ORIGINAL ARTICLE

Can ChatGPT pass the Turkish Orthopedics and Traumatology Board Examination? Turkish orthopedic surgeons versus artificial intelligence

Çağdaş Pamuk,[1] Abdullah Faruk Uyanık,[2] Ersin Kuyucu,[1] Meriç Uğurlar[3]

[1]Department of Orthopedics and Traumatology, Private Silivri Anadolu Hospital, İstanbul-*Türkiye*
[2]Department of Orthopedics and Traumatology, Prof. Dr. Cemil Taşçıoğlu City Hospital, İstanbul-*Türkiye*
[3]Department of Orthopedics and Traumatology, Nişantaşı University, İstanbul-*Türkiye*

ABSTRACT

BACKGROUND: Artificial intelligence has been shown to achieve successful outcomes in various orthopedic qualification examinations worldwide. This study aims to assess the performance of ChatGPT in the written section of the Turkish Orthopedics and Traumatology Board Examination, compare its results with those of candidates who took the exam, and determine whether ChatGPT is sufficient to achieve a passing score.

METHODS: This retrospective observational study evaluated whether ChatGPT achieved a passing grade on 400 publicly available questions from the Turkish orthopedics qualification exam over the past four years. ChatGPT's performance was compared with the mean scores of the candidates who took the exam.

RESULTS: A total of 627 candidates participated in the four exams included in the study, of whom 292 (46.5%) passed. ChatGPT received higher scores than 619 (98.7%) of the candidates. In all exams conducted between 2020 and 2023, ChatGPT achieved significantly higher scores than the mean exam success rate (p=0.012, p=0.012, p=0.002, p=0.005, respectively). Of the 400 questions analyzed, 36 (9%) included figures.

CONCLUSION: This is the first study to evaluate the performance of ChatGPT in the Turkish orthopedics proficiency exam. Our findings indicate that ChatGPT demonstrated high success in the Turkish Orthopedics and Traumatology Board Examination (TOTBE) written exam, achieving higher scores than the vast majority of candidates taking the exam (98.7%). ChatGPT performed well in the first part of the proficiency exam, where only theoretical knowledge is assessed. However, the human factor, which synthesizes both theoretical and practical knowledge, remains essential in daily medical practice.

Keywords: Artificial intelligence; ChatGPT-4o; orthopedics; education; trauma surgery.

INTRODUCTION

One of the most significant technological developments in recent years is the integration of artificial intelligence (AI) into daily life and its ability to perform tasks with remarkable accuracy.[1] With each new version, AI's capacity to execute complex tasks has improved, enabling highly accurate responses through guided commands. In light of these advancements, the pursuit of processing information, making the right decisions automatically, and reaching accurate results has gained significant importance in the medical field, leading to numerous studies in this direction.[2,3] Particularly in medical education, rapid access to information has become a crucial tool for synthesizing knowledge and achieving results quickly.[4,5] When

Cite this article as: Pamuk Ç, Uyanık AF, Kuyucu E, Uğurlar M. Can ChatGPT pass the Turkish Orthopedics and Traumatology Board Examination? Turkish orthopedic surgeons versus artificial intelligence. Ulus Travma Acil Cerrahi Derg 2025;31:310-315.

Address for correspondence: Çağdaş Pamuk
Department of Orthopedics and Traumatology, Private Silivri Anadolu Hospital, İstanbul, Türkiye
E-mail: cagdaspamuk@hotmail.com

Ulus Travma Acil Cerrahi Derg 2025;31(3):310-315 DOI: 10.14744/tjtes.2025.07724
Submitted: 14.01.2025 Revised: 14.01.2025 Accepted: 01.02.2025 Published: 03.03.2025

studies on AI in the orthopedic field are examined, it is evident that research is concentrated in key areas such as surgical education, accurate diagnosis, and determining the correct treatment indication.[6-8] Numerous studies have investigated whether AI provides accurate answers to frequently asked questions in search engines on various orthopedic topics and whether acceptable responses can be obtained in the evaluation of shoulder and hip pathologies. Additionally, research has assessed whether AI can serve as a reliable tool in orthopedic decision-making.[9-11] Although AI's capabilities in orthopedic surgery are improving daily, the most critical challenge remains the evaluation of its reliability. In orthopedic practice, as in all medical fields, the risk of incorrect decision-making and the spread of misinformation that could harm patients is unacceptable. Therefore, it is essential to assess the accuracy of AI tools from a scientific perspective. This is crucial not only for the development of AI technology in a more reliable manner but also for its effective integration into daily practice.[4,12,13]

The most useful and practical method for evaluating AI's competence in orthopedics is to assess whether it can successfully pass board examinations that currently evaluate the proficiency of orthopedic residents and specialists. Previous studies have examined whether AI can pass board examinations in various medical specialties, revealing that it often achieves results comparable to those of human candidates who take the exam.[14,15] In the field of orthopedics, AI's success has been evaluated in various orthopedic examinations, particularly the American Board of Orthopedic Surgery Examination. To our knowledge, while the performance of AI in major orthopedic board exams worldwide has been studied in the literature, the Turkish Orthopedics and Traumatology Board Examination (TOTBE) has not yet been evaluated.[16,17] In this study, we hypothesized that AI's performance in TOTBE, a globally standardized exam, would make a significant contribution to the literature and could yield results similar to those of human candidates. Therefore, the aim of this study is to assess the performance of ChatGPT (OpenAI®, San Francisco, USA) in the written section of TOTBE, compare its performance with that of candidates who took the exam, and determine whether ChatGPT is capable of passing TOTBE.

MATERIALS AND METHODS

In this retrospective observational study, 400 questions from four publicly available TOTBE exams conducted over the past four years were included. All questions were multiple-choice. Questions containing figures, pictures, and tables were also specifically included in the study. Each year's exam questions were uploaded to ChatGPT one by one and requested to be answered. ChatGPT's examination statistics were then compared with the general examination statistics of that year. Additionally, the questions were classified into the following categories for subgroup analysis: Basic-General Orthopedics, Trauma, Pediatric Orthopedics and Spine Surgery, Hand, Wrist, and Upper Extremity Surgery, Foot and Ankle Surgery, Sports Surgery-Arthroscopy, Orthopedic Oncology, and Adult Reconstructive Surgery.

The questions were entered into the chat session of ChatGPT version 4o (OpenAI®, San Francisco, USA), one by one, in Turkish, to avoid confusion, and the answers were recorded.

The primary study outcome was to determine the percentage of questions ChatGPT answered correctly and whether it achieved a sufficient score to pass the TOTBE exam. The secondary outcome was to compare ChatGPT's performance with the average exam success of the candidates who took the exam. The tertiary analysis included a subgroup analysis of ChatGPT's responses based on the subject category (Basic-General Orthopedics, Trauma, Pediatric Orthopedics and Spine Surgery, Hand, Wrist, and Upper Extremity Surgery, Foot and Ankle Surgery, Sports Surgery-Arthroscopy, Orthopedic Oncology, and Adult Reconstructive Surgery).

Ethical approval was not required for this study, as it involved

Table 1. Distribution of Turkish Orthopedic and Traumatology Board Examination (TOTBE) questions by topic

Year	2020	2021	2022	2023
Basic-General Orthopedics	17	18	18	18
Trauma	31	31	31	31
Pediatric Orthopedics and Spine Surgery	18	18	18	18
Hand, Wrist, and Upper Extremity Surgery	11	11	11	11
Foot and Ankle Surgery	5	5	5	5
Sports Surgery, Arthroscopy	6	6	6	6
Orthopedic Oncology	5	5	5	5
Adult Reconstructive Surgery	7	6	6	6
Total	100	100	100	100

TOTBE: Turkish Orthopedics and Traumatology Board Examination.

Table 2. Overall performance of candidates and ChatGPT

Year	Lowest Passing Score	Highest Score	Mean Score of Candidates	ChatGPT-4 Score	P-value (Mean vs. ChatGPT)
2020	55	78	53.2	75	0.012
2021	60	88	60.3	78	0.012
2022	60	80	56.6	76	0.002
2023	58	78	56.9	76	0.005

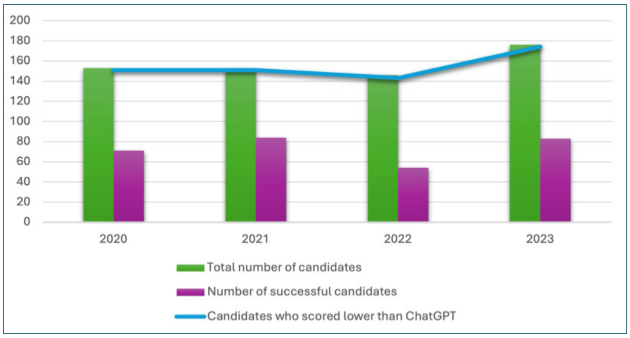

Figure 1. Distribution of candidates based on their exam success.

only the analysis of an online tool without human subject involvement.

Statistical Analysis

The number of questions to analyze was determined using a sample size calculator. With a margin of error of 5% and a response distribution rate of 50%, the minimum number of questions required to achieve a 95% confidence level was 78. Questions from TOTBE 2020 to 2023 were submitted to both candidates and the artificial intelligence model, totaling 100 questions. The responses from both groups were collected and their results compared. To analyze contingency tables, the Chi-squared exact test (SPSS Ver. 21 statistical package (SPSS Inc., Chicago, IL, USA)) was used. A p-value below 0.05 was considered significant.

RESULTS

Each of the TOTBE exams held between 2020 and 2023 consisted of 100 multiple-choice questions. The distribution of questions across subgroups remained approximately consistent each year (Table 1). A total of 400 questions were analyzed using ChatGPT and included in the study. The distribution of candidates' success rates by year, along with ChatGPT's performance in each respective year, is summarized in Table 2. In all exams from 2020 to 2023, ChatGPT demonstrated a significantly higher success rate than the mean exam success rate ($p=0.012$, $p=0.012$, $p=0.002$, $p=0.005$, respectively) (Table 2). A total of 627 candidates took the four exams included in the study, of whom 292 (46.5%) were successful. ChatGPT scored higher than 619 (98.7%) of the candidates (Fig. 1). The placement of ChatGPT within the distribution of all candidates' scores for the 2023 exam is shown in Figure 2. A subgroup analysis was conducted by comparing the 2023 exam statistics with ChatGPT's performance (Table 3). The analysis revealed that ChatGPT achieved significantly

Table 3. Comparison of ChatGPT performance by subgroups in the 2023 Turkish Orthopedic and Traumatology Board Examination (TOTBE)

Subgroup	Mean Score of Candidates (Net-Success Rate)	ChatGPT-4 Score (Net-Success Rate)	P-Value
Basic-General Orthopedics	9.9 (55%)	13 (72%)	0.018
Trauma	18.1 (58.3%)	24 (77.4%)	0.023
Pediatric Orthopedics and Spine Surgery	10.8 (60%)	14 (77.7%)	0.046
Hand, Wrist, and Upper Extremity Surgery	6.3 (57.2%)	9 (81.8%)	0.012
Foot and Ankle Surgery	3.1 (61.4%)	4 (80%)	0.176
Sports Surgery, Arthroscopy	4.4 (73%)	5 (83.3%)	0.085
Orthopedic Oncology	2.2 (54%)	3 (60%)	0.176
Adult Reconstructive Surgery	3.3 (54.3%)	4 (66.6%)	0.254
Total	56.9 (56.9%)	76 (76%)	0.005

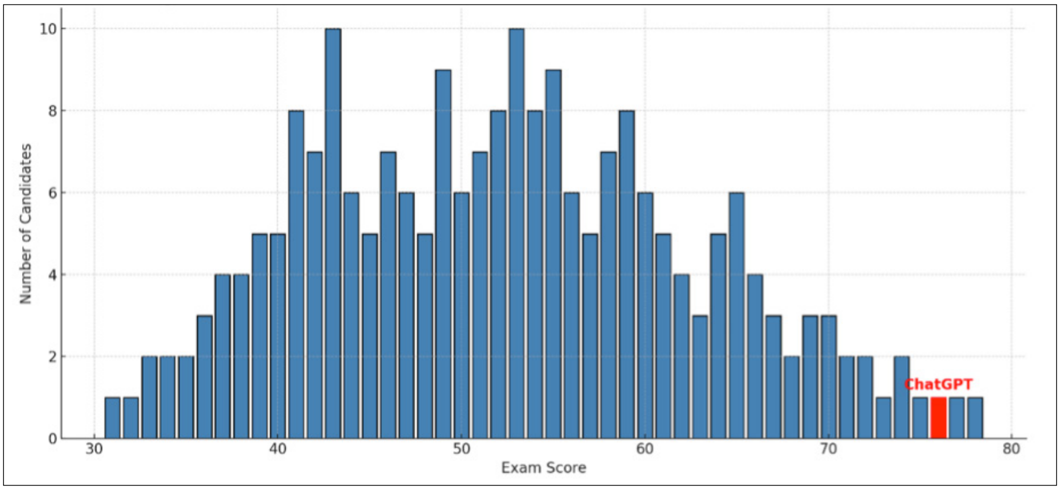

Figure 2. Distribution of candidates' exam scores in the 2023 Turkish Orthopedic and Traumatology Board Examination (TOTBE).

higher success rates in the fields of Basic-General Orthopedics, Trauma, Pediatric Orthopedics-Spine Surgery, and Hand, Wrist, and Upper Extremity Surgery (p=0.018, p=0.023, p=0.046, and p=0.012, respectively). Of the 400 questions included in the study, 36 (9%) contained figures. ChatGPT left 13 (36.1%) of these questions unanswered, answered 11 (30.5%) correctly, and 12 (33.3%) incorrectly.

DISCUSSION

The potential use of ChatGPT as a medical decision-making tool represents a groundbreaking technological innovation that has been the focus of extensive research in recent years.[4] The most reliable method for assessing ChatGPT's proficiency in accessing information across different medical specialties is to evaluate its performance in proficiency exams conducted worldwide. In recent years, ChatGPT's performance has been tested in various orthopedic qualification exams globally.[6,14,16-20] This study is the first to assess ChatGPT's performance in the Turkish Orthopedic and Traumatology Board Examination conducted in Türkiye. Our findings indicate that ChatGPT demonstrated high success in the written exam, achieving higher scores than the vast majority of candidates taking the exam (98.7%). Unlike previous studies, this research included questions containing figures, revealing that ChatGPT correctly answered 30.5% of these questions.[6,14,17] Additionally, the fact that the exam was conducted in Turkish distinguishes this study from prior research. Previous studies have shown that ChatGPT's scores in proficiency exams conducted in different countries around the world were close to or below the mean scores of human candidates. These evaluations were conducted in English, and the exam questions were not publicly available.[16,19,20] We believe that higher success rate observed in our study, conducted in Turkish, compared to other exams may be attributed to the availability of publicly accessible exam questions. Additionally, an analysis of the references cited by ChatGPT in answering the exam questions revealed that the Turkish Orthopedics and Traumatology Association Journal, which has been publishing review articles for many years, was frequently referenced. This suggests that Turkish knowledge and resources in the field of orthopedics and traumatology are highly comprehensive. It is well known that ChatGPT's algorithm has the capability to analyze extensive texts and utilize available resources through this process. In this regard, it has been observed that as different versions are updated and the amount of publicly available information increases, the accuracy of information retrieval also improves.[14,17]

A study comparing different versions of ChatGPT and evaluating performance on the French orthopedic proficiency exam found that the ChatGPT-4o version used in our study demonstrated higher success than its predecessors. As newer versions are introduced, their ability to analyze and synthesize information continues to improve.[14] We believe that the use of the latest version of ChatGPT in our study may have contributed to its higher success compared to previous research.

When examining the algorithms of AI models, it has been observed that the presence of more keywords enhances their analytical capabilities. Thus, the data format plays a crucial role in obtaining accurate answers. While performance declines in exams where a main question text is provided and a direct answer is required, success rates increase in multiple-choice exams.[2,21,22] As a data-processing tool, ChatGPT is still far from achieving high success in orthopedic qualification exams in a way that could replace human expertise, particularly in terms of excluding the human factor. This remains a significant deficiency of AI models. Our study found that 36.1% of the questions containing figures were left unanswered, indicating that the AI was unable to interpret visual data effectively. Research is ongoing to improve the ability of newer versions to analyze figures and images.

Having accurate and sufficient information is the most important game-changer in today's modern world. In previous studies, ChatGPT's responses to the most frequently searched questions on various orthopedic diseases were compared with the expertise of specialists in the field. Although ChatGPT provided acceptable answers, it was evident that expert opinion is still crucial when making medical decisions.[9,23] While our study observed a high success rate, ChatGPT was successful only in the first part of the proficiency exam, where theoretical knowledge is assessed. However, in daily medical practice, the human factor, combining theoretical and practical knowledge, remains essential.

In past studies, when the references and explanations provided by ChatGPT for incorrect answers were analyzed, it was found that the model could generate meaningful justifications, allowing the correct answer to be inferred. This raised questions about the quality of the exam questions themselves.[18,24] However, in our study, no such issue was observed, and the TOTBE questions in Turkish were deemed adequate in terms of clarity and design. From this perspective, we believe that ChatGPT can serve as a valuable tool for exam committees in evaluating question quality and statement accuracy.

While AI models significant advantages in scientific knowledge production, they also introduce ethical concerns. Due to their ability to analyze long texts, these models facilitate statistical analysis, graphic design, and complex calculations with ease. However, when it comes to generating new and original knowledge, ethical concerns arise, as these models cannot go beyond the repetition or rephrasing of existing knowledge. Current versions still fall short of advancing beyond synthesis and simplified evaluation of pre-existing knowledge.[4,25,26]

There are several limitations to our study. The high success rates observed may be attributed to the retrospective nature of the study and the fact that TOTBE questions are publicly available on search engines. However, when viewed from a different perspective on accessing accurate information, it has been demonstrated that ChatGPT can also play a significant role in education and training. Another limitation is that the exam performance of different ChatGPT versions was not compared. However, the objective of this study was to evaluate the proficiency of artificial intelligence in the exam, assessing the performance of the latest version was essential.

CONCLUSION

In conclusion, ChatGPT performed significantly above the mean score of the candidates who took the TOTBE written exam, demonstrating a high ability to access orthopedic information. An analysis of the references provided by ChatGPT for correctly answered questions revealed that Turkish orthopedic information sources are abundant and easily accessible. Therefore, we believe that ChatGPT can serve as an important tool both for accessing accurate information in orthopedic education and for the preliminary evaluation of question quality in exams prepared by educational committees. Further studies will be necessary to assess the performance of future versions of artificial intelligence in all components of the TOTBE.

Acknowledgments: The authors would like to thank the Turkish Orthopedics and Traumatology Association and the Turkish Orthopedics and Traumatology Education Council for their efforts in publicly publishing qualification exams and facilitating access to information.

Ethics Committee Approval: Ethical approval was not required for this study, as it involved only the analysis of an online tool without human subject involvement.

Peer-review: Externally peer-reviewed.

Authorship Contributions: Concept: Ç.P., A.F.U., E.K., M.U.; Design: Ç.P., A.F.U., E.K., M.U.; Supervision: Ç.P., E.K.; Resource: Ç.P., A.F.U.; Materials: Ç.P., A.F.U.; Data Collection and/or Processing: Ç.P., E.K.; Analysis and/or Interpretation: Ç.P., M.U.; Literature Review: Ç.P., M.U.; Writing: Ç.P., E.K.; Critical Review: Ç.P., A.F.U., E.K., M.U.

Conflict of Interest: The authors have no conflicts of interest to declare.

Financial Disclosure: The author declared that this study has received no financial support.

REFERENCES

1. Reis F, Lenz C, Gossen M, Volk HD, Drzeniek NM. Practical Applications of Large Language Models for Health Care Professionals and Scientists. JMIR Med Inform 2024;12:e58478. [CrossRef]
2. Wilson T, Lee KKW. How I Do It: Using ChatGPT to Create Clinical Scenarios for Surgical Learning and Oral Boards Preparation. J Surg Educ 2024;81:1623-31. [CrossRef]
3. Atik OŞ. Artificial intelligence: Who must have autonomy the machine or the human? Jt Dis Relat Surg 2024;35:1-2. [CrossRef]
4. Odri GA, Ji Yun Yoon D. Detecting generative artificial intelligence in scientific articles: Evasion techniques and implications for scientific integrity. Orthop Traumatol Surg Res 2023;109:103706. [CrossRef]
5. Atik OŞ. Artificial intelligence, machine learning, and deep learning in orthopedic surgery. Jt Dis Relat Surg 2022;33:484-5. [CrossRef]
6. Kung JE, Marshall C, Gauthier C, Gonzalez TA, Jackson JB 3rd. Evaluating ChatGPT Performance on the Orthopaedic In-Training Examination. JB JS Open Access 2023;8:e23.00056. [CrossRef]
7. Das D, Kumar N, Longjam LA, Sinha R, Deb Roy A, Mondal H, et al. Assessing the Capability of ChatGPT in Answering First- and Second-Order Knowledge Questions on Microbiology as per Competency-Based Medical Education Curriculum. Cureus 2023;15:e36034. [CrossRef]
8. Loftus TJ, Tighe PJ, Filiberto AC, Efron PA, Brakenridge SC, Mohr AM, Rashidi P, Upchurch GR Jr, Bihorac A. Artificial Intelligence and Surgical Decision-making. JAMA Surg 2020;155:148-58. [CrossRef]
9. Ayık G, Ercan N, Demirtaş Y, Yıldırım T, Çakmak G. Evaluation of ChatGPT-4o's answers to questions about hip arthroscopy from the patient perspective. Jt Dis Relat Surg 2025;36:193-9. [CrossRef]
10. Wu L, Yang XC, Wu J, Zhao X, Lu ZD, Li P. Short-term outcome of artificial intelligence-assisted preoperative three-dimensional planning of total hip arthroplasty for developmental dysplasia of the hip compared to

traditional surgery. Jt Dis Relat Surg 2023;34:571-82. [CrossRef]
11. Kumar V, Roche C, Overman S, Simovitch R, Flurin PH, Wright T, et al. Using machine learning to predict clinical outcomes after shoulder arthroplasty with a minimal feature set. J Shoulder Elbow Surg 2021;30:e225-36. [CrossRef]
12. Charles YP, Lamas V, Ntilikina Y. Artificial intelligence and treatment algorithms in spine surgery. Orthop Traumatol Surg Res 2023;109:103456. [CrossRef]
13. Guerrero DT, Asaad M, Rajesh A, Hassan A, Butler CE. Advancing Surgical Education: The Use of Artificial Intelligence in Surgical Training. Am Surg 2023;89:49-54. [CrossRef]
14. Maraqa N, Samargandi R, Poichotte A, Berhouet J, Benhenneda R. Comparing performances of french orthopaedic surgery residents with the artificial intelligence ChatGPT-4/4o in the French diploma exams of orthopaedic and trauma surgery. Orthop Traumatol Sur [CrossRef]
15. Howard FM, Li A, Riffon MF, Garrett-Mayer E, Pearson AT. Characterizing the Increase in Artificial Intelligence Content Detection in Oncology Scientific Abstracts From 2021 to 2023. JCO Clin Cancer Inform 2024;8:e2400077. [CrossRef]
16. Lum ZC. Can Artificial Intelligence Pass the American Board of Orthopaedic Surgery Examination? Orthopaedic Residents Versus ChatGPT. Clin Orthop Relat Res 2023;481:1623-30. [CrossRef]
17. Massey PA, Montgomery C, Zhang AS. Comparison of ChatGPT-3.5, ChatGPT-4, and Orthopaedic Resident Performance on Orthopaedic Assessment Examinations. J Am Acad Orthop Surg 2023;31:1173-9. [CrossRef]
18. Kasapovic A, Ali T, Babasiz M, Bojko J, Gathen M, Kaczmarczyk R, et al. Does the Information Quality of ChatGPT Meet the Requirements of Orthopedics and Trauma Surgery? Cureus 2024;16:e60318. [CrossRef]
19. Saad A, Iyengar KP, Kurisunkal V, Botchu R. Assessing ChatGPT's ability to pass the FRCS orthopaedic part A exam: A critical analysis. Surgeon 2023;21:263-6. [CrossRef]
20. Maroteau G, An JS, Murgier J, Hulet C, Ollivier M, Ferreira A. Evaluation of the impact of large language learning models on articles submitted to Orthopaedics & Traumatology: Surgery & Research (OTSR): A significant increase in the use of artificial intelligence in 2023. Orthop Traumatol Surg Res 2023;109:103720. [CrossRef]
21. Ali R, Tang OY, Connolly ID, Fridley JS, Shin JH, Zadnik Sullivan PL, et al. Performance of ChatGPT, GPT-4, and Google Bard on a Neurosurgery Oral Boards Preparation Question Bank. Neurosurgery 2023;93:1090-8. [CrossRef]
22. Blacker SN, Chen F, Winecoff D, Antonio BL, Arora H, Hierlmeier BJ, et al. An Exploratory Analysis of ChatGPT Compared to Human Performance With the Anesthesiology Oral Board Examination: Initial Insights and Implications. Anesth Analg 2024 Sep 13. [Epub ahead of print]. [CrossRef]
23. Yapar D, Demir Avcı Y, Tokur Sonuvar E, Eğerci ÖF, Yapar A. ChatGPT's potential to support home care for patients in the early period after orthopedic interventions and enhance public health. Jt Dis Relat Surg 2024;35:169-76. [CrossRef]
24. Keating M, Bollard SM, Potter S. Assessing the Quality, Readability, and Acceptability of AI-Generated Information in Plastic and Aesthetic Surgery. Cureus 2024;16:e73874. [CrossRef]
25. Gralha JG, Pimentel AS. Gotcha GPT: Ensuring the Integrity in Academic Writing. J Chem Inf Model 2024;64:8091-7. [CrossRef]
26. Bisi T, Risser A, Clavert P, Migaud H, Dartus J. What is the rate of text generated by artificial intelligence over a year of publication in Orthopedics & Traumatology: Surgery & Research? Analysis of 425 articles before versus after the launch of ChatGPT in November 2022. Orthop Traumatol Surg Res 2023;109:103694. [CrossRef]

ORİJİNAL ÇALIŞMA - *ÖZ*

ChatGPT Türk ortopedi ve travmatoloji yeterlilik sınavını geçebilir mi? Türk Ortopedi Cerrahları yapay zeka'ya karşı

AMAÇ: Yapay zekanın dünyada birçok ortopedi yeterlilik sınavında başarılı sonuçlar elde edebildiği bilinmektedir. Bu nedenle bu çalışma, ChatGPT'nin Türk Ortopedi ve Travmatoloji Yeterlilik Sınavı yazılı bölümündeki performansını değerlendirmeyi, bu performansı sınava giren adayların sonuçlarıyla karşılaştırmayı ve ChaGPT'nin sınavı geçmek için yeterli olup olmadığını araştırmayı amaçlamaktadır.
GEREÇ VE YÖNTEM: Bu retrospektif gözlemsel çalışmada, ChatGPT'nin son dört yılda Türkiye'de yapılan ve halka açık olarak yayınlanan ortopedi yeterlilik sınavında sorulan 400 soruda geçer not alıp almadığı değerlendirilmiştir. ChatGPT'nin performansı, sınava giren adayların ortalama puanlarıyla karşılaştırılmıştır.
BULGULAR: Çalışmaya dahil edilen dört sınava toplam n=627 aday katılmış ve bunların n=292'si (%46.5) başarılı olmuştur. ChatGPT'nin adayların n=619'undan (%98.7) daha yüksek puan aldığı görülmüştür. 2020-2023 yılları arasındaki tüm sınavlarda ChatGPT, ortalama sınav başarısının önemli ölçüde üzerinde bir başarı gösterdi (sırasıyla p= 0.012, p= 0.012, p=0.002, p=0.005). Çalışmaya dahil edilen 400 sorunun 36'sının (%9) şekil içeren sorulardan oluştuğu görüldü.
SONUÇ: Bu, Türkiye'de yapılan ortopedi yeterlilik sınavında ChatGPT performansını değerlendirmek için yapılan ilk çalışmadır. Çalışmamızda, ChatGPT'nin Türk ortopedi ve travmatoloji yeterlilik yazılı sınavında yüksek başarı gösterdiği ve sınava giren adayların büyük çoğunluğundan (%98.7) daha yüksek puanlar aldığı görüldü. ChatGPT, yalnızca teorik bilginin ölçüldüğü yeterlilik sınavının ilk bölümünde başarılı oldu. Ancak teorik ve pratik bilginin sentezi olan insan faktörü, günlük tıbbi uygulamada hala çok önemli bir yer tutmaktadır.

Anahtar sözcükler: Yapay zeka; ChatGPT-4o; ortopedi; eğitim; travma cerrahisi.

Ulus Travma Acil Cerrahi Derg 2025;31(3):310-315 DOI: 10.14744/tjtes.2025.07724