

Original Paper

# Assessment of ChatGPT-3.5's Knowledge in Oncology: Comparative Study with ASCO-SEP Benchmarks

Roupen Odabashian<sup>1\*</sup>, MD; Donald Bastin<sup>2\*</sup>, MD; Georden Jones<sup>3\*</sup>, PhD; Maria Manzoor, MD; Sina Tangestaniapour, MD; Malke Assad<sup>4</sup>, MD; Sunita Lakhani<sup>5</sup>, MD; Maritsa Odabashian<sup>3,6</sup>, Bsc (c); Sharon McGee<sup>7,8\*</sup>, MD, PhD

<sup>1</sup>Department of Oncology, Barbara Ann Karmanos Cancer Institute, Wayne State University, Detroit, MI, United States

<sup>2</sup>Department of Medicine, Division of Internal Medicine, The Ottawa Hospital and the University of Ottawa, Ottawa, ON, Canada

<sup>3</sup>Mary A Rackham Institute, University of Michigan, Ann Arbor, MI, United States

<sup>4</sup>Department of Plastic Surgery, University of Pittsburgh Medical Center, Pittsburgh, PA, United States

<sup>5</sup>Department of Medicine, Division of Internal Medicine, Jefferson Abington Hospital, Philadelphia, PA, United States

<sup>6</sup>The Ottawa Hospital Research Institute, Ottawa, ON, Canada

<sup>7</sup>Department of Medicine, Division of Medical Oncology, The Ottawa Hospital and the University of Ottawa, Ottawa, ON, Canada

<sup>8</sup>Cancer Therapeutics Program, Ottawa Hospital Research Institute, Ottawa, ON, Canada

\*these authors contributed equally

**Corresponding Author:**

Roupen Odabashian, MD

Department of Oncology, Barbara Ann Karmanos Cancer Institute, Wayne State University

4100 John R St

Detroit, MI, 48201

United States

Phone: 1 (313) 745 3000 ext 7731

Fax: 1 313 576 8120

Email: [roupen.odabashian@mclaren.org](mailto:roupen.odabashian@mclaren.org)

## Abstract

**Background:** ChatGPT (Open AI) is a state-of-the-art large language model that uses artificial intelligence (AI) to address questions across diverse topics. The American Society of Clinical Oncology Self-Evaluation Program (ASCO-SEP) created a comprehensive educational program to help physicians keep up to date with the many rapid advances in the field. The question bank consists of multiple choice questions addressing the many facets of cancer care, including diagnosis, treatment, and supportive care. As ChatGPT applications rapidly expand, it becomes vital to ascertain if the knowledge of ChatGPT-3.5 matches the established standards that oncologists are recommended to follow.

**Objective:** This study aims to evaluate whether ChatGPT-3.5's knowledge aligns with the established benchmarks that oncologists are expected to adhere to. This will furnish us with a deeper understanding of the potential applications of this tool as a support for clinical decision-making.

**Methods:** We conducted a systematic assessment of the performance of ChatGPT-3.5 on the ASCO-SEP, the leading educational and assessment tool for medical oncologists in training and practice. Over 1000 multiple choice questions covering the spectrum of cancer care were extracted. Questions were categorized by cancer type or discipline, with subcategorization as treatment, diagnosis, or other. Answers were scored as correct if ChatGPT-3.5 selected the answer as defined by ASCO-SEP.

**Results:** Overall, ChatGPT-3.5 achieved a score of 56.1% (583/1040) for the correct answers provided. The program demonstrated varying levels of accuracy across cancer types or disciplines. The highest accuracy was observed in questions related to developmental therapeutics (8/10; 80% correct), while the lowest accuracy was observed in questions related to gastrointestinal cancer (102/209; 48.8% correct). There was no significant difference in the program's performance across the predefined subcategories of diagnosis, treatment, and other ( $P=.16$ , which is greater than .05).

**Conclusions:** This study evaluated ChatGPT-3.5's oncology knowledge using the ASCO-SEP, aiming to address uncertainties regarding AI tools like ChatGPT in clinical decision-making. Our findings suggest that while ChatGPT-3.5 offers a hopeful outlook for AI in oncology, its present performance in ASCO-SEP tests necessitates further refinement to reach the requisite competency levels. Future assessments could explore ChatGPT's clinical decision support capabilities with real-world clinical

scenarios, its ease of integration into medical workflows, and its potential to foster interdisciplinary collaboration and patient engagement in health care settings.

(*JMIR AI* 2024;3:e50442) doi: [10.2196/50442](https://doi.org/10.2196/50442)

## KEYWORDS

artificial intelligence; ChatGPT-3.5; language model; medical oncology

## Introduction

OpenAI released ChatGPT, a pioneering artificial intelligence (AI) language model, in late 2022. ChatGPT-3 is an AI chatbot that can comprehend user input and react to it in a manner that is natural and human-like [1]. The program was trained on a large body of data sourced from the internet, including textbooks, articles, social media posts, and web-based forums, up to the last quarter of 2021 [2]. It works by analyzing user input text to generate a response using a probabilistic distribution of words and phrases derived from its training data. To date, it has significantly impacted numerous disciplines, including law, health care, and medical education [3-6]. Large language models like ChatGPT-3.5 represent a significant advancement in the preceding class of deep learning-based models, by facilitating the interpretation, processing, and production of natural language [7].

The use of AI has rapidly emerged as a promising approach in the health care industry, where it has been applied to medical imaging analysis, drug discovery, and patient monitoring [8]. Recent research has evaluated ChatGPT-3.5's abilities to respond to standardized questions from professional examinations for law and the United States Medical Licensing Examination (USMLE) [3,4]. ChatGPT-3.5 was able to achieve passing grades on these examinations while providing logical and informative explanations. Additionally, studies have been conducted to assess ChatGPT's capabilities in responding to international medical licensing examinations from countries such as Italy, France, Spain, the United Kingdom, and India. The success rates observed ranged between 22% and 73% [9].

AI and Chat GPT showcase substantial promise in augmenting medical consultations, offering preliminary diagnostic suggestions, and providing a vast knowledge base for medical practitioners and patients alike [10]. However, while it embarks on a path toward a more integrated health care AI system, several limitations hinder its full potential. The model's reliance on historical data without the ability to access real-time patient data can lead to outdated or inaccurate information dissemination. Additionally, its inability to comprehend nuanced human emotions and the ethical implications surrounding patient data privacy remain significant hurdles [11].

AI has displayed a notable deficiency in grasping context and nuance, elements that are fundamental for delivering safe and effective patient care [12]. Furthermore, analyzing the prospects of job automation in health care, Frey and Osborne [13] have projected that while administrative roles within the sector, such as health information technicians, exhibit a high likelihood of automation at 91%, the odds plummet to a mere 0.42% for the

automation of roles held by physicians and surgeons. This stark contrast underscores the intricate nature of medical practice, which extends beyond the mere application of codified knowledge. Additionally, there is a burgeoning discussion around the ethical dimensions of using conversational AI in medical practice. The crux of the issue revolves around the substantial volume of high-quality data required to train these models. Present-day algorithms are often honed on biased data sets, inheriting not just the availability, selection, and confirmation biases inherent in the data but also displaying a propensity to exacerbate these biases [14]. Looking ahead, the evolving capabilities of AI hint at the potential for tackling more sophisticated tasks, such as orchestrating experiments or future clinical trials [15] or engaging in peer review processes [16].

The American Society of Clinical Oncology Self-Evaluation (ASCO-SEP) program created a comprehensive educational program to help physicians keep up to date with the many rapid advances in the field. The question bank consists of multiple choice questions (MCQs) addressing the many facets of cancer care, including diagnosis, treatment, and supportive care. It is intended to evaluate participants' knowledge and give them feedback to direct future learning. The program is largely regarded as the leading resource for cancer specialists seeking to gain and maintain professional licensure in the field of medical oncology [17].

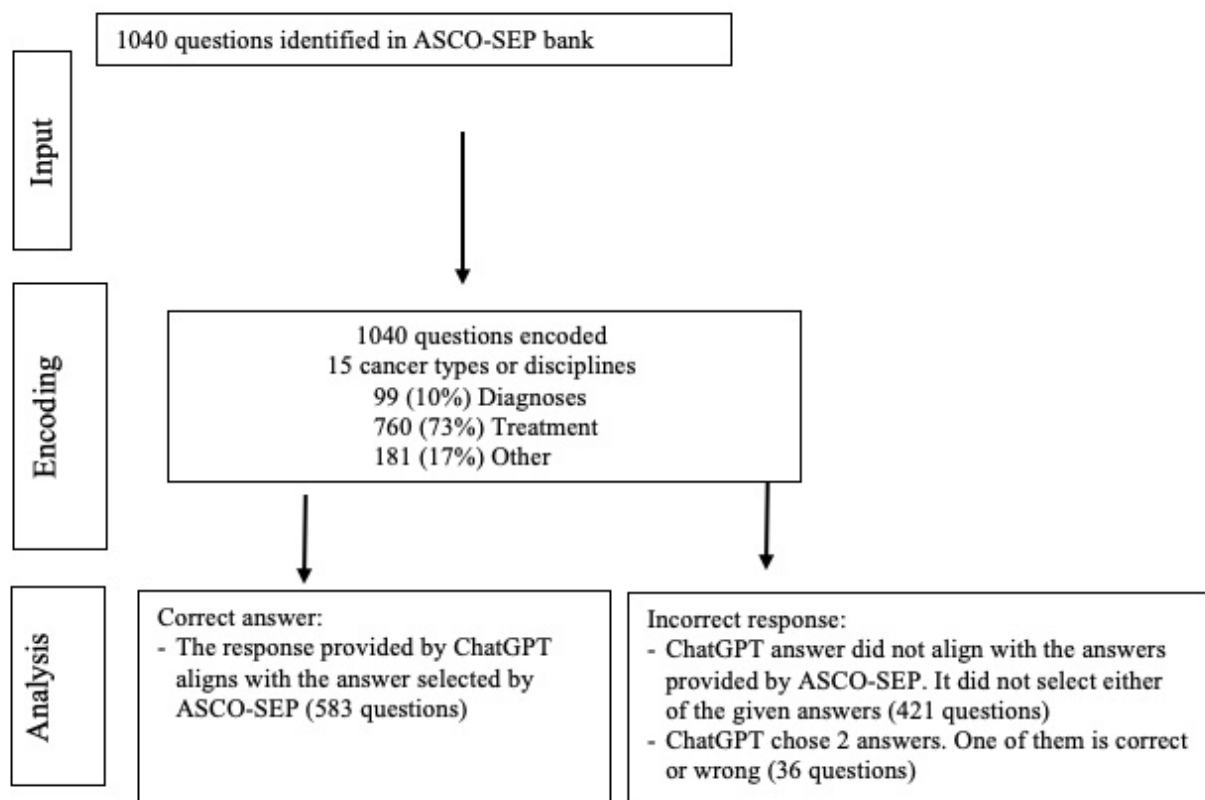
However, the evolving complexities of oncological care necessitate additional tools that can aid oncologists in clinical decision-making. By assessing ChatGPT-3's ability to answer ASCO-SEP questions, this study's objective is to understand ChatGPT's potential to serve as a supportive instrument in clinical decisions, offering instantaneous insights for health care providers, and to identify novel and efficient educational aids in oncology, with a specific emphasis on their role in clinical decisions.

## Methods

### Input Data

Questions were sourced from ASCO-SEP, which consists of approximately 1000 MCQs covering the spectrum of cancer care. The question bank was accessed from February 2023 to March 2023. As ChatGPT-3.5 can only generate responses to textual data, the study excluded questions with images, tables, or other non-textual content. Questions consisted of an information stem followed by a specific question with 3-5 possible answers (A-E), along with their corresponding letter choices, only 1 of which was correct. [Figure 1](#) illustrates the workflow for data sourcing, input, encoding, and analysis.

**Figure 1.** Schematic of sourcing, encoding, and scoring procedures. ASCO-SEP: American Society of Clinical Oncology Self-Evaluation Program.



Before proceeding with the analysis, a random spot check was performed. For this, a random subset of the ASCO-SEP questions was selected, and their answers, explanations, or related content were manually cross-referenced with Google's index to ensure that they were not present before January 1, 2022, the last date accessible to the ChatGPT training data.

During this study, we used the free version of ChatGPT-3.5. At that time, ChatGPT-4 and its associated plugins were not yet available.

### Encoding

We imported individual ASCO-SEP questions, including the information stem and multiple-choice response options, into the ChatGPT-3.5 interface. The questions were formatted to include the question stem, followed by each potential response on a separate line. We did not change the structure of the questions given to ChatGPT-3.5 and entered them in the original format provided by ASCO-SEP without altering the phrasing or the wording. A new conversation session was started in ChatGPT for each question. We did not provide ChatGPT-3.5 with any prompts and offered only one opportunity to answer each question.

### Data Analysis

Selected questions were grouped by cancer type or discipline (eg, breast, lung, and colon cancer) with further

subcategorization based on content such as treatment, diagnosis, or other. ChatGPT was deemed to have responded correctly if it chose the correct answer as defined and provided by ASCO-SEP. The study team did not define or determine the correct answer. The program was not asked to provide justifications or references for answers. No point was assigned if ChatGPT-3.5 provided an answer that was not from the options given. Questions where ChatGPT-3.5 chose 2 possible answers or chose multiple answers and did not commit to a single best answer were also considered wrong, even if 1 of the responses was correct.

For statistical analysis, data were logged, scored, and analyzed in Excel (Microsoft Corp). Specifically, a chi-square test was performed to determine if there was a significant difference in the distribution of correct answers across different categories or groups.

## Results

A total of 1040 questions were extracted from the ASCO-SEP question bank.

The questions covered 15 cancer types or disciplines. The largest portion focused on breast (223/1040; 21.4%) and gastrointestinal (209/1040; 20%) cancers, with  $\leq 1\%$  (13/1040) covering central nervous system malignancies, developmental therapeutics, and prevention/epidemiology (Table 1).

**Table 1.** Question distribution and proportions by cancer type or specialty area.

Cancer type or discipline	Number of questions (N=1040), n (%)
Breast cancer	223 (21.4)
Gastrointestinal cancer	209 (20)
Thoracic oncology	137 (13.1)
Hematological malignancies	121 (11.6)
Genitourinary cancer	97 (9)
Melanoma and skin cancer	43 (4)
Sarcoma	36 (3)
Head and neck	36 (3)
Gynecologic cancers	36 (3)
General oncology	29 (3)
Supportive and palliative care	28 (3)
Genetics and genomics	17 (2)
Central nervous system	13 (1)
Developmental therapeutics	10 (1)
Prevention and epidemiology	5 (0.5)

Varying levels of accuracy were observed in ChatGPT-3.5's performance in answering questions based on different cancer types or disciplines (Table 2). The highest accuracy was achieved in questions related to developmental therapeutics (8/10; 80% correct), while the lowest accuracy was observed for questions related to gastrointestinal cancer (102/209; 48.8% correct).

**Table 2.** Accuracy rates by cancer type or specialty area.

Cancer type or discipline	Discipline-specific accuracy rates, n/N (%)
Developmental therapeutics	8/10 (80)
Central nervous system	10/13 (77)
Melanoma and skin cancer	28/43 (65)
Genetics and genomics	11/17 (65)
General oncology	18/29 (62)
Gynecologic cancers	22/36 (61)
Supportive and palliative care	17/28 (61)
Prevention and epidemiology	3/5 (60)
Head and neck	21/36 (58)
Breast cancer	130/223 (58.3)
Sarcoma	20/36 (57)
Thoracic oncology	77/137 (56)
Hematological malignancies	66/121 (55)
Genitourinary cancer	49/97 (51)
Gastrointestinal cancer	102/209 (48.8)
Total	583/1040 (56.1)

Questions were further subcategorized as "diagnosis," "treatment," and "other," with the latter covering topics such as biostatistics, cancer staging, and treatment complications. Out of the total questions, 73.1% (760/1040) were related to cancer treatment, 10% (99/1040) focused on diagnosis, and the remaining 17.4% (181/1040) were categorized as "other" (Table

3). Accuracy based on subcategory also varied, with 55% (418/760) of treatment questions, 63% (62/99) of diagnosis questions, and 56.9% (103/181) of "other" questions answered correctly (Table 2). There was no significant difference in the program's performance across the predefined subcategories of

diagnosis, treatment, and other ( $P=.16$ , which is greater than  $.05$ ).

**Table 3.** ChatGPT-3.5 performance on questions per subcategory.

Category	Number of questions, n (%)	Overall accuracy, n/N (%)	<i>P</i> value <sup>a</sup>
Treatment	760 (73.1)	418/760 (55)	.16
Diagnosis	99 (10)	62/99 (63)	.16
Other	181 (17.4)	103/181 (56.9)	.16
Overall	1040 (100)	583/1040 (56.3)	.16

<sup>a</sup>Chi-square test.

Overall, ChatGPT-3.5 achieved a score of 56.3% (583/1040) for correct answers provided across all categories. Of note, responses were marked as incorrect if ChatGPT-3.5 provided 2 or more answers, even if 1 of those answers was correct (37/1040, 3%; [Figure 1](#)).

## Discussion

### Overview

In this study, we evaluated the performance of ChatGPT-3.5 in answering ASCO-SEP questions designed for medical oncologists in training and practice to support licensure and ongoing medical education. To facilitate a fair and rigorous assessment, spot checks were performed to ensure answers were not present in the program training data, and questions were entered in separate sessions to avoid grounding bias. Furthermore, questions were presented in their original format, as seen by physicians, with no changes made to prompt the program.

Over 1000 questions were posed to the program, spanning the spectrum of cancer care, with an overall score of 56.3% (583/1040) achieved. While promising, this is, however, below the accepted threshold of 70% that is required by ASCO-SEP to claim CME credits using their question bank [18].

Since the launch of ChatGPT-3.5, several studies have evaluated the program's performance on medical examinations. A notable study conducted by Kung et al [3] assessed ChatGPT-3.5's performance on the USMLE taken by US medical students. The results showed that ChatGPT-3.5 performed at, or near, the passing threshold for all 3 examinations. Specifically, the accuracy rates for USMLE Steps 1, 2 CK, and 3 were 68.0%, 58.3%, and 62.4%, respectively, which are acceptable passing scores. Gilson et al [19] reported similar results, where ChatGPT-3.5 scored 60% on USMLE test questions. It is worth noting that although the authors used questions published on the USMLE website after the training date cutoff for ChatGPT, which is late 2021, many of these questions were similar to those published in previous years. Moreover, these questions were discussed on web-based forums, which may explain the higher scores achieved [20]. Additionally, previous studies have evaluated ChatGPT-3.5's performance in microbiology [21] and pathology [22] and have shown promising outcomes in these fields with an accuracy rate of 80%.

Several factors might explain why ChatGPT's performs differently on USMLE compared to ASCO-SEP questions. First,

the ASCO-SEP is tailored for medical oncologists, delving deep into cancer care, while USMLE caters to a broader set of medical students, covering general medical knowledge. Given that ChatGPT-3.5's training data spans a wide range of topics, it's plausible that the content aligns more with the generalized medical queries of USMLE than the specialized focus of ASCO-SEP. Additionally, the structure and phrasing of questions play a critical role, potentially influencing AI's response accuracy. The questions within the USMLE typically features keywords that assist students in selecting an answer from the provided options. Conversely, the ASCO-SEP presents more specialized questions, challenging physicians' ability to discern first- and second-line treatments for a specified condition [23]. For instance, in 1 of the numerous subreddits [24] available web-based that was likely included in ChatGPT's training data set [25] students discuss how certain keywords aid them in answering examination questions. These data might have assisted ChatGPT in responding to USMLE questions in a previous paper that tested ChatGPT's performance on the USMLE [3,19]. However, such keywords are not used or discussed among physicians engaging with ASCO-SEP questions.

There are additional possible explanations for the observed performance of ChatGPT-3.5 in this study. One key factor is the comprehensive data set of over 1000 questions used, which allowed for a more thorough and holistic evaluation of the program's performance compared to previous studies [3,19,26,27]. Another contributing factor may be the dynamic and rapid scientific and clinical advances that occur in the field of oncology, which ChatGPT-3.5 could not fully tackle given that its training data is limited to pre-2022 internet data, with restricted access to key databases in the field like PubMed [28].

ChatGPT-3.5 demonstrated varying levels of accuracy in answering questions across the different cancer types and disciplines. Questions related to developmental therapeutics had the highest accuracy rate (80%, 8/10); however, the limited question sample size may not have allowed a complete assessment. Indeed, ChatGPT-3.5's lowest score was achieved in gastrointestinal cancer, which contained one of the largest numbers of questions in the bank (102/209, 48.8%), suggesting that broader assessments may identify more knowledge gaps. This study did not, however, find any significant difference in ChatGPT-3.5's performance across the subcategories of diagnosis, treatment, and others.

While ChatGPT-3.5 is not yet fully dependable for complex decision-making in medical oncology, it shows promise in the field. In recent years, we have witnessed significant progress in neural networks, and the future of health care is becoming increasingly multimodal. Oncologists now rely on more than just text-based information when prescribing treatments. They consider a wide range of factors, including diverse image types, genomic data, and social determinants of health. However, in the past, developing multimodal machine learning models seemed like an overly ambitious goal. Thankfully, the landscape has changed, and we have seen exciting advancements in this area through various publications in 2022 and 2023 [29,30]. These studies have showcased the potential applications of multimodal models in the field of oncology, bringing us closer to a more comprehensive and holistic approach to cancer care.

Based on its performance in this study, we do not think that AI can aid oncologists in clinical decision-making at this time. However, it may excel in other tasks in the field [31]. Experts might look to language-generating AI to reduce the burden on humans who create questions and explanations for tests. However, it should be noted that ChatGPT-3.5 is not a useful tool without human supervision at this point, given its potential to fabricate references that may sound plausible but are incorrect [14,32,33]. Oncologists can also use it for administrative tasks such as drafting notes [34] or crafting communication messages for patients [11]. Additionally, while a previous study by Johnson et al [35] demonstrated that ChatGPT can be used by patients to answer common cancer myths and questions, the questions used in this study were already featured on the National Cancer Institute's webpage and were likely part of ChatGPT's training data [25] and fewer questions were used. We can infer from this study that the answers provided by ChatGPT still require review by an oncologist to ascertain their accuracy.

In the future, AI has the potential to assist oncologists in critical aspects such as determining optimal chemotherapy dosages [36] and aiding in diagnostics within fields like radiology and pathology [37]. By leveraging the capabilities of these advanced language models, health care professionals can access valuable

insights and support in making informed decisions regarding treatment plans. Moreover, patients can also reap the advantages of AI-driven technologies by receiving more accurate diagnoses and tailored treatment approaches, ultimately leading to improved outcomes and enhanced patient care [38].

This study does, however, have several important limitations. First, as ASCO-SEP only consists of MCQs, we did not challenge ChatGPT-3.5 with any other question formats (eg, open-ended), which may have yielded different results. Furthermore, MCQs may not fully reflect the complexity of clinical scenarios that oncologists face in their practice. Second, we did not test the variability of the answers provided by ChatGPT. Each question was presented to ChatGPT 3.5 only once, and the first answer was scored given that previous studies showed high consistency of ChatGPT answers [39]. Finally, we could have performed a qualitative assessment of ChatGPT-3.5 answers to gain insights into the etiology of its errors as a guide to future required improvements.

## Conclusions

In conclusion, this study explored the capacity of ChatGPT-3.5's knowledge in medical oncology using the ASCO-SEP. We aimed to bridge the knowledge gaps surrounding the efficacy of AI-driven tools like ChatGPT-3.5 in supporting clinical decision-making. Our assessment revealed that while ChatGPT-3.5 shows promise for the future of AI in oncology, its current performance on ASCO-SEP underscores a pressing need for further refinement to meet the competency standards in this complex field.

Future evaluations of ChatGPT could extend to assessing its capability in clinical decision support, gauging its accuracy in real-life clinical scenarios, and its ease of integration into medical workflows. Evaluating GPT-4 as a resource to aid oncologists in clinical decision-making, an aspect not available during the tenure of this study, could significantly contribute to the field. The tool's facilitation of interdisciplinary collaboration among health care professionals and its impact on patient engagement and communication are other potential areas of investigation.

## Conflicts of Interest

None declared.

## References

1. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. *Future Healthc J.* 2019;6(2):94-98. [FREE Full text] [doi: [10.7861/futurehosp.6-2-94](https://doi.org/10.7861/futurehosp.6-2-94)] [Medline: [31363513](https://pubmed.ncbi.nlm.nih.gov/31363513/)]
2. Brown T, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P, et al. Language models are few-shot learners. *Adv Neural Inf Process Syst.* 2020;33:1877-1901.
3. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. *PLOS Digit Health.* 2023;2(2):e0000198. [FREE Full text] [doi: [10.1371/journal.pdig.0000198](https://doi.org/10.1371/journal.pdig.0000198)] [Medline: [36812645](https://pubmed.ncbi.nlm.nih.gov/36812645/)]
4. CHatGPT goes law school. University of Minnesota Law School. 2023. URL: <https://twin-cities.umn.edu/news-events/chatgpt-goes-law-school#:~:text=MINNEAPOLIS%2FST.achieved%20low%20but%20passing%20grades> [accessed 2023-12-08]
5. Khan RA, Jawaid M, Khan AR, Sajjad M. ChatGPT—reshaping medical education and clinical management. *Pak J Med Sci.* 2023;39(2):605-607. [FREE Full text] [doi: [10.12669/pjms.39.2.7653](https://doi.org/10.12669/pjms.39.2.7653)] [Medline: [36950398](https://pubmed.ncbi.nlm.nih.gov/36950398/)]

6. King MR. A conversation on artificial intelligence, chatbots, and plagiarism in higher education. *Cell Mol Bioeng*. 2023;16(1):1-2. [FREE Full text] [doi: [10.1007/s12195-022-00754-8](https://doi.org/10.1007/s12195-022-00754-8)] [Medline: [36660590](https://pubmed.ncbi.nlm.nih.gov/36660590/)]
7. Large language models. Wikipedia. 2023. URL: [https://en.wikipedia.org/wiki/Large\\_language\\_model](https://en.wikipedia.org/wiki/Large_language_model) [accessed 2023-03-14]
8. Bajwa J, Munir U, Nori A, Williams B. Artificial intelligence in healthcare: transforming the practice of medicine. *Future Healthc J*. 2021;8(2):e188-e194. [FREE Full text] [doi: [10.7861/fhj.2021-0095](https://doi.org/10.7861/fhj.2021-0095)] [Medline: [34286183](https://pubmed.ncbi.nlm.nih.gov/34286183/)]
9. Alfertshofer M, Hoch CC, Funk PF, Hollmann K, Wollenberg B, Knoedler S, et al. Sailing the seven seas: a multinational comparison of ChatGPT's performance on medical licensing examinations. *Ann Biomed Eng*. 2023 [FREE Full text] [doi: [10.1007/s10439-023-03338-3](https://doi.org/10.1007/s10439-023-03338-3)] [Medline: [37553555](https://pubmed.ncbi.nlm.nih.gov/37553555/)]
10. Asch D. An interview with ChatGPT about health care. *NEJM Catal Innov Care Deliv*. 2023;4(2) [FREE Full text]
11. Dave T, Athaluri SA, Singh S. ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. *Front Artif Intell*. 2023;6:1169595. [FREE Full text] [doi: [10.3389/frai.2023.1169595](https://doi.org/10.3389/frai.2023.1169595)] [Medline: [37215063](https://pubmed.ncbi.nlm.nih.gov/37215063/)]
12. Rich AS, Gureckis TM. Lessons for artificial intelligence from the study of natural stupidity. *Nat Mach Intell*. 2019;1(4):174-180. [FREE Full text] [doi: [10.1038/s42256-019-0038-z](https://doi.org/10.1038/s42256-019-0038-z)]
13. Frey CB, Osborne MA. The future of employment: how susceptible are jobs to computerisation? *Technol Forecast Soc Change*. 2017;114:254-280. [FREE Full text] [doi: [10.1016/j.techfore.2016.08.019](https://doi.org/10.1016/j.techfore.2016.08.019)]
14. Homolak J. Opportunities and risks of ChatGPT in medicine, science, and academic publishing: a modern promethean dilemma. *Croat Med J*. 2023;64(1):1-3. [FREE Full text] [doi: [10.3325/cmj.2023.64.1](https://doi.org/10.3325/cmj.2023.64.1)] [Medline: [36864812](https://pubmed.ncbi.nlm.nih.gov/36864812/)]
15. Melnikov AA, Nautrup HP, Krenn M, Dunjko V, Tiersch M, Zeilinger A, et al. Active learning machine learns to create new quantum experiments. *Proc Natl Acad Sci U S A*. 2018;115(6):1221-1226. [FREE Full text] [doi: [10.1073/pnas.1714936115](https://doi.org/10.1073/pnas.1714936115)] [Medline: [29348200](https://pubmed.ncbi.nlm.nih.gov/29348200/)]
16. van Dis EAM, Bollen J, Zuidema W, van Rooij R, Bockting CL. ChatGPT: five priorities for research. *Nature*. 2023;614(7947):224-226. [doi: [10.1038/d41586-023-00288-7](https://doi.org/10.1038/d41586-023-00288-7)] [Medline: [36737653](https://pubmed.ncbi.nlm.nih.gov/36737653/)]
17. ASCO-SEP®. American Society of Clinical Oncology® Store. 2023. URL: <https://shop.asco.org/asco-sep/#:~:text=ASCO%2DSEP%2%AE%20is%20a,and%20cancer%20in%20elderly%20patients> [accessed 2023-03-15]
18. ASCO-SEP 6th edition self-evaluation. American Society of Clinical Oncology® Education. 2023. URL: <https://education.asco.org/product-details/asco-sep-6th-edition-self-evaluation> [accessed 2023-05-05]
19. Gilson A, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA, et al. How does ChatGPT perform on the United States medical licensing examination? The implications of large language models for medical education and knowledge assessment. *JMIR Med Educ*. 2023;9:e45312. [FREE Full text] [doi: [10.2196/45312](https://doi.org/10.2196/45312)] [Medline: [36753318](https://pubmed.ncbi.nlm.nih.gov/36753318/)]
20. Explanations for the 2020-2022 official step 2 CK practice questions. Ben White. 2023. URL: <https://www.benwhite.com/medicine/explanations-for-the-2020-2021-official-step-2-ck-practice-questions/> [accessed 2023-10-22]
21. Das D, Kumar N, Longjam LA, Sinha R, Roy AD, 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(3):e36034. [FREE Full text] [doi: [10.7759/cureus.36034](https://doi.org/10.7759/cureus.36034)] [Medline: [37056538](https://pubmed.ncbi.nlm.nih.gov/37056538/)]
22. Sinha RK, Roy AD, Kumar N, Mondal H. Applicability of ChatGPT in assisting to solve higher order problems in pathology. *Cureus*. 2023;15(2):e35237. [FREE Full text] [doi: [10.7759/cureus.35237](https://doi.org/10.7759/cureus.35237)] [Medline: [36968864](https://pubmed.ncbi.nlm.nih.gov/36968864/)]
23. Chan MW, Eppich WJ. The keyword effect: a grounded theory study exploring the role of keywords in clinical communication. *AEM Educ Train*. 2020;4(4):403-410. [FREE Full text] [doi: [10.1002/aet.2.10424](https://doi.org/10.1002/aet.2.10424)] [Medline: [33150283](https://pubmed.ncbi.nlm.nih.gov/33150283/)]
24. Keywords/Buzzwords on step. Reddit. 2023. URL: [https://www.reddit.com/r/step1/comments/o56ho3/keywordsbuzzwords\\_on\\_step/?rdt=55189](https://www.reddit.com/r/step1/comments/o56ho3/keywordsbuzzwords_on_step/?rdt=55189) [accessed 2023-10-22]
25. Schade M. How ChatGPT and our language models are developed. OpenAI. 2023. URL: <https://help.openai.com/en/articles/7842364-how-chatgpt-and-our-language-models-are-developed> [accessed 2023-10-27]
26. Takagi S, Watari T, Erabi A, Sakaguchi K. Performance of GPT-3.5 and GPT-4 on the Japanese medical licensing examination: comparison study. *JMIR Med Educ*. 2023;9:e48002. [FREE Full text] [doi: [10.2196/48002](https://doi.org/10.2196/48002)] [Medline: [37384388](https://pubmed.ncbi.nlm.nih.gov/37384388/)]
27. Kumah-Crystal Y, Mankowitz S, Embi P, Lehmann CU. ChatGPT and the clinical informatics board examination: the end of unproctored maintenance of certification? *J Am Med Inform Assoc*. 2023;30(9):1558-1560. [doi: [10.1093/jamia/ocad104](https://doi.org/10.1093/jamia/ocad104)] [Medline: [37335851](https://pubmed.ncbi.nlm.nih.gov/37335851/)]
28. Arif TB, Munaf U, Ul-Haque I. The future of medical education and research: Is ChatGPT a blessing or blight in disguise? *Med Educ Online*. 2023;28(1):2181052. [FREE Full text] [doi: [10.1080/10872981.2023.2181052](https://doi.org/10.1080/10872981.2023.2181052)] [Medline: [36809073](https://pubmed.ncbi.nlm.nih.gov/36809073/)]
29. Boehm KM, Khosravi P, Vanguri R, Gao J, Shah SP. Harnessing multimodal data integration to advance precision oncology. *Nat Rev Cancer*. 2022;22(2):114-126. [FREE Full text] [doi: [10.1038/s41568-021-00408-3](https://doi.org/10.1038/s41568-021-00408-3)] [Medline: [34663944](https://pubmed.ncbi.nlm.nih.gov/34663944/)]
30. Foersch S, Glasner C, Woerl AC, Eckstein M, Wagner DC, Schulz S, et al. Multistain deep learning for prediction of prognosis and therapy response in colorectal cancer. *Nat Med*. 2023;29(2):430-439. [FREE Full text] [doi: [10.1038/s41591-022-02134-1](https://doi.org/10.1038/s41591-022-02134-1)] [Medline: [36624314](https://pubmed.ncbi.nlm.nih.gov/36624314/)]
31. Liu J, Wang C, Liu S. Utility of ChatGPT in clinical practice. *J Med Internet Res*. 2023;25:e48568. [FREE Full text] [doi: [10.2196/48568](https://doi.org/10.2196/48568)] [Medline: [37379067](https://pubmed.ncbi.nlm.nih.gov/37379067/)]
32. David Smerdon. X. 2023. URL: <https://twitter.com/dsmerdon/status/1618816703923912704> [accessed 2023-04-30]

33. Teresa Kubacka. X. 2023. URL: [https://twitter.com/paniterka\\_ch/status/1599893718214901760](https://twitter.com/paniterka_ch/status/1599893718214901760) [accessed 2023-04-30]
34. Patel SB, Lam K. ChatGPT: the future of discharge summaries? *Lancet Digit Health*. 2023;5(3):e107-e108. [FREE Full text] [doi: [10.1016/S2589-7500\(23\)00021-3](https://doi.org/10.1016/S2589-7500(23)00021-3)] [Medline: [36754724](https://pubmed.ncbi.nlm.nih.gov/36754724/)]
35. Johnson SB, King AJ, Warner EL, Aneja S, Kann BH, Bylund CL. Using ChatGPT to evaluate cancer myths and misconceptions: artificial intelligence and cancer information. *JNCI Cancer Spectr*. 2023;7(2):pkad015. [FREE Full text] [doi: [10.1093/jncics/pkad015](https://doi.org/10.1093/jncics/pkad015)] [Medline: [36929393](https://pubmed.ncbi.nlm.nih.gov/36929393/)]
36. Londhe VY, Bhasin B. Artificial intelligence and its potential in oncology. *Drug Discov Today*. 2019;24(1):228-232. [FREE Full text] [doi: [10.1016/j.drudis.2018.10.005](https://doi.org/10.1016/j.drudis.2018.10.005)] [Medline: [30342246](https://pubmed.ncbi.nlm.nih.gov/30342246/)]
37. Luchini C, Pea A, Scarpa A. Artificial intelligence in oncology: current applications and future perspectives. *Br J Cancer*. 2022;126(1):4-9. [FREE Full text] [doi: [10.1038/s41416-021-01633-1](https://doi.org/10.1038/s41416-021-01633-1)] [Medline: [34837074](https://pubmed.ncbi.nlm.nih.gov/34837074/)]
38. Huang S, Cai N, Pacheco PP, Narrandes S, Wang Y, Xu W. Applications of Support Vector Machine (SVM) learning in cancer genomics. *Cancer Genomics Proteomics*. 2018;15(1):41-51. [FREE Full text] [doi: [10.21873/cgp.20063](https://doi.org/10.21873/cgp.20063)] [Medline: [29275361](https://pubmed.ncbi.nlm.nih.gov/29275361/)]
39. Suárez A, García VDF, Algar J, Gómez Sánchez M, de Pedro ML, Freire Y. Unveiling the ChatGPT phenomenon: evaluating the consistency and accuracy of endodontic question answers. *Int Endod J*. 2023:108-113. [FREE Full text] [doi: [10.1111/iej.13985](https://doi.org/10.1111/iej.13985)] [Medline: [37814369](https://pubmed.ncbi.nlm.nih.gov/37814369/)]

## Abbreviations

**AI:** artificial intelligence

**ASCO-SEP:** American Society of Clinical Oncology Self-Evaluation Program

**MCQ:** multiple choice question

**USMLE:** United States Medical Licensing Examination

*Edited by K El Emam; submitted 02.07.23; peer-reviewed by D Hu, S Matsuda, J Luo; comments to author 28.07.23; revised version received 05.10.23; accepted 19.11.23; published 12.01.24*

*Please cite as:*

*Odabashian R, Bastin D, Jones G, Manzoor M, Tangestaniapour S, Assad M, Lakhani S, Odabashian M, McGee S*

*Assessment of ChatGPT-3.5's Knowledge in Oncology: Comparative Study with ASCO-SEP Benchmarks*

*JMIR AI 2024;3:e50442*

*URL: <https://ai.jmir.org/2024/1/e50442>*

*doi: [10.2196/50442](https://doi.org/10.2196/50442)*

*PMID:*

©Roupen Odabashian, Donald Bastin, Georden Jones, Maria Manzoor, Sina Tangestaniapour, Malke Assad, Sunita Lakhani, Maritsa Odabashian, Sharon McGee. Originally published in JMIR AI (<https://ai.jmir.org>), 12.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR AI, is properly cited. The complete bibliographic information, a link to the original publication on <https://www.ai.jmir.org/>, as well as this copyright and license information must be included.