information on best practices, likely outcomes, and risk factors based on patient-specific data and prevalent medical knowledge[5,6]. In addition, Ghanem et al. suggested that ChatGPT can enhance patients' care by providing guidelines for rehabilitation, answering common queries, and offering suggestions for physiotherapy and recovery after surgery[7].

These studies highlight ChatGPT's potential as a reliable and accessible educational resource, for both orthopaedic physicians and patients. They have demonstrated ChatGPT's unprecedented ability to generate accurate and, in many cases, excellent responses that align closely with expert recommendations across diverse orthopaedic surgery settings[8-10]. Yet, none looked at the references behind the provided orthopaedic information despite the warning by Alkaissi and McFarlane about "artificial hallucinations" in ChatGPT[11]. Indeed, the earlier model of ChatGPT has been observed to occasionally make up references despite its rapid processing and synthesis of vast information[11]. This tendency to generate plausible but factually inaccurate sources raises significant concerns about its reliability and safety in the field of orthopaedic surgery, which presents complex challenges owing to its various approaches to procedures, radiographical images, and clinical intricacies. While ChatGPT seems to have made its entrance into the world of orthopaedic surgery, ensuring the accuracy of its orthopaedic references is paramount, particularly when the credibility of the content relies vastly on the sources used.

This study aims to assess the ability of ChatGPT-4, the latest iteration of the GPT series, to provide accurate scientific references supporting the well-established Airway, Breathing, Circulation, Disability, Exposure (ABCDE) approach to trauma protocol. These guidelines constitute a cornerstone of trauma surgery and emergency medical care, including but not limited to orthopaedic trauma, and have been the gold standard of trauma care for decades[12]. Hence, we determined that the ABCDE approach would serve as a useful challenge for ChatGPT given the vast resources available on the subject on the internet.

Materials and Methods

ChatGPT-4

ChatGPT is a large language model that was developed by OpenAI (2022). Powered by billions of data parameters, ChatGPT can provide nearly instantaneous responses to even the most complex queries[13]. For the purpose of this study, the most advanced version of ChatGPT, version 4.0, also known as ChatGPT Plus or ChatGPT-4, was used to better gauge ChatGPT's ability and avoid underperformance.

Model Input

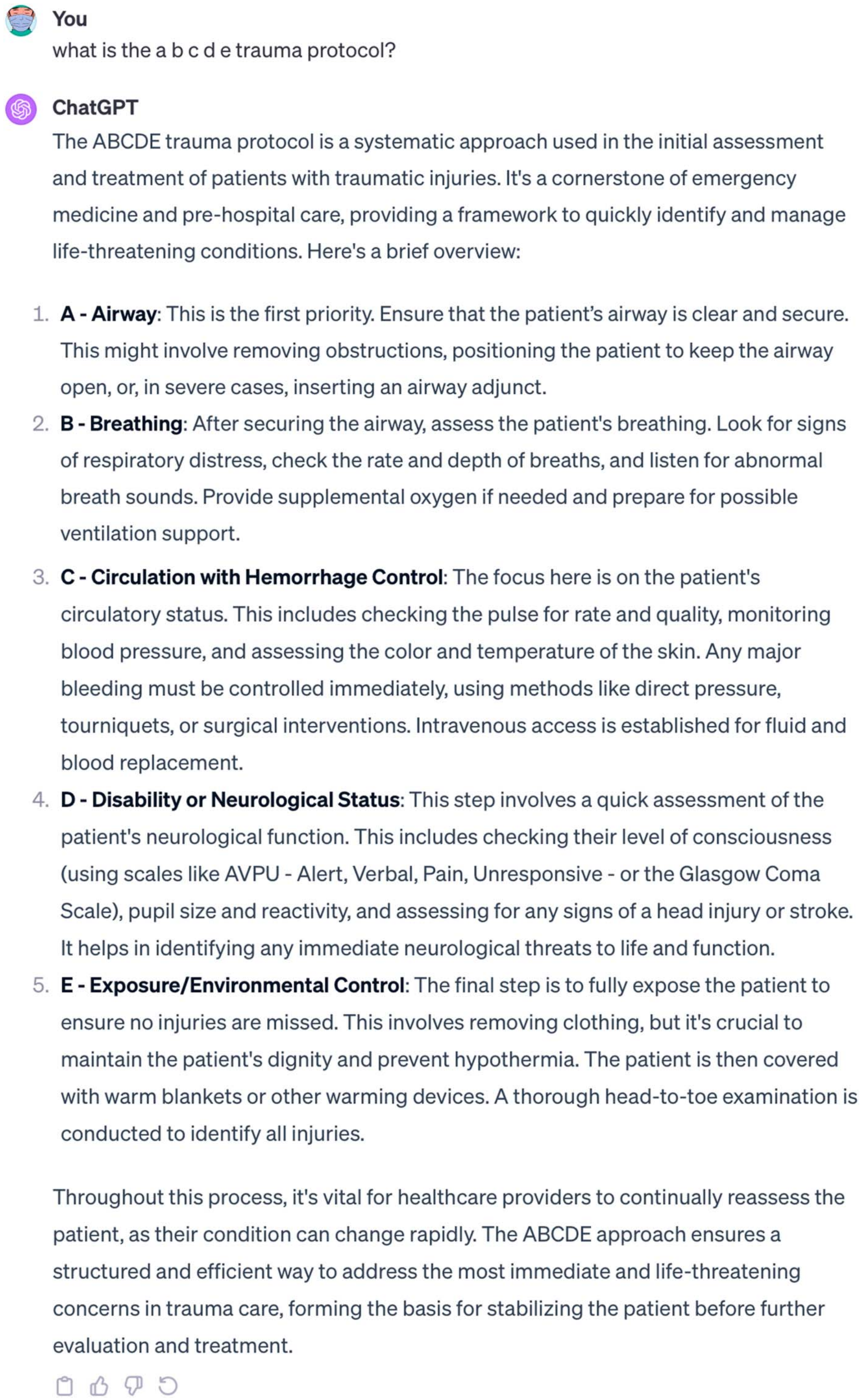
On November 27, 2023, we asked ChatGPT version 4.0 (https://chat.openai.com) about the ABCDE approach to trauma protocol (Fig. 1). The ChatGPT-4–generated answer was structured into 5 categories—(1) airway management, (2) breathing and ventilation, (3) circulation assessment, (4) disability or neurological status assessment, and (5) exposure and environmental control in trauma care. The content of each category was carefully reviewed and, expectedly, as per recent publications, deemed accurate and in alignment with the American College of Surgeons and American Academy of Orthopaedic Surgeons trauma guidelines. Then, ChatGPT-4 was asked to provide scientific references for each of the 5 steps (Fig. 2): "*Please provide PubMed-indexed references to support each of the 5 steps mentioned above.*" The task was considered complete after ChatGPT-4 had provided 6 references per category. We chose 6 references per category to ensure a larger and more representative sample size (n = 30) and a subsequent comprehensive assessment of ChatGPT-4's ability to provide accurate and relevant sources while maintaining a manageable scope for detailed analysis and verification. The 30 ChatGPT-generated references were recorded into a Microsoft Excel document 2023 (version 16.73; Microsoft) (Table I).

Assessment of References

Two independent reviewers—1 orthopaedic trauma surgeon and 1 orthopaedic trauma postdoctoral research fellow—classified the 30 references provided by ChatGPT-4 on a grading scale, from 0 (nonexistent reference) to 2 (accurate reference) (Table II). This was defined as the "reference accuracy score." A reference was considered "nonexistent" if the title, authors, journal, and year of publication did not match an existing reference and the reviewers were unable to find the publication online on PubMed, Google Scholar, Cochrane, Embase, or Scopus. "Inaccurate" references provided either false or incomplete information and required substantial modification—i.e., these references were either missing an author or had the wrong year of publication, title, volume, pages, or PubMed ID (PMID). "Accurate" references provided satisfactory information, in line with PubMed, and did not require any clarification. A reference was considered "true" only if it was deemed "accurate" with a perfect score of 2 and "false" if it was inaccurate (score of 1) or nonexistent (score of 0).

Statistical Analyses

Cohen's Kappa coefficient was used to examine the agreement of the accuracy scores of the ChatGPT-4–generated references between reviewers. Scores from both independent observers were used to calculate the mean ChatGPT-4 performance for



You
what is the a b c d e trauma protocol?

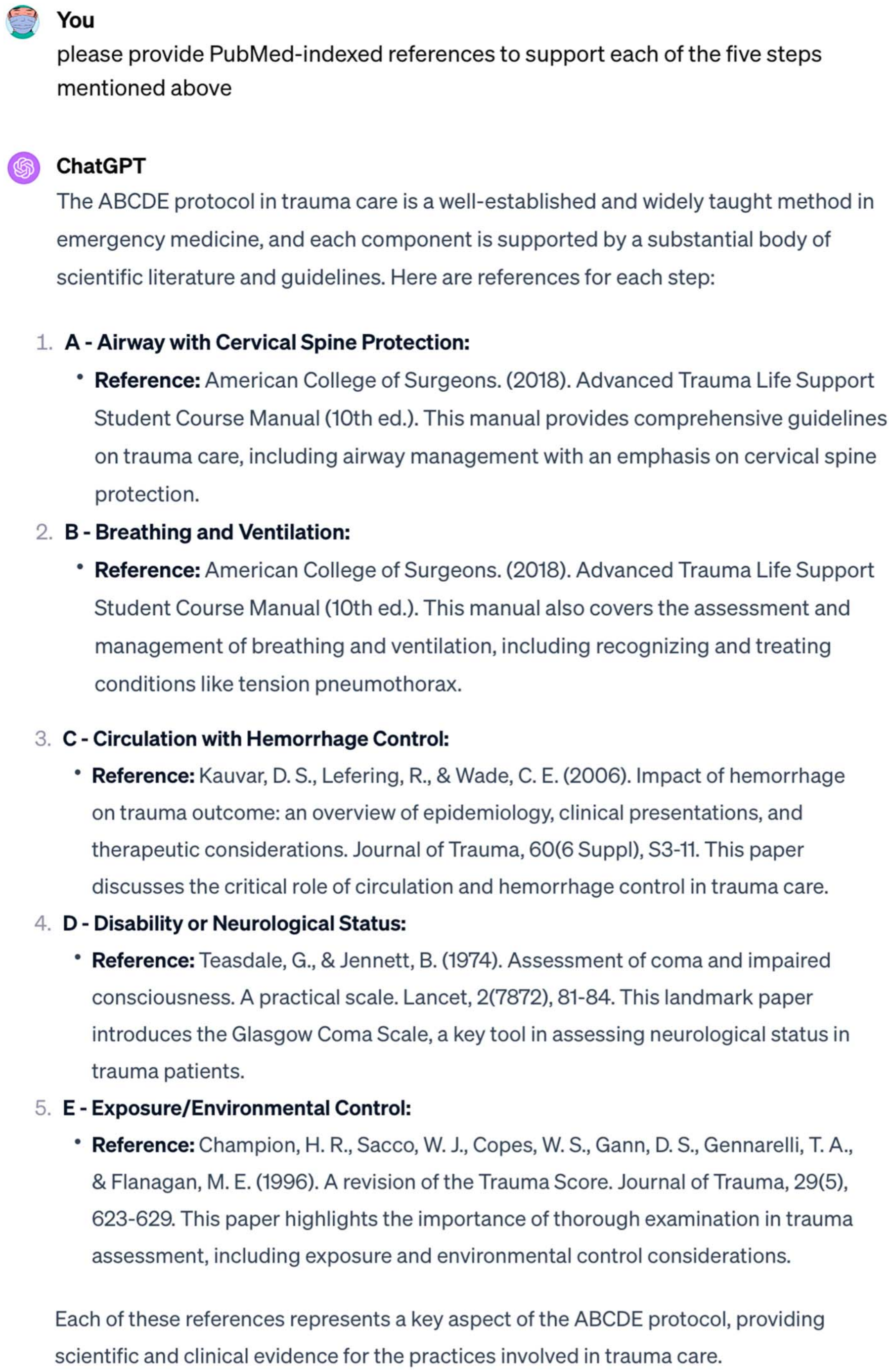
ChatGPT
The ABCDE trauma protocol is a systematic approach used in the initial assessment and treatment of patients with traumatic injuries. It's a cornerstone of emergency medicine and pre-hospital care, providing a framework to quickly identify and manage life-threatening conditions. Here's a brief overview:

1. **A - Airway**: This is the first priority. Ensure that the patient's airway is clear and secure. This might involve removing obstructions, positioning the patient to keep the airway open, or, in severe cases, inserting an airway adjunct.
2. **B - Breathing**: After securing the airway, assess the patient's breathing. Look for signs of respiratory distress, check the rate and depth of breaths, and listen for abnormal breath sounds. Provide supplemental oxygen if needed and prepare for possible ventilation support.
3. **C - Circulation with Hemorrhage Control**: The focus here is on the patient's circulatory status. This includes checking the pulse for rate and quality, monitoring blood pressure, and assessing the color and temperature of the skin. Any major bleeding must be controlled immediately, using methods like direct pressure, tourniquets, or surgical interventions. Intravenous access is established for fluid and blood replacement.
4. **D - Disability or Neurological Status**: This step involves a quick assessment of the patient's neurological function. This includes checking their level of consciousness (using scales like AVPU - Alert, Verbal, Pain, Unresponsive - or the Glasgow Coma Scale), pupil size and reactivity, and assessing for any signs of a head injury or stroke. It helps in identifying any immediate neurological threats to life and function.
5. **E - Exposure/Environmental Control**: The final step is to fully expose the patient to ensure no injuries are missed. This involves removing clothing, but it's crucial to maintain the patient's dignity and prevent hypothermia. The patient is then covered with warm blankets or other warming devices. A thorough head-to-toe examination is conducted to identify all injuries.

Throughout this process, it's vital for healthcare providers to continually reassess the patient, as their condition can change rapidly. The ABCDE approach ensures a structured and efficient way to address the most immediate and life-threatening concerns in trauma care, forming the basis for stabilizing the patient before further evaluation and treatment.

Fig. 1
ChatGPT-generated answer (version 4.0) regarding the ABCDE approach to trauma protocol. ABCDE = Airway, Breathing, Circulation, Disability, Exposure, and ChatGPT = Chat Generative Pretrained Transformer.

each reference, and the means were used for all statistical analyses. Descriptive statistics were used to summarize the mean reference accuracy of ChatGPT-4. One-way analysis of variance was used to compare the variance of means between the accuracy scores across the 5 categories. A p-value of 0.05 was set to determine statistical significance. Statistical tests and analyses were performed using R software (version 4.3.0.; R Foundation for Statistical Computing).



You

please provide PubMed-indexed references to support each of the five steps mentioned above

ChatGPT

The ABCDE protocol in trauma care is a well-established and widely taught method in emergency medicine, and each component is supported by a substantial body of scientific literature and guidelines. Here are references for each step:

1. **A - Airway with Cervical Spine Protection:**
   - **Reference:** American College of Surgeons. (2018). Advanced Trauma Life Support Student Course Manual (10th ed.). This manual provides comprehensive guidelines on trauma care, including airway management with an emphasis on cervical spine protection.
2. **B - Breathing and Ventilation:**
   - **Reference:** American College of Surgeons. (2018). Advanced Trauma Life Support Student Course Manual (10th ed.). This manual also covers the assessment and management of breathing and ventilation, including recognizing and treating conditions like tension pneumothorax.
3. **C - Circulation with Hemorrhage Control:**
   - **Reference:** Kauvar, D. S., Lefering, R., & Wade, C. E. (2006). Impact of hemorrhage on trauma outcome: an overview of epidemiology, clinical presentations, and therapeutic considerations. Journal of Trauma, 60(6 Suppl), S3-11. This paper discusses the critical role of circulation and hemorrhage control in trauma care.
4. **D - Disability or Neurological Status:**
   - **Reference:** Teasdale, G., & Jennett, B. (1974). Assessment of coma and impaired consciousness. A practical scale. Lancet, 2(7872), 81-84. This landmark paper introduces the Glasgow Coma Scale, a key tool in assessing neurological status in trauma patients.
5. **E - Exposure/Environmental Control:**
   - **Reference:** Champion, H. R., Sacco, W. J., Copes, W. S., Gann, D. S., Gennarelli, T. A., & Flanagan, M. E. (1996). A revision of the Trauma Score. Journal of Trauma, 29(5), 623-629. This paper highlights the importance of thorough examination in trauma assessment, including exposure and environmental control considerations.

Each of these references represents a key aspect of the ABCDE protocol, providing scientific and clinical evidence for the practices involved in trauma care.

Fig. 2

ChatGPT-generated scientific references to support each of the 5 steps of the ABCDE approach to the trauma protocol. ABCDE = Airway, Breathing, Circulation, Disability, Exposure, and ChatGPT = Chat Generative Pretrained Transformer.

Results

ChatGPT-4 had a mean reference accuracy score of 66.7% (ratio of the actual points awarded [40] to the potential perfect score [60], expressed as a percentage) (Table III). Excellent agreement was observed between raters for the 30 ChatGPT-4–generated references, with a Cohen's Kappa coefficient of 0.89. ChatGPT-4's overall accuracy was graded as “inaccurate providing false or incomplete information and requiring substantial modification” or “accurate requiring no modification” with a mean reference accuracy score ranging from 1 to 1.67 for each category (Table III). When assessing each of the 30 ChatGPT-generated references, 13 (43.3%)

TABLE I ChatGPT-Generated Title, Explanation, and References for Each Step of the Airway, Breathing, Circulation, Disability, Exposure Approach to Trauma

ChatGPT Title	ChatGPT Explanation	ChatGPT References
A: Airway		
Airway with cervical spine protection	This is the first and crucial step. The responder checks if the patient's airway is open and free from obstructions. It is vital to ensure the patient is able to breathe and that their cervical spine is protected, especially in cases of suspected spinal injuries. Techniques such as the head-tilt-chin-lift or jaw-thrust maneuver are used to open the airway if it is blocked. This step may also involve inserting airway adjuncts such as oropharyngeal or nasopharyngeal airways to keep the passage open	American College of Surgeons. (2018). Advanced Trauma Life Support Student Course Manual (10th ed)
		Carney N, Totten AM, O'Reilly C, et al. (2017). Guidelines for the Management of Severe Traumatic Brain Injury, Fourth Edition. *Neurosurgery*, 80(1), 6-15
		Walls RM, & Murphy MF. (2008). Manual of Emergency Airway Management (3rd ed). Philadelphia: Lippincott Williams & Wilkins
		Thiboutot F, Nicole PC, Trépanier CA, Turgeon AF, & Lessard BM. (2009). Prehospital Airway Management in the Acutely Injured Patient: The Role of Surgical Cricothyrotomy Revisited. *Journal of Trauma: Injury, Infection, and Critical Care*, 67(3), 583-586 [PubMed ID: 19741395]
		Sakles JC, Mosier JM, Patanwala AE, & Arcaris B[missing author]. (2016). First Pass Success Without Hypoxemia is Increased with the Use of Apneic Oxygenation During RSI in the Emergency Department. *Academic Emergency Medicine*, 23(6), 703-710 [PubMed ID: 26836726]
		Crewdson K, Lockey DJ, & Røislien J. (2017). The impact of pre-hospital intubation with and without sedation on outcome in trauma patients with a GCS of 8 or less. *Resuscitation*, 118, 80-86 [PubMed ID: 28736051]
B: Breathing		
Breathing and ventilation	The focus is on the patient's breathing effectiveness. The health care provider assesses the breathing rate, depth, and pattern. They look for symmetry in chest movements, listen for breath sounds on both sides, and check for signs of respiratory distress such as cyanosis (bluish skin due to lack of oxygen). Treatment at this stage may include administering supplemental oxygen, providing ventilation support, or performing procedures such as needle decompression if a tension pneumothorax (a type of collapsed lung) is suspected	American College of Surgeons. (2018). Advanced Trauma Life Support Student Course Manual (10th ed)
		Zideman DA, Singletary EM, De Buck EDJ, et al. (2015). European Resuscitation Council Guidelines for Resuscitation 2015: Section 9. First aid. *Resuscitation*, 95, 278-287
		Kortbeek JB, Al Turki SA, Ali J, et al. (2008). Advanced trauma life support, 8th edition, the evidence for change. *Journal of Trauma*, 64(6), 1638-1650
		Leigh-Smith S, & Harris T. (2005). Tension pneumothorax—time for a re-think? *Emergency Medicine Journal*, 22(1), 8-16 [PubMed ID: 15611534]

continued

TABLE I (*continued*)

ChatGPT Title	ChatGPT Explanation	ChatGPT References
		Galvagno SM, Nahmias JT, & Young DA. (2020). Advanced Trauma Life Support® Update 2019: Management and Applications for Adults and Special Populations. *Anesthesiology Clinics*, 38(2), 291-303 [PubMed ID: 32471636]
		Zeng J, Qian S, Zheng M, Wang Y, Zhou J, & Qiu H. (2019). The Epidemiology and Prehospital Management of Traumatic Pneumothorax. *Prehospital Emergency Care*, 23(4), 575-583 [PubMed ID: 30431398]
C: Circulation		
Circulation with hemorrhage control	This step involves evaluating the patient's circulatory status. It includes monitoring heart rate, blood pressure, and skin condition (color, temperature, and moisture). The assessment identifies signs of shock, such as rapid heart rate, low blood pressure, and cold, clammy skin. Controlling any external bleeding is a priority, which might involve applying pressure, tourniquets, or using hemostatic agents. Intravenous access is established for fluid and blood product administration if needed to maintain circulation	Kauvar DS, Lefering R, & Wade CE. (2006). Impact of hemorrhage on trauma outcome: an overview of epidemiology, clinical presentations, and therapeutic considerations. *Journal of Trauma*, 60(6 Suppl), S3-11
		Holcomb JB, Tilley BC, Baraniuk S, et al. (2015). Transfusion of plasma, platelets, and red blood cells in a 1:1:1 vs a 1:1:2 ratio and mortality in patients with severe trauma: The PROPPR randomized clinical trial. *JAMA*, 313(5), 471-482
		Pusateri AE, Weiskopf RB, Bebarta V, et al. (2013). Tranexamic acid and trauma: current status and knowledge gaps with recommended research priorities. *Shock*, 39(2), 121-126
		Kragh JFJr, Walters TJ, Baer DG, Fox CJ, Wade CE, Salinas J, Holcomb JB. (2009). Practical use of emergency tourniquets to stop bleeding in major limb trauma. *Journal of Trauma*, 64(2 Suppl), S38-49; discussion S49-50 [PubMed ID: 19177521]
		Cannon JW, Khan MA, Raja AS, Cohen MJ, Como JJ, Cotton BA, & Yelon JA. (2017). Damage control resuscitation in patients with severe traumatic hemorrhage. *Journal of Trauma and Acute Care Surgery*, 82(3), 605-617 [PubMed ID: 28030486]
		Moore HB, Moore EE, Chapman MP, McVaney K, Bryskiewicz G, Blechar R & Silliman CC. (2018). Plasma-first resuscitation to treat haemorrhagic shock during emergency ground transportation in an urban area: a randomised trial. *The Lancet*, 392(10144), 283-291 [PubMed ID: 30032977]
D: Disability		
Disability or neurological status	In this phase, the patient's neurological function is evaluated. This includes assessing the level of consciousness, usually with the Glasgow Coma Scale, which scores eye, verbal, and motor responses. Pupil size and reaction are also checked. This step helps in identifying any immediate neurological impairments that may need urgent attention	Teasdale G, & Jennett, B. (1974). Assessment of coma and impaired consciousness. A practical scale. *Lancet*, 2(7872), 81-84

continued

TABLE I (*continued*)

ChatGPT Title	ChatGPT Explanation	ChatGPT References
		Ropper AH. (2014)Traumatic brain injury-A review. *New England Journal of Medicine*, 370(10), 898-901
		Marion DW, & Carlier PM. (1994). Problems with initial Glasgow Coma Scale assessment caused by prehospital treatment of patients with head injuries: results of a national survey. *Journal of Trauma*, 36(1), 89-95
		Majdan M, Mauritz W, Wilbacher I, Brazinova A, Rusnak M, & Leitgeb J. (2013). Glasgow Coma Scale score at intensive care unit discharge predicts the 1-year outcome of patients with severe traumatic brain injury. *European Journal of Trauma and Emergency Surgery*, 39(3), 285-292 [PubMed ID: 26815230]
		Spaite DW, [missing author], Bobrow BJ, Keim SM, Barnhart B, Chikani V, Gaither JB, & Vadeboncoeur TF (2017). Association of out-of-hospital hypotension depth and duration with traumatic brain injury outcomes. *JAMA Neurology*, 74(10), 1206-1214 [PubMed ID: 28846763]
		Sun M, Meurer WJ, Domeier RM, Shields JF, & Layton KF. (2017). The cause of death in trauma: A study in brain injured patients who died during emergency medical rescue. *Prehospital Emergency Care*, 21(5), 583-590 [PubMed ID: 28394703]
E: Exposure		
Exposure/ environmental control	Finally, the patient is fully exposed for a thorough physical examination to ensure no injuries are missed. However, it is essential to maintain the patient's privacy and protect them from environmental conditions. For instance, the patient should be covered with blankets to prevent hypothermia, which is a risk, especially in a cold environment or if the patient is in shock	Champion HR, Sacco WJ, Copes WS, Gann DS, Gennarelli TA, & Flanagan, ME. (1996). A revision of the Trauma Score. *Journal of Trauma*, 29(5), 623-629
		Gentilello LM, Jurkovich GJ, Stark MS, Hassantash SA, & O'Keefe GE. (2002). Is hypothermia in the victim of major trauma protective or harmful? A randomized, prospective study. *Annals of Surgery*, 236(4), 492-500
		Seekamp A, Regel G, Tscherne H. (1998). Hypothermia in patients with multiple injuries. *Injury*, 29 Suppl 2, B7-12
		Lapostolle F, Sebbah JL, Couvreur J, Koch FX, Savary D, Tazarourte K, & Adnet, F. (2007). Risk [factors for onset] of hypothermia in trauma victims: the importance of the prehospital phase. *Prehospital Emergency Care*, 11(4), 460-467 [PubMed ID: 17907033]
		Ireland S, Endacott R, Cameron P, Fitzgerald M, & Paul E. (2018). The incidence and significance of accidental hypothermia in major trauma–A prospective observational study. *Resuscitation*, 122, 11-17 [PubMed ID: 29141153]
		Søreide K. Clinical and translational aspects of hypothermia in major trauma patients: from pathophysiology to prevention, prognosis and potential preservation. *Injury*. 2014;45(4):647-654. doi:10.1016/ j.injury.2012.12.023. PMID: 23352151

TABLE II Standardized Grading System Adopted for the Assessment of ChatGPT-Generated References

Reference Accuracy Score	Reference Accuracy Description
0	Nonexistent reference
1	Inaccurate reference providing false or incomplete information and requiring substantial modification
2	Accurate reference requiring no modification

were considered "accurate" with a perfect score of 2 and were hence categorized as "true." 17 (56.7%) were categorized as "false": 13 (43.3%) were considered "inaccurate" with a score of 1 and 4 (13.3%) were found to be "nonexistent," hence receiving a score of 0 (Table III, Fig. 3). For example, in the "airway" category, ChatGPT-4 provided a reference with an accurate title and year and journal of publication, yet was missing an author name and had the wrong PMID number. In other instances, such as in "breathing and ventilation," the ChatGPT-4–generated reference had the correct title and authorship but a different year of publication, journal volume, and associated pages, as well as PMID. Such references were considered inaccurate and were attributed a score of 1 (Table III). Across the 5 different categories, ChatGPT-4 scored an average of 1.5 on "airway management," 1.42 on "breathing and ventilation," 1.67 on "circulation assessment," 1 on "disability or neurological status assessment," and 1.08 on "exposure and environmental control." There was no statistically significant difference in means between ChatGPT-4's references' accuracy across the categories ($p = 0.437$) (Table III).

Discussion

This is the first study to assess the accuracy of ChatGPT-4–generated references in orthopaedic surgery. It encompassed 30 references associated with the ABCDE trauma protocol. While the ChatGPT-generated content about the trauma protocol closely aligned with the current literature and recommendations, the associated references were more often than not nonexistent or inaccurate. In fact, the AI language model ChatGPT-4 performed below average, with more than half of the references considered "false" and an average reference accuracy score of only 66.7% as judged by our independent reviewers.

It seems that, when pressed for references, ChatGPT-4 provides what researchers in the generative AI field refer to as a "hallucination," fabricating a bibliographic citation that is plausible but does not correspond to an actual scholarly work. In this study, such errors ranged from incorrect PMIDs and year of publication to fabricated author lists and titles. Such erroneous references act as a cautionary signal to all ChatGPT users, including orthopaedic experts, who are considering incorporating ChatGPT into their medical decision making. If the references used to generate answers are flawed, it may compromise the quality and trustworthiness of the information provided by AI models. These findings further underscore OpenAI's disclaimer found under each ChatGPT chat: "ChatGPT can make mistakes. Consider checking important information."

When stratifying the references based on the different steps of the trauma protocol, there was no statistically significant difference between the accuracy of the references. ChatGPT demonstrated consistency in providing equally as accurate or inaccurate references across topics. This suggests that the ChatGPT-generated references are merely the result of its software internal workflow, which might rely more on pattern recognition and emulation, rather than actual literature search and review.

While the few studies that looked at the accuracy of ChatGPT-generated references were performed in other fields of medicine, such as rheumatology, plastic surgery, anesthesia, and radiology, the scholars have raised similar concerns about the language model's tendency to cite works that do not actually exist[14-17]. Compared with previous studies on medical references provided by ChatGPT, our findings also demonstrate the presence of fabricated citations (56.7%), albeit at different rates compared with the most recent publication by Bhattacharyya et al. (93%), and a higher percentage of authentic references (43.3% vs 7%, respectively)[18]. This much-improved performance, albeit insufficient, might be attributed to the fact that our study relied on the more advanced ChatGPT-4 instead of prior versions including ChatGPT 3.5 and earlier[13,18]. This version was chosen due to its decreased hallucination effect, superior performance and consistency despite its monthly subscription cost, and its longer processing time to generate a response. In addition, our study focused on the ABCDE approach to trauma—a very well-established topic with clear guidelines and abundant online references.

It is imperative to recognize that the reference accuracy scores reported in this study are inherently tied to the grading system adopted by our methodology. While the predefined criteria for evaluating the veracity and relevance of ChatGPT-4's references have been set per our research team's good judgment and latest review of the literature, it is important to acknowledge the element of subjectivity in such a grading scheme. Different researchers, using alternative grading scales or criteria that might prioritize different aspects of reference accuracy or relevance, could potentially report a different performance. Therefore, while our findings provide valuable insights into the reliability of ChatGPT-4's references within the context of orthopaedic trauma surgery, they should be interpreted with an understanding that the results are influenced by the specific evaluative framework used. Using the current grading scale, there was excellent agreement between our 2 independent observers, indicating a high level of consistency in their evaluations. This finding suggests high validity, increased reliability, and decreased bias. Indeed, the dual-reviewer approach underscores the study's commitment to rigor and accuracy in

TABLE III Accuracy Scores of ChatGPT-Generated References as Compared With Their Associated PubMed References

ChatGPT Title	ChatGPT References	PubMed References	Mean Accuracy Score*	Average Score	p-value*
Airway with cervical spine protection	American College of Surgeons. (2018). Advanced Trauma Life Support Student Course Manual (10th ed)	—	2	1.5	0.437
	Carney N, Totten AM, O'Reilly C, et al. (2017). Guidelines for the Management of Severe Traumatic Brain Injury, Fourth Edition. *Neurosurgery*, 80(1), 6-15	Carney N, Totten AM, O'Reilly C, et al. Guidelines for the Management of Severe Traumatic Brain Injury, Fourth Edition. *Neurosurgery*. 2017;80(1):6-15. doi:10.1227/NEU.0000000000001432	2		
	Walls RM & Murphy MF. (2008). Manual of Emergency Airway Management (3rd ed). Philadelphia: Lippincott Williams & Wilkins	—	2		
	Thiboutot F, Nicole PC, Trépanier CA, Turgeon AF, & Lessard BM. (2009). Prehospital Airway Management in the Acutely Injured Patient: The Role of Surgical Cricothyrotomy Revisited. *Journal of Trauma: Injury, Infection, and Critical Care*, 67(3), 583-586. [PubMed ID: 19741395]	Gerich TG, Schmidt U, Hubrich V, Lobenhoffer HP, Tscherne H. Prehospital airway management in the acutely injured patient: the role of surgical cricothyrotomy revisited. *J Trauma*. 1998;45(2):312-314. doi:10.1097/00005373-199808000-00017; PMID: 9715188	1		
	Sakles JC, Mosier JM, Patanwala AE, & Arcaris B. [missing author] (2016). First Pass Success Without Hypoxemia is Increased with the Use of Apneic Oxygenation During RSI in the Emergency Department. *Academic Emergency Medicine*, 23(6), 703-710. [PubMed ID: 26836726]	Sakles JC, Mosier JM, Patanwala AE, Arcaris B, Dicken JM. First Pass Success Without Hypoxemia Is Increased With the Use of Apneic Oxygenation During Rapid Sequence Intubation in the Emergency Department. *Acad Emerg Med*. 2016;23(6):703-710. doi:10.1111/acem.12931; PMID: 26836712	1		
	Crewdson K, Lockey DJ, & Røislien J. (2017). The impact of pre-hospital intubation with and without sedation on outcome in trauma patients with a GCS of 8 or less. *Resuscitation*, 118, 80-86. [PubMed ID: 28736051]	Hoffmann M, Czorlich P, Lehmann W, et al. The Impact of Prehospital Intubation With and Without Sedation on Outcome in Trauma Patients With a GCS of 8 or Less. *J Neurosurg Anesthesiol*. 2017;29(2):161-167. doi:10.1097/ANA.0000000000000275; PMID: 26797107	1		
Breathing and ventilation	American College of Surgeons. (2018). Advanced Trauma Life Support Student Course Manual (10th ed)		2	1.42	
	Zideman DA, Singletary EM, De Buck EDJ, et al. (2015). European Resuscitation Council Guidelines for Resuscitation 2015: Section 9. First aid. *Resuscitation*, 95, 278-287	Zideman DA, De Buck ED, Singletary EM, et al. European Resuscitation Council Guidelines for Resuscitation 2015 Section 9. First aid. *Resuscitation*. 2015;95:278-287. doi:10.1016/j.resuscitation.2015.07.031	1.5		
	Kortbeek JB, Al Turki SA, Ali J, et al. (2008). Advanced trauma life support, 8th edition, the evidence for change. *Journal of Trauma*, 64(6), 1638-1650	Kortbeek JB, Al Turki SA, Ali J, et al. Advanced trauma life support, 8th edition, the evidence for change. *J Trauma*. 2008;64(6):1638-1650. doi:10.1097/TA.0b013e3181744b03	2		
	Leigh-Smith S, & Harris T. (2005). Tension pneumothorax—time for a re-think? *Emergency Medicine Journal*, 22(1), 8-16. [PubMed ID: 15611534]	Leigh-Smith S, Harris T. Tension pneumothorax: time for a re-think? *Emerg Med J*. 2005 Jan;22(1):8-16. doi: 10.1136/emj.2003.010421. PMID: 15611534; PMCID: PMC1726546	2		

continued

TABLE III (*continued*)

ChatGPT Title	ChatGPT References	PubMed References	Mean Accuracy Score*	Average Score	p-value*
	Galvagno SM, Nahmias JT, & Young DA. (2020). Advanced Trauma Life Support® Update 2019: Management and Applications for Adults and Special Populations. *Anesthesiology Clinics*, 38(2), 291-303. [PubMed ID: 32471636]	Galvagno SM Jr, Nahmias JT, Young DA. Advanced Trauma Life Support® Update 2019: Management and Applications for Adults and Special Populations. *Anesthesiol Clin*. 2019; 37(1):13-32. doi:10.1016/j.anclin.2018.09.009; PMID: 30711226	1		
	Zeng J, Qian S, Zheng M, Wang Y, Zhou J, & Qiu H. (2019). The Epidemiology and Prehospital Management of Traumatic Pneumothorax. *Prehospital Emergency Care*, 23(4), 575-583. [PubMed ID: 30431398]	Alghnam S, Aldahnim MH, Aldebasi MH, et al. The incidence and predictors of pneumothorax among trauma patients in Saudi Arabia. Findings from a level-I trauma center. *Saudi Med J*. 2020;41(3):247-252. doi:10.15537/smj.2020.3.24917; PMID: 32114596	0		
Circulation with hemorrhage control	Kauvar DS, Lefering R, & Wade CE. (2006). Impact of hemorrhage on trauma outcome: an overview of epidemiology, clinical presentations, and therapeutic considerations. *Journal of Trauma*, 60(6 Suppl), S3-11	Kauvar DS, Lefering R, Wade CE. Impact of hemorrhage on trauma outcome: an overview of epidemiology, clinical presentations, and therapeutic considerations. *J Trauma*. 2006;60(6 Suppl):S3-S11. doi:10.1097/01.ta.0000199961.02677.19	2	1.67	
	Holcomb JB, Tilley BC, Baraniuk S, et al. (2015). Transfusion of plasma, platelets, and red blood cells in a 1:1:1 vs a 1:1:2 ratio and mortality in patients with severe trauma: The PROPPR randomized clinical trial. *JAMA*, 313(5), 471-482	Holcomb JB, Tilley BC, Baraniuk S, et al. Transfusion of plasma, platelets, and red blood cells in a 1:1:1 vs a 1:1:2 ratio and mortality in patients with severe trauma: the PROPPR randomized clinical trial. *JAMA*. 2015;313(5):471-482. doi:10.1001/jama.2015.12	2		
	Pusateri AE, Weiskopf RB, Bebarta V, et al. (2013). Tranexamic acid and trauma: current status and knowledge gaps with recommended research priorities. *Shock*, 39(2), 121-126	Pusateri AE, Weiskopf RB, Bebarta V, et al. Tranexamic acid and trauma: current status and knowledge gaps with recommended research priorities. *Shock*. 2013;39(2):121-126. doi:10.1097/SHK.0b013e318280409a	2		
	Kragh JF Jr, Walters TJ, Baer DG, Fox CJ, Wade CE, Salinas J, Holcomb JB. (2009). Practical use of emergency tourniquets to stop bleeding in major limb trauma. *Journal of Trauma*, 64(2 Suppl), S38-49; discussion S49-50. [PubMed ID: 19177521]	Kragh JF Jr, Walters TJ, Baer DG, et al. Practical use of emergency tourniquets to stop bleeding in major limb trauma. *J Trauma*. 2008;64(2 Suppl):S38-S50. doi:10.1097/TA.0b013e31816086b1; PMID: 18376170	1		
	Cannon JW, Khan MA, Raja AS, Cohen MJ, Como JJ, Cotton BA, & Yelon JA. (2017). Damage control resuscitation in patients with severe traumatic hemorrhage. [incomplete title] *Journal of Trauma and Acute Care Surgery*, 82(3), 605-617. [PubMed ID: 28030486]	Cannon JW, Khan MA, Raja AS, et al. Damage control resuscitation in patients with severe traumatic hemorrhage: A practice management guideline from the Eastern Association for the Surgery of Trauma. *J Trauma Acute Care Surg*. 2017; 82(3):605-617. doi:10.1097/TA.0000000000001333; PMID: 28225743	1		
	Moore HB, Moore EE, Chapman MP, McVaney K, Bryskiewicz G, Blechar R & Silliman CC (2018). Plasma-first resuscitation to treat haemorrhagic shock during emergency ground transportation in an urban area: a randomised trial. *The Lancet*, 392(10144), 283-291. [PubMed ID: 30032977]	Moore HB, Moore EE, Chapman MP, et al. Plasma-first resuscitation to treat haemorrhagic shock during emergency ground transportation in an urban area: a randomised trial. *Lancet*. 2018;392(10144):283-291. doi:10.1016/S0140-6736(1831553-8)	2		

continued

TABLE III (*continued*)

ChatGPT Title	ChatGPT References	PubMed References	Mean Accuracy Score*	Average Score	p-value*
Disability or neurological status	Teasdale G, & Jennett B. (1974). Assessment of coma and impaired consciousness. A practical scale. *Lancet*, 2(7872), 81-84	Teasdale G, Jennett B. Assessment of coma and impaired consciousness. A practical scale. *Lancet*. 1974;2(7872):81-84. doi:10.1016/s0140-6736(7491639-0)	2	1	
	Ropper AH. (2014). Traumatic brain injury–A review. *New England Journal of Medicine*, 370(10), 898-901	Crooks CY, Zumsteg JM, Bell KR. Traumatic brain injury: a review of practice management and recent advances. *Phys Med Rehabil Clin N Am*. 2007;18(4):681-vi. doi:10.1016/j.pmr.2007.06.005	0		
	Marion DW, & Carlier PM. (1994). Problems with initial Glasgow Coma Scale assessment caused by prehospital treatment of patients with head injuries: results of a national survey. *Journal of Trauma*, 36(1), 89-95	Marion DW, Carlier PM. Problems with initial Glasgow Coma Scale assessment caused by prehospital treatment of patients with head injuries: results of a national survey. *J Trauma*. 1994;36(1):89-95. doi:10.1097/00005373-199401000-00014	2		
	Majdan M, Mauritz W, Wilbacher I, Brazinova A, Rusnak M, & Leitgeb J. (2013). Glasgow Coma Scale score at intensive care unit discharge predicts the 1-year outcome of patients with severe traumatic brain injury. *European Journal of Trauma and Emergency Surgery*, 39(3), 285-292. [PubMed ID: 26815230]	Leitgeb J, Mauritz W, Brazinova A, et al. Glasgow Coma Scale score at intensive care unit discharge predicts the 1-year outcome of patients with severe traumatic brain injury. *Eur J Trauma Emerg Surg*. 2013;39(3):285-292. doi:10.1007/s00068-013-0269-3; PMID: 23762202	1		
	Spaite DW, [missing author], Bobrow BJ, Keim SM, Barnhart B, Chikani V, Gaither JB, & Vadeboncoeur TF. (2017). Association of out-of-hospital hypotension depth and duration with traumatic brain injury outcomes. *JAMA Neurology*, 74(10), 1206-1214. [PubMed ID: 28846763]	Spaite DW, Hu C, Bobrow BJ, et al. Association of Out-of-Hospital Hypotension Depth and Duration With Traumatic Brain Injury Mortality. *Ann Emerg Med*. 2017;70(4):522-530.e1. doi:10.1016/j.annemergmed.2017.03.027; PMID: 28559036	1		
	Sun M, Meurer WJ, Domeier RM, Shields JF, & Layton KF. (2017). The cause of death in trauma: A study in brain injured patients who died during emergency medical rescue. *Prehospital Emergency Care*, 21(5), 583-590. [PubMed ID: 28394703]	Goldberg SA, Rojanasarntikul D, Jagoda A. The prehospital management of traumatic brain injury. *Handb Clin Neurol*. 2015;127:367-378. doi:10.1016/B978-0-444-52892-6.00023-4	0		
Exposure/ environmental control	Champion HR, Sacco WJ, Copes WS, Gann DS, Gennarelli TA, & Flanagan ME. (1996). A revision of the Trauma Score. *Journal of Trauma*, 29(5), 623-629	Champion HR, Sacco WJ, Copes WS, Gann DS, Gennarelli TA, Flanagan ME. A revision of the Trauma Score. *J Trauma*. 1989;29(5):623-629. doi:10.1097/00005373-198905000-00017	1.5	1.08	
	Gentilello LM, Jurkovich GJ, Stark MS, Hassantash SA, & O'Keefe GE. (2002). Is hypothermia in the victim of major trauma protective or harmful? A randomized, prospective study. *Annals of Surgery*, 236(4), 492-500	Gentilello LM, Jurkovich GJ, Stark MS, Hassantash SA, O'Keefe GE. Is hypothermia in the victim of major trauma protective or harmful? A randomized, prospective study. *Ann Surg*. 1997 Oct;226(4):439-47; discussion 447-9. doi: 10.1097/00000658-199710000-00005. PMID: 9351712; PMCID: PMC1191057	1		
	Seekamp A, Regel G, Tscherne H. (1998). Hypothermia in patients with multiple injuries. *Injury*, 29 Suppl 2, B7-12	Segers MJ, Diephuis JC, van Kesteren RG, van der Werken C. Hypothermia in trauma patients. *Unfallchirurg*. 1998;101(10):742-749	0		

continued

TABLE III (*continued*)

ChatGPT Title	ChatGPT References	PubMed References	Mean Accuracy Score*	Average Score	p-value*
	Lapostolle F, Sebbah JL, Couvreur J, Koch FX, Savary D, Tazarourte K & Adnet F. (2007). Risk [factors for onset] of hypothermia in trauma victims: the importance of the prehospital phase. *Prehospital Emergency Care*, 11(4), 460-467. [PubMed ID: 17907033]	Lapostolle F, Sebbah JL, Couvreur J, et al. Risk factors for onset of hypothermia in trauma victims: the HypoTraum study. *Crit Care*. 2012;16(4):R142. Published 2012 Jul 31. doi: 10.1186/cc11449; PMID: 22849694	1		
	Ireland S, Endacott R, Cameron P, Fitzgerald M, & Paul E. (2018). The incidence and significance of accidental hypothermia in major trauma–A prospective observational study. *Resuscitation*, 122, 11-17. [PubMed ID: 29141153]	Ireland S, Endacott R, Cameron P, Fitzgerald M, Paul E. The incidence and significance of accidental hypothermia in major trauma: a prospective observational study. *Resuscitation.* 2011;82(3):300-306. doi:10.1016/ j.resuscitation.2010.10.016; PMID: 21074927	1		
	Søreide K. Clinical and translational aspects of hypothermia in major trauma patients: from pathophysiology to prevention, prognosis and potential preservation. *Injury*. 2014;45(4):647-654. doi:10.1016/j.injury.2012.12.023. PMID: 23352151	Søreide K. Clinical and translational aspects of hypothermia in major trauma patients: from pathophysiology to prevention, prognosis and potential preservation. *Injury*. 2014;45(4):647-654. doi:10.1016/j.injury.2012.12.027	2		

*Mean accuracy score represents the average of the grades provided by the 2 independent reviewers. †p-value calculated using one-way analysis of variance.

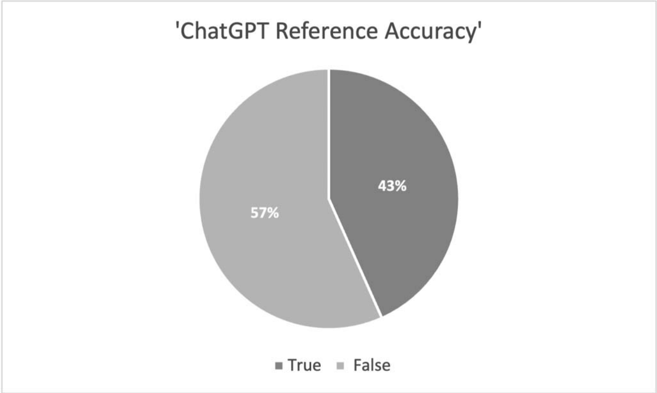


Fig. 3

Pie chart showing the accuracy of the ChatGPT-generated references categorized by "true" or "false." ChatGPT = Chat Generative Pretrained Transformer.

evaluating the ChatGPT-generated responses, which is particularly important in the setting of medical education and decision making, where the reliability of reference materials can significantly affect clinical outcomes and educational integrity.

Several limitations should be acknowledged in our study. First, ChatGPT is a text-based AI in which the generated output is often only as good as the provided input. While this might also explain the different accuracy performances across studies, it also suggests that an experienced ChatGPT user might be closer to successfully gauging ChatGPT's performance compared with a one-time user. Thus, comparing and reproducing studies might be difficult as each investigator inputs a different prompt. The lead investigator has led multiple ChatGPT studies and thus has significant experience with the AI language model, which may have contributed to its superior performance when compared with other studies. Second, ChatGPT-4 has recently gained real-time internet access as of September 2023 and can now constantly modify its information inventory[19]. Therefore, the exact same query might yield a different answer depending on the date and iteration of the model that is accessed. Currently, ChatGPT-4's latest update goes back to April 2023. This suggests that as improvements and updates arise, studies might find an enhanced efficacy and reliability of ChatGPT-4 with higher performance scores. Third, because ChatGPT's training data are not public, it is unclear which and how journal articles and webpages contributed to ChatGPT-4's answer to the prompt and fabricated references[20]. This is further restricted by ChatGPT's use of publicly available external information, which often excludes high-impact journals due to subscription fees. Finally, OpenAI cautions that the ChatGPT model can rapidly produce responses that sound believable but may be either incorrect or nonsensical[13,21]. This is a known issue with ChatGPT and similar large language models, which sometimes generate fabricated information to back their statements. Therefore, users must possess a certain degree of prior knowledge about the topic at hand to identify potential misinformation. At this time, pattern recognition embedded in ChatGPT cannot replace critical review of the literature by physicians. Nonetheless, the engaging and comprehensive replies offered by ChatGPT can be valuable for knowledgeable users, especially when they verify the information against reliable and established sources.

This is the first published evaluation of ChatGPT-4–generated references of the ABCDE approach to trauma by an orthopaedic investigator. With 57% of references being inaccurate or nonexistent, ChatGPT-4 has fallen short in providing reliable and reproducible references. This "hallucination" makes it hard for orthopaedic surgeons to rely solely on ChatGPT-4. Only if used cautiously, with cross-referencing, can this large language model act as a great adjunct learning tool that can enhance comprehensiveness as well as knowledge rehearsal and manipulation. We encourage future studies to explore ways to improve the capabilities of such AI systems to reference relevant literature in the field of orthopaedic surgery to support their embedded data and provide health care workers with a reliable interactive medical resource. ■

Diane Ghanem, MD[1]
Alexander R. Zhu, BA[2]
Whitney Kagabo, MD[1]
Greg Osgood, MD, FAOA[1]
Babar Shafiq, MD, FAOA[1]

[1]Department of Orthopaedic Surgery, The Johns Hopkins Hospital, Baltimore, Maryland

[2]School of Medicine, The Johns Hopkins University, Baltimore, Maryland

E-mail address for D. Ghanem: dghanem1@jh.edu

References

1. Lee P, Bubeck S, Petro J. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. New Engl J Med. 2023;388(13):1233-9.

2. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, Madriaga M, Aggabao R, Diaz-Candido G, Maningo J, Tseng V. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. PLOS Digit Health. 2023;2(2):e0000198.

3. Kung JE, Marshall C, Gauthier C, Gonzalez TA, Jackson JB. Evaluating ChatGPT performance on the orthopaedic in-training examination. JBJS Open Access. 2023; 8(3):e23.00056.

4. Ghanem D, Covarrubias O, Raad M, LaPorte D, Shafiq B. ChatGPT performs at the level of a third-year orthopaedic surgery resident on the orthopaedic in-training examination. JBJS Open Access. 2023;8(4):e23.00103.

5. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and Valid concerns. Healthcare. 2023;11(6):887.

6. Kuroiwa T, Sarcon A, Ibara T, Yamada E, Yamamoto A, Tsukamoto K, Fujita K. The potential of ChatGPT as a Self-Diagnostic tool in common orthopedic diseases: exploratory study. J Med Internet Res. 2023;25:e47621. doi.

7. Ghanem D, Shu H, Bergstein V, Marrache M, Love A, Hughes A, Sotsky R, Shafiq B. Educating patients on osteoporosis and bone health: can "ChatGPT" provide high-quality content? Eur J Orthop Surg Traumatol. 2024;34(5):2757-65.

8. Draschl A, Hauer G, Fischerauer SF, Kogler A, Leitner L, Andreou D, Leithner A, Sadoghi P. Are ChatGPT's free-text responses on periprosthetic joint infections of the hip and knee reliable and useful? J Clin Med. 2023;12(20):6655.

9. Anastasio AT, Mills FB, Karavan MP, Adams SB. Evaluating the quality and usability of artificial intelligence–generated responses to common patient questions in foot and ankle surgery. Foot Ankle Orthop. 2023;8(4):24730114231209919.

10. Chatterjee S, Bhattacharya M, Pal S, Lee SS, Chakraborty C. ChatGPT and large language models in orthopedics: from education and surgery to research. J Exp orthopaedics. 2023;10(1):128.

11. Alkaissi H, McFarlane SI. Artificial hallucinations in ChatGPT: implications in scientific writing. Cureus. 2023;15(2):e35179.

12. Thim T, Krarup NH, Grove EL, Rohde CV, Lofgren B. Initial assessment and treatment with the Airway, breathing, circulation, disability, exposure (ABCDE) approach. Int J Gen Med. 2012;117:117.

13. OpenAI. GPT-4 technical report. 2023. Available at: http://arxiv.org/abs/2303.08774. Accessed December 20, 2023.

14. Hueber AJ, Kleyer A. Quality of citation data using the natural language processing tool ChatGPT in rheumatology: creation of false references. RMD Open. 2023;9(2):e003248.

15. Xie Y, Seth I, Rozen WM, Hunter-Smith DJ. Evaluation of the artificial intelligence chatbot on breast reconstruction and its efficacy in surgical research: a case study. Aesthet Plast Surg. 2023;47(6):2360-9.

16. De Cassai A, Dost B. Concerns regarding the uncritical use of ChatGPT: a critical analysis of AI-generated references in the context of regional anesthesia. Reg Anesth Pain Med. 2024;49(5):378-80.

17. Wagner MW, Ertl-Wagner BB. Accuracy of information and references using ChatGPT-3 for retrieval of clinical radiological information. Can Assoc Radiol J. 2024; 75(1):69-73.

18. Bhattacharyya M, Miller VM, Bhattacharyya D, Miller LE. High rates of fabricated and inaccurate references in ChatGPT-generated medical content. Cureus. 2023; 15(5):e39238.

19. OpenAI. ChatGPT Plugins. Available at: https://openai.com/blog/chatgpt-plugins#safety-considerations. Accessed December 27, 2023.

20. Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, Neelakantan A, Shyam P, Sastry G, Askell A, Agarwal S, Herbert-Voss A, Krueger G, Henighan T, Child R, Ramesh A, Ziegler DM, Wu J, Winter C, Hesse C, Chen M, Sigler E, Litwin M, Gray S, Chess B, Clark J, Berner C, McCandlish S, Radford A, Sutskever I, Amodei D. Language models are few-shot learners; 2020. Available at: http://arxiv.org/abs/2005.14165. Accessed December 26, 2023.

21. OpenAI. Introducing ChatGPT. Available at: https://openai.com/blog/chatgpt. Accessed December 26, 2023.