

Exploring the Usage of ChatGPT Among Medical Students in the United States

Janice S. Zhang , Christine Yoon, Darnell K. Adrian Williams  and Adi Pinkas

Albert Einstein College of Medicine, Bronx, New York, USA.

Journal of Medical Education and Curricular Development
Volume 11: 1–7
© The Author(s) 2024
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/23821205241264695



ABSTRACT

OBJECTIVES: Chat Generative Pretrained Transformer (ChatGPT) is a large language model developed by OpenAI that has gained widespread interest. It has been cited for its potential impact on health care and its beneficial role in medical education. However, there is limited investigation into its use among medical students. In this study, we evaluated the frequency of ChatGPT use, motivations for use, and preference for ChatGPT over existing resources among medical students in the United States.

METHODS: Data was collected from an original survey consisting of 14 questions assessing the frequency and usage of ChatGPT in various contexts within medical education. The survey was distributed via email lists, group messaging applications, and classroom lectures to medical students across the United States. Responses were collected between August and October 2023.

RESULTS: One hundred thirty-one participants completed the survey and were included in the analysis. Of the total, 48.9% respondents responded that they have used ChatGPT in medical studies. Among ChatGPT users, 43.7% of respondents report using ChatGPT weekly, several times per week, or daily. ChatGPT is most used for writing, revising, editing, and summarizing purposes. 37.5% and 41.3% of respondents reported using ChatGPT more than 25% of the working time for these tasks respectively. Among respondents who have not used ChatGPT, more than 50% of respondents reported they were extremely unlikely or unlikely to use ChatGPT across all surveyed scenarios. ChatGPT users report they are more likely to use ChatGPT over directly asking professors or attendings (45.3%), textbooks (42.2%), and lectures (31.7%), and least likely to be used over popular flashcard application Anki (11.1%) and medical education videos (9.5%).

CONCLUSIONS: ChatGPT is an increasingly popular resource among medical students, with many preferring ChatGPT over other traditional resources such as professors, textbooks, and lectures. Its impact on medical education will only continue to grow as its capabilities improve.

KEYWORDS: Large language model, Medical education, ChatGPT

RECEIVED: March 7, 2024. ACCEPTED: June 8, 2024

TYPE: Original Research Article

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

DECLARATION OF CONFLICTING INTERESTS: The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Adi Pinkas, Department of Pathology, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Forchheimer Building, Room 620SE, Bronx, NY 10461 USA. Email: adi.pinkas@einsteinmed.edu

Introduction

Artificial intelligence (AI) is the study of computer systems capable of performing complex tasks and learning from experience, allowing for abilities traditionally associated with human intelligence from pattern recognition to complex problem-solving and decision-making.^{1,2} These capabilities allow AI to have countless applications in a large range of fields. One type of AI of particular interest is generative AI, which can produce novel text, images, videos, and other types of data based on patterns detected from large amounts of training input data. Large language models (LLMs) are a type of generative AI trained on large volumes of text to learn statistically significant relationships among words for text generation.² LLMs have garnered increasing interest for their applications in a broad range of fields, such as healthcare and education.^{3,4}

One example of an LLM is Chat Generative Pretrained Transformer (ChatGPT), a tool developed by OpenAI with the ability to generate human-like responses to text input.^{1,5} ChatGPT has been cited for its potential impact on healthcare, including its ability to answer patient questions,⁶ to generate

postoperative patient instructions,⁷ and to provide guidelines for breast cancer screening.^{8,9} In addition, ChatGPT has been reported to have several beneficial roles in medical education, including creating summaries and flashcards to facilitate learning, generating quizzes and cases, and providing research assistance.^{3,10} However, there are also concerns regarding the ethical issues of advanced AI tools, including bias, plagiarism, data privacy, and security issues.^{1,11,12}

Several studies have investigated the utility of ChatGPT in healthcare education, research, and practice. A systematic review categorized the benefits/applications of ChatGPT as follows: (1) educational benefits in healthcare education (eg, generation of realistic and variable clinical vignettes, customized clinical cases with immediate feedback based on the student's needs, enhanced communication skills); (2) benefits in academic/scientific writing (eg, text generation, summarization, translation, and literature review in scientific research).¹ Other studies have explored the early applications of ChatGPT in medical practice, education, and research,⁴ as well as its potential for use in undergraduate medical



education,¹³ and its performance on the United States Medical Licensing Examination.¹⁴

Few studies have been conducted on medical students' perceptions of AI in medical education. One study conducted in the United Arab Emirates focused on perceptions and potential uses of ChatGPT and AI in recently graduated medical students.¹⁵ Another study conducted in Germany, Austria, and Switzerland explored perceptions of AI use in medicine and AI ethics in medical education.¹⁶ One study in Malaysia among healthcare students focused on practices and frequency of ChatGPT use.¹⁷

To our knowledge, no investigation into ChatGPT's specific uses among medical students in the United States has been done. Therefore, this study aims to survey the frequency of ChatGPT use among medical students and investigate the motivations for its use, from learning new material to conducting research projects and looking up information in a clinical setting, as well as comparison of ChatGPT with other common resources for medical education. The study's underlying hypothesis is that a minority of medical students have used ChatGPT as a tool for studying, research, and advising. The study provides both quantitative and qualitative insight into ChatGPT use among medical students to characterize its potential for integration into medical education.

Methods

Study Design

The survey was developed based on informal interviews with several students who were familiar with using ChatGPT in medical school as well as educators interested in investigating its impact. A literature review conducted on the growing utility of ChatGPT in medical education supported these questions. The survey consists of 14 questions and is included in Supplemental File S1. The first set of questions assesses whether the respondent has used ChatGPT in the past and if they have used ChatGPT in their medical studies. For those who have used ChatGPT, questions on frequency of ChatGPT use in various areas of medical studies were constructed using a variation of a 5-point Likert scale supplemented with percentages for clarification (never [0% of the working time], rarely [0-25% of the working time], sometimes [25-50% of the working time], often [50-75% of the working time], always [>75% of the working time]).

For those respondents who have not used ChatGPT, we assessed the likelihood of using ChatGPT in the same areas in question using a 5-point Likert scale (extremely unlikely, unlikely, neutral, likely, extremely likely). The second set of questions assessed the likelihood of using ChatGPT over existing medical education resources, measured by the same 5-point Likert scale. The final section of the survey included demographic questions (age, gender identity, medical school, year in medical school, and specialty interest). Each question allowed a single response. Respondents were able to review and change answers throughout the survey.

The electronic questionnaire was constructed using Google Forms. The electronic questionnaire was tested for usability and technical functionality among the study team members. Participants were asked to participate in the open, voluntary, confidential survey through the university-wide email list, class-wide group chats, and classroom lectures. No incentives were offered. Inclusion criteria included students currently enrolled in an MD or MD/PhD program in the United States. Respondents who did not meet these criteria were excluded. The study was determined to be exempt by Albert Einstein College of Medicine's Institutional Review Board (IRB# 2023-15003) for all study respondents. A waiver of signed documentation of consent was approved per 45 CFR 46.117. Responses were collected between August 2023 and October 2023. All responses were automatically saved into Google Drive. No participant identifiers were saved for analysis. Incomplete questionnaires were included in the analysis. The questions and reporting of this study conform to the Checklist for Reporting Results of Internet E-Survey.¹⁸ The checklist is included in Supplemental File S2.

Participants

Invitations to participate in the study were distributed to students at Albert Einstein College of Medicine, Bronx, NY through the university-wide email list, class-wide group messaging applications, and classroom lectures. The survey link was also distributed to students at nine other medical schools through personal contacts of the study team members. The survey link provided a brief background and the objectives of the study. The invitation also notified participants that participation was voluntary and anonymous.

Statistical Analysis

Participant responses were directly captured from Google Forms into exportable data files. Survey results were collected and evaluated using an analysis of descriptive statistics and logistic regression analysis. A 95% confidence interval was utilized in the analysis to determine statistical significance with any *P*-value less than .05. Descriptive statistics were presented with tables and visualized with stacked bar plots using frequencies and percentages. Tables showed the frequencies and percentages of all survey options available to respondents. Missing responses were counted but not included in table percentages. All analyses were performed using Microsoft Excel (Version 16.661) and R (Version 4.2.3).

Results

Respondent Demographics

Of the 740 students at Albert Einstein College of Medicine who were invited to participate, 91 (12.3%) completed the survey. Forty students from other institutions responded to the survey, some of whom chose not to specify their medical school. Table 1 shows the demographic characteristics of

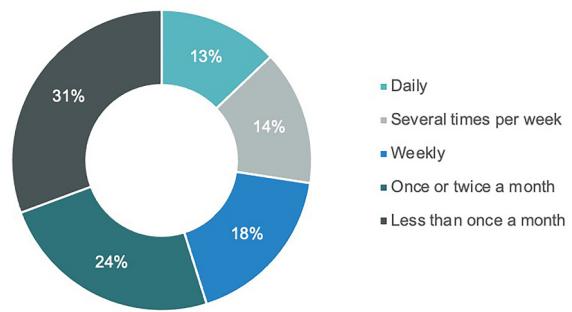
Table 1. Demographic characteristics of respondents.

	ChatGPT non-users	ChatGPT users	Total	P-value
Age				.2377
N (N missing)	66 (6)	65 (2)	131	
M (SD)	25.6 (2.5)	25.2 (1.9)	25.4 (2.2)	
Min–max	21–33	22–32	21–33	
Median	25	25	25	
Gender				.3056
Woman	37	33	70	
Man	26	30	56	
Nonbinary/nonconforming	2		2	
Prefer not to respond	1	2	3	
Medical school year				<.0001
First year (first year preclinical or equivalent)	17	7	24	
Second year (second year preclinical or equivalent)	6	28	34	
Third year (first year clinical or equivalent)	26	16	42	
Fourth year (second year clinical or equivalent)	16	12	28	
Medical School				.6890
Albert Einstein College of Medicine	48	43	91	
Other	13	15	28	
N/A	5	7	12	
Total	66	65	131	

respondents who have used ChatGPT in their medical studies and those who have not. 48.9% of respondents responded that they have used ChatGPT in medical studies.

Frequency of Use Among Students Using ChatGPT

Among respondents who have used ChatGPT in medical studies, 56.3% of respondents reported using ChatGPT less than once or twice per month. 43.7% of respondents reported using ChatGPT weekly, several times per week, or daily. Figure 1 illustrates these responses. Thirty-nine percent of respondents also reported using ChatGPT more than 25% of their working time to learn new material. Figure 2 shows students' frequency of use in several areas of medical education.

**Figure 1.** Overall frequency of ChatGPT use for medical studies. ChatGPT: Chat Generative Pretrained Transformer.

ChatGPT is commonly used for writing and editing purposes. 37.5% and 41.3% of respondents reported using ChatGPT more than 25% of their working time for writing (personal statements, proposals) and revising, editing, and summarizing respectively. Although 54.7% of respondents have never used ChatGPT in research projects (including understanding research topics and generating new ideas), nearly one-third of respondents reported using it sometimes, or even more frequently. ChatGPT is most infrequently used in writing clinical notes and other medical documentation and for looking up information on the wards or in a clinical setting. 79.4% and 63.5% reported never using ChatGPT in these scenarios respectively. Lastly, 73.4% of respondents have never utilized ChatGPT for advising or advice, including but not limited to finding study resources or career path information.

Likelihood of Use Among Students Not Using ChatGPT

Participants who responded that they had not used ChatGPT for medical studies were asked a series of questions on how likely they were to use ChatGPT. These questions correlated with scenarios asked in the frequency questions among students who had previously used ChatGPT. The responses are presented in Figure 3. More than 50% of respondents reported they were extremely unlikely or unlikely to use ChatGPT across all surveyed scenarios. Most notably, 86.6% of respondents were extremely unlikely or unlikely to use ChatGPT for studying for in-house exams and 82.1% of respondents responded similarly to using ChatGPT for writing clinical notes or clinical documentation. These participants were likely or extremely likely to use ChatGPT in revising, editing, and summarizing (29.9%) and for research purposes (22.7%).

Preference of ChatGPT Over Other Resources

Figure 4 compares the preferences of those who have used ChatGPT ("yes") to those who have not ("no"). Among respondents who have used ChatGPT in medical studies, ChatGPT is more likely to be used over directly asking professors or attendings (45.3%), textbooks (42.2%), and lectures (31.7%), and least likely to be used over Anki, a popular open-

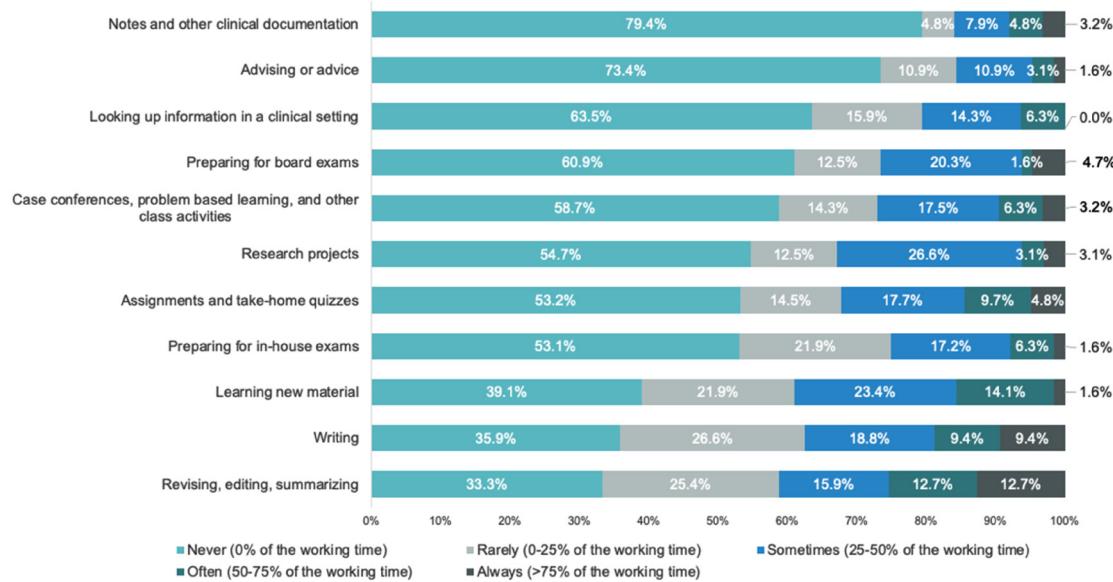


Figure 2. Frequency of ChatGPT use among users in medical school activities.
ChatGPT: Chat Generative Pretrained Transformer.

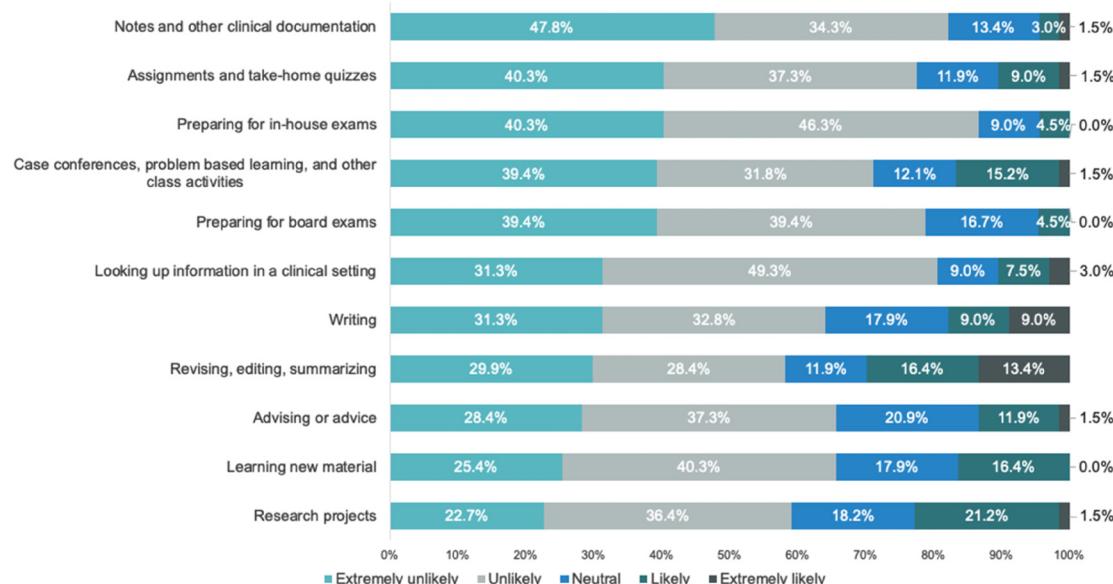


Figure 3. Likelihood of ChatGPT use among non-users in medical school activities.
ChatGPT: Chat Generative Pretrained Transformer.

sourced flashcard application among medical students built on spaced repetition algorithms, (11.1%) and medical education videos (9.5%). On the other hand, respondents who have not used ChatGPT reported that ChatGPT would most likely be used over asking (19.7%), textbook (22.7%), and lecture (9.2%), and least likely to be used over UpToDate (1.5%), board review books (1.5%), journal databases (0%), and medical education videos (0%).

In comparing the two groups, the greatest difference was found to be in preference for ChatGPT over Google searches.

Among ChatGPT users, 53.1% reported unlikely or extremely unlikely to use ChatGPT over conducting a Google search compared to 89.4% among nonusers. Furthermore, 80.0% of non-users reported unlikely or extremely unlikely to use ChatGPT over lecture materials while only 55.6% of ChatGPT users responded similarly.

Discussion

The present study of 131 students provides insight into the frequency and current uses of ChatGPT among medical students

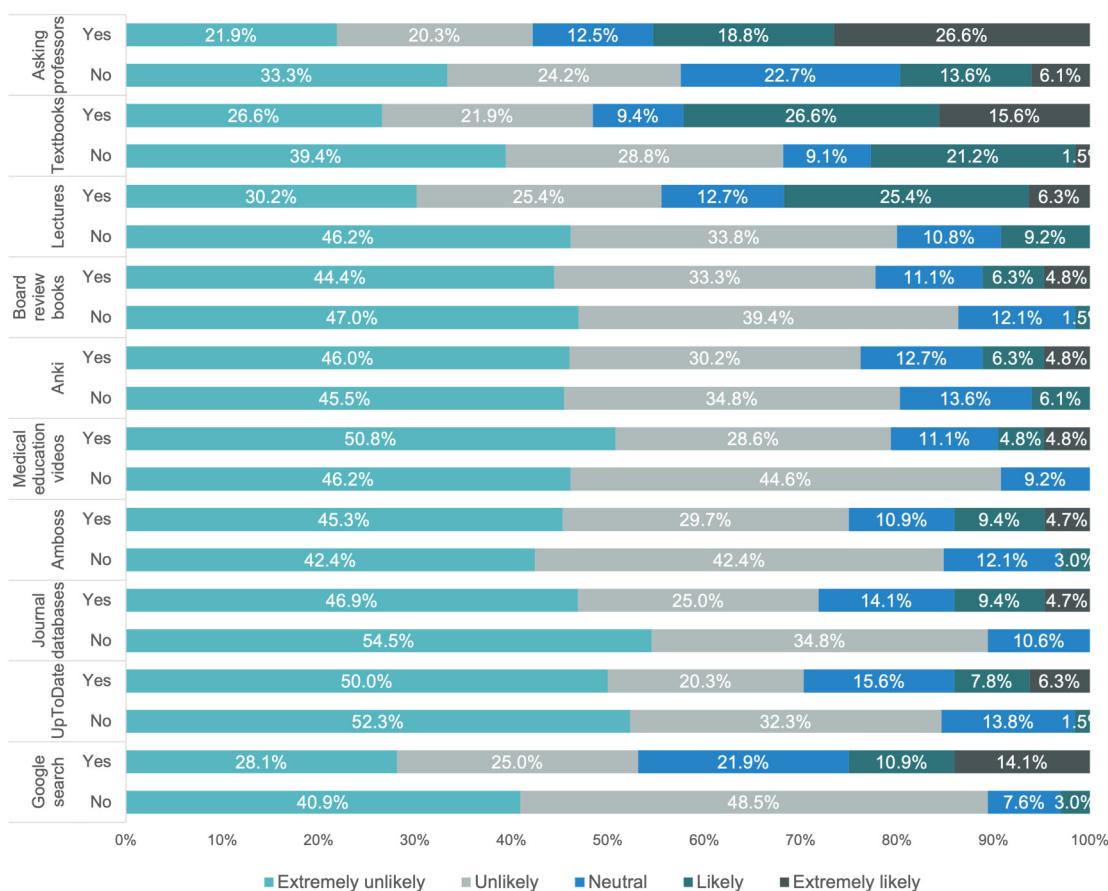


Figure 4. Comparing preference for ChatGPT over other medical education resources between ChatGPT users and non-users. ChatGPT: Chat Generative Pretrained Transformer.

in the United States. Furthermore, the perspectives of those who do not use ChatGPT are explored to uncover areas in which students may be more likely to incorporate AI tools.

In the present study, we reveal that ChatGPT use among medical students is prevalent, but frequency may be varied. 48.9% of students responded that they have used ChatGPT in medical studies. Among respondents who have used ChatGPT for medical studies, many (43.7%), but not most, of them use the resource weekly to daily, especially for learning new material and writing. A European study reports that 28% of survey students using ChatGPT in a medical context and 36.6% of students in Malaysia have used ChatGPT for academic activities.^{16,17} Furthermore, the data among healthcare students in Malaysia report less than 20% of students reported using ChatGPT weekly to daily.¹⁷ Notably, their survey data also suggests that while a significant proportion of university students are aware of ChatGPT's capabilities, there is still a sizable portion of students who lack knowledge about its various functions and limitations. An earlier study among medical students in Jordan found only 11.3% of students had used ChatGPT, a rate considerably lower than recent studies.¹⁹ Given the relative novelty of AI chatbot technology, the frequency of ChatGPT may continue to grow as students become more familiar with its use cases.

We find that ChatGPT users are most commonly using the tool for writing, editing purposes, and grasping new information. Furthermore, a larger percentage of ChatGPT non-users also responded they are likely or extremely like to use ChatGPT for these purposes. A similar finding in the United Arab Emirates revealed that the students were most likely to utilize ChatGPT to complete writing assignments and write case reports among other tasks.¹⁵ Given LLM's strength in querying vast amounts of text data and organizing conceptually appropriate responses, students may readily adopt ChatGPT in this area.

Many reports have recognized ChatGPT as a valuable study companion. It can offer concise explanations and summaries of complex medical concepts in an easy-to-understand format.³ For instance, ChatGPT may help students better understand medical and research literature by summarizing papers in a user-friendly format and answering specific clarification questions. Other use cases of ChatGPT in medical education include the generation of differential diagnoses given a clinical presentation, review of interactive practice cases, and individualized tutoring through practice questions.¹³ In clinical practice, ChatGPT may also help students improve communication skills. For instance, ChatGPT can provide sample cases, dialogs, and explanations for the practice of medical terminology, history-taking, ethical

principles, and effective doctor–patient communication.²⁰ Lastly, a significant portion of surveyed recent medical school graduates plan to use ChatGPT during residency for exploration of medical topics and research and examination preparation,¹⁵ indicating its growing role in future medical practice.

Given ChatGPT’s myriad of capabilities, understanding where students prefer ChatGPT over existing resources is critical for medical educators and administrators. Such information indicates how AI tools are changing learning practices and consequently, areas for curricular change. Among both groups of respondents, ChatGPT is more likely to be used over asking a professor or consulting textbooks or lectures, but less likely to be used compared to UpToDate, Amboss (learning resource platform for medical students),²¹ and Google. These observations may be attributed to ChatGPT’s ability to offer instant and personalized responses, such as summaries and explanations, in a conversational and more user-friendly format. As a result, this accessibility may be more attractive than traditional textbooks, lectures, and waiting for responses from professors or colleagues. Furthermore, this preference for ChatGPT may fit into the broader shift in educational paradigms toward more digital, personalized, and accessible learning tools, which include established resources such as UpToDate, Amboss, and Google. Unlike textbooks, professors, and lectures, these digital resources allow for browsable, objective answers to specific questions efficiently. Unlike ChatGPT however, UpToDate, Amboss, and Google display pages with large funds of knowledge specific to a user’s query that is up to the user to explore and decipher. In the future, as resources such as UpToDate and Amboss become integrated with LLMs, these trends may change to indicate more use and reliance on resources like ChatGPT.

When used correctly, ChatGPT can serve as a powerful tool, but it is far from infallible; such novel technology presents its potential challenges. Students should learn to discern when generated information aligns with current medical knowledge and practices. The application is not necessarily sourced with the most up-to-date information, especially with guidelines and protocols in medicine. This level of competency is similar to what is expected of physicians and researchers when evaluating scientific literature and trusted sources, and medical schools must consider integrating technology-related orientation programs that include training on effective AI interaction. Ethical considerations of ChatGPT in medical education and healthcare also present unprecedented challenges. The risk of data and algorithmic bias, which are present in the design and training of AI models, may interfere with current medical curricula.^{1,22} Discussions of academic dishonesty have also emerged among educators, noting the potential for students to submit plagiarized work, offering unfair advantages, and citing factual errors.²³ Little consensus has been made from the perspective of medical school administrators on the ideal regulatory plan.^{12,13,22} It is evident that AI will play a role in almost all students’ and physicians’ professional lives,

regardless of society’s familiarity with AI, and ensuring future medical professionals and educators can harness its potential to enhance learning and problem-solving safely is key.

The strengths of this study include the detailed nature of survey questions focused on capturing a variety of potential use cases for ChatGPT for medical students. This study is limited by its low effective response rate, at 12.3% at the primary institution, likely due to high message burden, no incentives, and limited reminders. Furthermore, our survey was distributed primarily at one urban medical school in the Northeast United States. The survey was shared among several students at other medical schools, capturing students from nine other medical schools, but was not formally distributed at these institutions due to time and personnel constraints. Such factors limit this study’s generalizability. The questionnaire was independently developed and not validated elsewhere, limiting its generalizability. Pilot testing was only performed among the four members of the research group. The assumption was made that users were utilizing GPT-3.5, the model’s free version; however, no differentiation was made on whether respondents used GPT-3.5 or GPT-4, the paid version with increased model performance. Lastly, response bias is another potential limitation due to the survey-based setup.

As LLMs gain popularity within medicine, education, and research, further large-scale studies are needed to better understand national attitudes toward ChatGPT among medical students. As ChatGPT’s capabilities continue to expand and enhance learning, medical schools would benefit from incorporating its use into medical education, rather than attempting to censor it. Future directions for this study include surveying a larger pool of medical students throughout the United States and investigating differences of use among various aspects of medical school curricula. ChatGPT use should also be surveyed among interns and residents, who may also be using the resource for their learning, clinical practice, and research. Within these contexts, it is also critical to discuss how AI tools can be used productively, ethically, and legally.

Conclusion

ChatGPT is an increasingly popular resource among medical students for learning, writing, and assisting research. This article demonstrates the current trends of ChatGPT use among surveyed medical students, which may continue to grow as its use becomes more popular and ChatGPT’s capabilities continue to improve. Medical educators should understand that ChatGPT may serve as a useful and valuable tool for personalized learning for students and consider the trends in its use to improve upon their curriculum to best serve their students.

Author Contributions

Conception and design: JS Zhang, C Yoon, and A Pinkas; administrative support: JS Zhang and A Pinkas; provision of study materials or patients: all authors; collection and assembly

of data: all authors; data analysis and interpretation: all authors; manuscript writing: all authors; final approval of manuscript: all authors.

Availability of Data and Materials

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

Ethics Approval and Consent to Participate

This study was determined exempt by the IRB of Albert Einstein College of Medicine (IRB# 2023-15003) per 45 CFR 46.104. A waiver of signed documentation of consent was approved per 45 CFR 46.117. All methods were performed following the relevant guidelines and regulations.

ORCID iDs

Janice S. Zhang  <https://orcid.org/0000-0003-3759-9689>

Darnell K. Adrian Williams  <https://orcid.org/0000-0003-1577-8168>

Supplemental Material

Supplemental material for this article is available online.

REFERENCES

1. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. *Healthcare*. 2023;11(6):887.
2. Clusmann J, Kolbinger FR, Muti HS, et al. The future landscape of large language models in medicine. *Communications Medicine*. 2023;3(1):141.
3. Khan RA, Jawaid M, Khan AR, Sajjad M. ChatGPT - reshaping medical education and clinical management. *Pak J Med Sci*. 2023;39(2):605-607.
4. Sam S. Early applications of ChatGPT in medical practice, education and research. *Clin Med*. 2023;23(3):278.
5. AI O. Introducing ChatGPT. <https://openai.com/index/chatgpt/>
6. Ayers JW, Poliak A, Dredze M, et al. Comparing physician and artificial intelligence Chatbot responses to patient questions posted to a public social media forum. *JAMA Intern Med*. 2023;183(6):589-596.
7. Ayoub NF, Lee Y-J, Grimm D, Balakrishnan K. Comparison between ChatGPT and Google search as sources of postoperative patient instructions. *JAMA Otolaryngology–Head & Neck Surgery*. 2023;149(6):556-558.
8. Saraju A, Bruemmer D, Van Iterson E, Cho L, Rodriguez F, Laffin L. Appropriateness of cardiovascular disease prevention recommendations obtained from a popular online chat-based artificial intelligence model. *JAMA*. 2023;329(10):842-844.
9. Haver HL, Ambinder EB, Bahl M, Oluyemi ET, Jeudy J, Yi PH. Appropriateness of breast cancer prevention and screening recommendations provided by ChatGPT. *Radiology*. 2023;307(4):e230424.
10. Moons P, Van Bulck L. ChatGPT: can artificial intelligence language models be of value for cardiovascular nurses and allied health professionals. *Eur J Cardiovasc Nurs*. 2023;22(7):e55-e59.
11. Biswas S. ChatGPT and the future of medical writing. *Radiology*. 2023;307(2):e223312.
12. Shen Y, Heacock L, Elias J, et al. ChatGPT and other large language models are double-edged swords. *Radiology*. 2023;307(2):e230163.
13. Tsang R. Practical applications of ChatGPT in undergraduate medical education. *J Med Educ Curric Dev*. 2023;10:23821205231178449.
14. Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. *PLOS Digital Health*. 2023;2(2):e0000198.
15. Alkhaaldi SMI, Kassab CH, Dimassi Z, et al. Medical student experiences and perceptions of ChatGPT and artificial intelligence: cross-sectional study. Original Paper. *JMIR Med Educ*. 2023;9:e51302.
16. Weidener L, Fischer M. Artificial intelligence in medicine: cross-sectional study among medical students on application, education, and ethical aspects. Original Paper. *JMIR Med Educ*. 2024;10:e51247.
17. George Pallivathukal R, Kyaw Soe HH, Donald PM, Samson RS, Hj Ismail AR. ChatGPT for academic purposes: survey among undergraduate healthcare students in Malaysia. *Cureus*. 2024;16(1):e53032.
18. Eysenbach G. Improving the quality of web surveys: the checklist for reporting results of internet E-surveys (CHERRIES). *J Med Internet Res*. 2004;6(3):e34.
19. Sallam M, Salim NA, Barakat M, et al. Assessing health Students' attitudes and usage of ChatGPT in Jordan: validation study. *JMIR Med Educ*. 2023;9:e48254.
20. Seetharaman R. Revolutionizing medical education: can ChatGPT boost subjective learning and expression? *J Med Syst*. 2023;47(1):61.
21. Amboss. <https://www.amboss.com/us>
22. Wang C, Liu S, Yang H, Guo J, Wu Y, Liu J. Ethical considerations of using ChatGPT in health care. *J Med Internet Res*. 2023;25:e48009.
23. Cotton DRE, Cotton PA, Shipway JR. Chatting and cheating: ensuring academic integrity in the era of ChatGPT. *Innov Educ Teach Int*. 2024;61(2):228-239.