

Viewpoint

Large Language Models in Medical Education: Opportunities, Challenges, and Future Directions

Alaa Abd-alrazaq¹, PhD; Rawan AlSaad^{1,2}, PhD; Dari Alhuwail³, PhD; Arfan Ahmed¹, PhD; Padraig Mark Healy⁴, MSc; Syed Latifi⁴, PhD; Sarah Aziz¹, MSc; Rafat Damseh⁵, PhD; Sadam Alabed Alrazak⁶, BSc; Javaid Sheikh¹, MD

¹AI Center for Precision Health, Weill Cornell Medicine-Qatar, Doha, Qatar

²College of Computing and Information Technology, University of Doha for Science and Technology, Doha, Qatar

³Information Science Department, College of Life Sciences, Kuwait University, Kuwait, Kuwait

⁴Office of Educational Development, Division of Medical Education, Weill Cornell Medicine-Qatar, Doha, Qatar

⁵Department of Computer Science and Software Engineering, United Arab Emirates University, Abu Dhabi, United Arab Emirates

⁶Department of Mechanical & Industrial Engineering, Faculty of Applied Science and Engineering, University of Toronto, Toronto, ON, Canada

Corresponding Author:

Alaa Abd-alrazaq, PhD

AI Center for Precision Health

Weill Cornell Medicine-Qatar

PO Box 5825, Doha Al Luqta St

Ar-Rayyan

Doha, NA

Qatar

Phone: 974 55708549

Email: alaa_alzoubi88@yahoo.com

Abstract

The integration of large language models (LLMs), such as those in the Generative Pre-trained Transformers (GPT) series, into medical education has the potential to transform learning experiences for students and elevate their knowledge, skills, and competence. Drawing on a wealth of professional and academic experience, we propose that LLMs hold promise for revolutionizing medical curriculum development, teaching methodologies, personalized study plans and learning materials, student assessments, and more. However, we also critically examine the challenges that such integration might pose by addressing issues of algorithmic bias, overreliance, plagiarism, misinformation, inequity, privacy, and copyright concerns in medical education. As we navigate the shift from an information-driven educational paradigm to an artificial intelligence (AI)-driven educational paradigm, we argue that it is paramount to understand both the potential and the pitfalls of LLMs in medical education. This paper thus offers our perspective on the opportunities and challenges of using LLMs in this context. We believe that the insights gleaned from this analysis will serve as a foundation for future recommendations and best practices in the field, fostering the responsible and effective use of AI technologies in medical education.

(*JMIR Med Educ* 2023;9:e48291) doi: [10.2196/48291](https://doi.org/10.2196/48291)

KEYWORDS

large language models; artificial intelligence; medical education; ChatGPT; GPT-4; generative AI; students; educators

Introduction

We are witnessing a significant paradigm shift in the field of artificial intelligence (AI) due to the emergence of large-scale self-supervised models that can be leveraged to automate a wide variety of downstream tasks. These models are now referred to as *foundation models*, with many notable examples, such as OpenAI's GPT-4 [1] and DALL-E [2], Meta's SAM (Segment Anything Model) [3] and LLaMA [4], and Google's LaMDA (Language Models for Dialog Applications) [5] and large-scale

ViT (Vision Transformer) [6]. These models are trained on massive amounts of data and are capable of performing tasks related to natural language processing, computer vision, robotic manipulation, and computer-human interaction. Language-based foundation models, or large language models (LLMs), can understand and generate natural language text, allowing them to engage in human-like conversations, with coherent and contextually appropriate responses to user prompts. Remarkably, due to the advancement of these large-scale AI systems, they

are now able to generate human-like content (eg, texts, images, codes, audio, and videos).

The Generative Pre-trained Transformers (GPT) series models launched by OpenAI are examples of foundation models that are based on generative AI (ie, AI models used to generate new content, such as texts, images, codes, audio, and videos, based on the training data they have been exposed to). OpenAI launched the first model of the GPT series (GPT-1) in 2018, followed by GPT-2 in 2019, GPT-3 in 2020, ChatGPT in 2022, and GPT-4 in 2023, with each iteration representing significant improvements over the previous one. GPT-4 is one of the most advanced AI-based chatbots available today. GPT-4 is an advanced multimodal foundation model that has state-of-the-art performance in generating human-like text based on user prompts [1]. Unlike previous GPT series models (eg, ChatGPT, GPT-3, and GPT-2), which accept only text inputs, GPT-4 can process image inputs, in addition to text inputs, to return textual responses [1]. Furthermore, GPT-4 has a larger model size (more parameters); has been trained on a larger amount of data; and can generate more detailed responses (more than 25,000 words), with a high level of fidelity [7]. Based on rigorous experimentation, GPT-4 capabilities demonstrate improved reasoning, creativity, safety, and alignment and the ability to process complex instructions [1]. As a result, GPT-4 is now actively used by millions of users for language translation, sentiment analysis, image captioning, text summarization, question-answering systems, named entity recognition, content moderation, text paraphrasing, personalized recommendations, text completion and prediction, programming code generation and debugging, and so forth.

Undoubtedly, the versatility and capabilities of current generative AI and LLMs (eg, GPT-4) will revolutionize various domains, with one of particular interest being medical education. The integration of such technologies into medical education

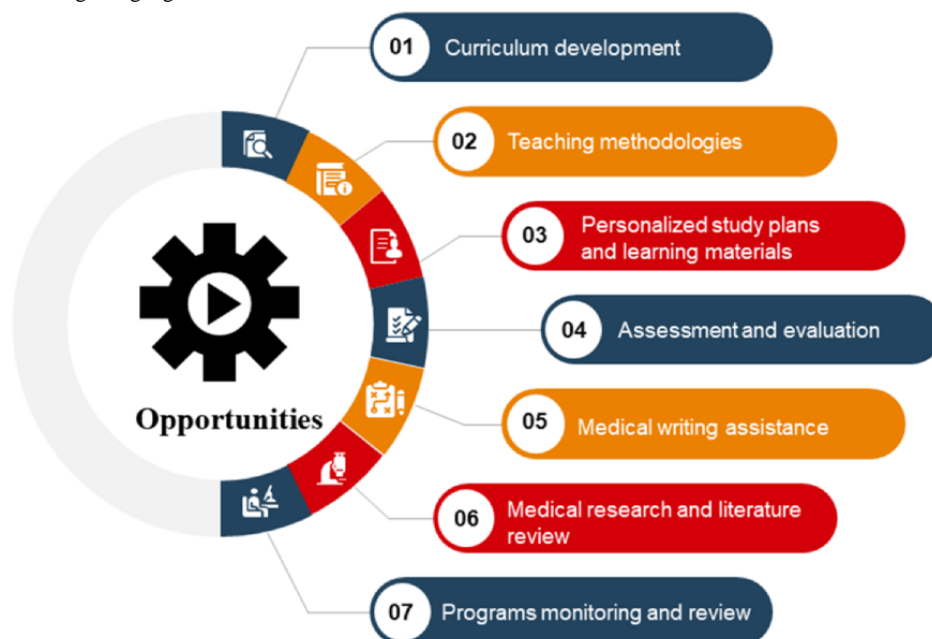
offers numerous opportunities for enhancing students' knowledge, skills, and competence. For instance, LLMs can be used to produce clinical case studies, act as virtual test subjects or virtual patients, facilitate and accelerate research outputs, develop course plans, and provide personalized feedback and assistance. However, their adoption in medical education presents serious challenges, such as plagiarism, misinformation, overreliance, inequity, privacy, and copyright issues. In order to shift medical education practices from being information-driven to being AI-driven through the use of LLMs, it is essential to acknowledge and address the concerns and challenges associated with the adoption of LLMs. This is necessary to ensure that students and educators understand how to use these tools effectively and appropriately to fully leverage their potential. To this end, the objective of this paper is to explore the opportunities, challenges, and future directions of using LLMs in medical education. This paper uses GPT-4 as a case study to discuss these opportunities and challenges, as it is a state-of-the-art generative LLM that was available at the time of writing.

Opportunities

Overview

LLMs have the potential to significantly impact all phases of medical education programs, offering numerous benefits in various aspects, including curriculum planning, delivery, assessments, programmatic enhancements, and research [8-30]. This section elucidates and illustrates the specific opportunities and applications of LLMs that can be leveraged to deliver a more efficient, effective, personalized, and engaging medical education system that is better equipped to prepare future health care professionals. [Figure 1](#) shows the main opportunities of LLMs in medical education.

Figure 1. Opportunities of large language models in medical education.



Curriculum Development

Medical curriculum planning is a complex process that requires careful consideration of various factors, including educational objectives, teaching methodologies, assessment strategies, and resource allocation [31]. LLMs, like GPT-4, can play a significant role in enhancing this process by conducting needs assessments and analyses and providing expert-level knowledge and insights on various medical topics, helping educators identify content gaps and ensure comprehensive coverage of essential subjects [8,17]. Additionally, GPT-4 can assist in developing measurable learning objectives for each phase of a medical program curriculum and customizing it to meet the diverse needs of individual learners, fostering personalized and adaptive learning experiences. By analyzing students' performance data, LLMs can suggest targeted interventions and recommend specific resources to address learning gaps and optimize educational outcomes [16,17]. Furthermore, this integration of LLMs and GPT-4 into medical curriculum planning can support faculty in designing, updating, or modifying a medical curriculum; LLMs can provide suggestions for course content, learning objectives, and teaching methodologies based on the emerging trends and best practices in medical education, freeing up more time for faculty to focus on other teaching aspects [32-34].

Teaching Methodologies

LLMs, like GPT-4, can be used to augment existing teaching methodologies in medical education programs, enhancing the overall learning experience for students. For example, LLMs can supplement lecture content by providing real-time clarifications, additional resources, and context to complex topics, ensuring a deeper understanding for students [35,36]. For small-group sessions, GPT-4 can facilitate discussions by generating thought-provoking questions, encouraging peer-to-peer interactions, and fostering an engaging and collaborative learning environment. For virtual patient simulations, LLMs can create realistic virtual patient scenarios, ask questions, interpret responses, and provide feedback, allowing students to practice clinical reasoning, decision-making, and communication skills in a safe and controlled setting. For interactive medical case studies, GPT-4 can generate case studies that are tailored to specific learning objectives and guide students through the diagnostic process, treatment options, and ethical considerations, thereby allowing students to interactively explore both common conditions and rare conditions, which can help to prepare them for real-world clinical practice [37]. An example of using ChatGPT (GPT-4) to create interactive case studies for medical students is included in [Multimedia Appendix 1](#). For clinical rotations, as virtual mentors, LLMs can help students apply theoretical knowledge to real-world situations by offering instant feedback and personalized guidance to reinforce learning and address misconceptions.

Personalized Study Plans and Learning Materials

By leveraging the power of LLMs and generative AI tools, students can input information about their individual strengths, weaknesses, goals, and preferences to generate study plans that are tailored to their specific needs. This level of personalization

ensures that each student's unique learning style and pace are taken into account, leading to more efficient and effective learning [38]. Moreover, LLMs, like GPT-4, can also generate personalized learning materials, including concise summaries, flash cards, and practice questions, that target specific areas where a student needs improvement. An example of using an LLM, like ChatGPT, to provide personalized explanations of medical terminology (ie, *aphthous stomatitis*) to students at different levels (premedical students, year 2 medical students, and year 4 medical students) is presented in [Multimedia Appendix 2](#). Tailored resources can help students focus on the most relevant content, optimizing their study time and enhancing knowledge retention. Furthermore, an iterative feedback loop could be established wherein students use LLM-generated materials and provide feedback, which is then used to fine-tune the LLM's outputs. Over time, this could lead to increasingly accurate and effective personalized learning materials.

Assessment and Evaluation

LLMs and GPT-4 can play a significant role in designing comprehensive assessment plans and enhancing the evaluation process in medical education [14,18,26]. They can be utilized to (1) develop comprehensive, well-rounded assessment plans that incorporate formative and summative evaluations, competency-based assessments, and effective feedback mechanisms; (2) align assessment methods with learning objectives by analyzing learning objectives and suggesting appropriate assessment methods that accurately measure students' progress toward achieving the desired competencies; and (3) provide prompt feedback and rubrics by automating the process of providing timely and actionable feedback to students, identifying areas of strength and weakness, and offering targeted suggestions for improvement. Additionally, GPT-4 can assist in the creation of transparent and consistent grading rubrics, ensuring that students understand the expectations and criteria for success.

Medical Writing Assistance

LLMs have become valuable tools in medical writing, offering a range of benefits to medical students and medical researchers [37,39-43]. LLMs, like GPT-4, can assist medical students and educators in selecting appropriate language, terminology, and phrases for use in their writing, ensuring accuracy and readability for their intended audience. Furthermore, LLMs can provide guidance on writing style and formatting, helping students to improve the clarity and coherence of their work. By leveraging these chatbots' capabilities, medical students can streamline their writing process and produce high-quality work, resulting in time that can be reallocated to other aspects of their studies.

Medical Research and Literature Review

LLMs are valuable tools for medical research and literature reviews, providing a faster, more efficient, and more accurate means of gathering and analyzing data [26,28,44-46]. With the ability to access, extract, and summarize relevant information from scientific literature, electronic medical records, and other sources, these chatbots enable medical students and researchers to quickly and efficiently gather the information they need for

their reports, papers, and research articles. By leveraging the data extraction capabilities of LLMs, medical students and researchers can more easily access and analyze the vast amounts of information available to them (Multimedia Appendix 3). This ensures that their research is grounded in accurate and reliable data, allows them to make well-informed conclusions based on their findings, and frees up valuable time and resources that can be directed toward other important aspects of the research process. Moreover, when writing research papers, medical students can use LLMs for help with generating outlines and drafting introductions or conclusions; LLMs can also suggest possible ways to discuss and analyze results (Multimedia Appendix 4).

Program Monitoring and Review

LLMs and generative AI tools, when integrated into curriculum management systems, have enormous potential to transform the monitoring and review of medical education programs. By analyzing data collected through various sources, including student feedback, testing results, and program delivery data,

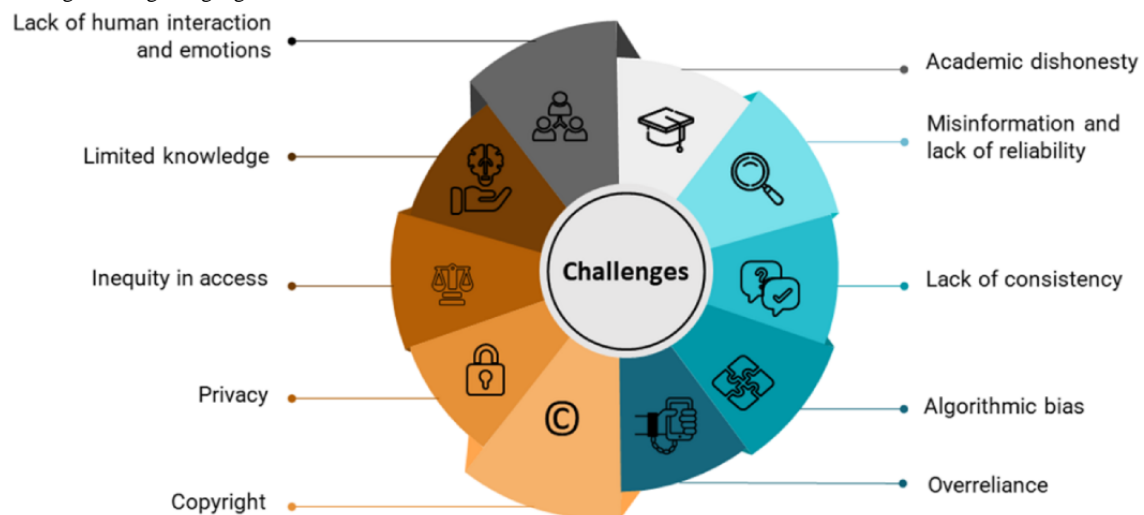
LLMs like GPT-4 can provide program leaders with valuable insights into the efficacy of their programs. LLMs can identify areas of improvement, monitor trends in student performance, and provide benchmarks against which program performance can be evaluated. LLMs can also analyze national health priorities and community needs to help programs adapt and adjust their objectives and allocation of resources accordingly. By leveraging these tools, program leaders can gain insights and make data-driven decisions that enhance the quality and effectiveness of medical education programs.

Challenges

Overview

Despite the abovementioned opportunities that LLMs and generative AI tools can provide, they have limitations in medical education. These challenges and limitations are discussed in the following subsections. Figure 2 shows the main challenges of LLMs in medical education.

Figure 2. Challenges of large language models in medical education.



Academic Dishonesty

The ability of LLMs to respond to short-answer and multiple-choice exam questions can be exploited for cheating purposes [47]. As mentioned earlier, LLMs can write medical essays that are difficult to distinguish from human-generated essays, which may increase plagiarism. Although several tools (eg, GPTZero, Originality.AI, OpenAI AI Text Classifier, and Turnitin AI Writing Detector) have been developed to detect AI-generated text, students may still be able to make their AI-generated essays undetectable to such tools. Specifically, a study demonstrated that adding 1 word (“amazing”) to an AI-generated text reduced the fake level (ie, generated by AI) detected by a tool from 99% to 24% [48]. Although this is just 1 example, it still increases and highlights apprehensions regarding the effectiveness of such tools in detecting and preventing plagiarism.

Misinformation and Lack of Reliability

Although recent LLMs (eg, GPT-4) have significantly reduced hallucinations in comparison with earlier models [1], due to

inaccurate training data, recent LLMs still generate incorrect or inaccurate information that is convincingly written. Given the authoritative writing style generated by these systems, students may find it challenging to differentiate between genuine knowledge and unverified information. As a result, they may not scrutinize the validity of information and end up believing inaccurate or deceptive information [49]. Further, such misinformation may make LLMs untrustworthy among users and thus may decrease the adoption of LLMs. As an example of misinformation, studies showed that LLMs, such as GPT-4, either include citations that do not exist in generated articles or include citations that are irrelevant to the topic [41,50-52]. This raises the question of how to guarantee that generative AI tools and LLMs remain assistive technologies and not propagators of false or misleading health information.

Lack of Consistency

Recent LLMs and generative AI tools generate different outputs for the same prompt. Although this feature may be helpful in some cases, it has several disadvantages [53]. First, generating different responses to the same prompt may prevent educators

from detecting whether the text was generated by AI. Second, this feature may produce contradicting responses on the same topic. Finally, this feature may generate responses with different qualities. For example, in a study [48], 3 researchers at the same location asked an LLM-based chatbot the exact same question at the same time, but they received 3 different responses of different quality. Specifically, the first researcher received a more up-to-date, complete, and organized response compared to the responses that the second and third researchers received [48]. Accordingly, one may inquire about the methods to guarantee fair access, for all users (students and educators), to identical, up-to-date, and high-quality learning materials.

Algorithmic Bias

Given that recent LLMs (eg, GPT-4) are trained on a large corpus of text data from the internet (eg, websites, books, news articles, scientific papers, and movie subtitles), it is likely that they are trained on biased or unrepresentative data. OpenAI has acknowledged that GPT-4 may still generate biased responses like earlier GPT models, thereby reinforcing social biases and stereotypes [1]. For example, if an LLM was trained on data related to disease among a certain ethnic group, then it is likely that it generates responses (eg, essays, exams, and clinical case scenarios) that are biased toward that group. According to a study [54], an LLM that was trained on a vast corpus of internet text demonstrated gender bias in its output.

Overreliance

As mentioned earlier, recent generative AI tools (eg, GPT-4) have a tendency to make up facts and present incorrect information in more convincing and believable ways [1]. This may cause users to excessively trust generative AI tools, thereby increasing the risk of overreliance. Therefore, the use of generative AI tools may hinder the development of new skills or even lead to the loss of skills that are foundational to medical student development, such as critical thinking, problem-solving, and communication. In other words, the ease with which generative AI tools can provide answers could lead to a decrease in students' motivation to conduct independent investigations and arrive at their own conclusions or solutions. This raises the question of how generative AI tools can be used to improve rather than reduce critical thinking and problem-solving in students.

Lack of Human Interaction and Emotions

Current LLMs are unable to deliver the same degree of human interaction as an actual educator or tutor. This is because, at present, (1) their capabilities are restricted to a textual interface, (2) they are incapable of recognizing the physical gestures or movements of students and educators, and (3) they cannot reveal any emotions. The absence of human interaction can negatively affect students who prefer a personal connection with their educator. According to a study conducted by D'Mello and colleagues [55], students who engaged with a virtual tutor that imitated human-like emotional behavior demonstrated superior learning outcomes compared to those who engaged with a virtual tutor that lacked such behavior. Hence, it is worth considering ways to humanize generative AI tools not just in their ability

to think and provide responses but also in terms of exhibiting emotions and possessing a distinctive personality.

Limited Knowledge

LLMs, like GPT-4, depend on the data used for training, which cover a wide range of general information but might not always encompass the latest or most specialized medical knowledge. This constraint impacts the reliability and precision of the information generated by LLMs in medical education environments, where accuracy and expertise are essential [26]. Moreover, the knowledge base of most LLMs is presently static, which means that they cannot learn and adjust in real time as new medical information emerges. However, the field of medicine is constantly evolving, with novel research findings, guidelines, and treatment protocols being regularly introduced [56]. Additionally, the restricted knowledge of current LLMs in medical education could result in a superficial understanding of complex medical concepts, lacking the necessary depth and context for effective learning. For instance, while GPT-4 can produce text that seems coherent and factually correct at first glance, it may not always capture the subtleties and complexities of medical knowledge, thus falling short in providing comprehensive and accurate guidance for medical students and educators.

Inequity in Access

Generative AI tools and LLMs may increase the inequity among students and educators, given that these tools are not equally accessible to all of them. For example, although most generative AI tools can communicate in several languages, in addition to English, and outperform earlier chatbots in this aspect, their proficiency in each language varies based on the amount and quality of training data available for each language [1]; thus, students and educators who are not proficient in English are less likely to use them. Further, generative AI tools may be less accessible to (1) those who are not familiar with using technologies or AI tools; (2) those who do not have access to the necessary technology (eg, internet and computers); (3) those who cannot afford subscription fees (eg, US \$20/month for GPT-4); and (4) those with disabilities, such as blindness or motor impairment.

Privacy

When communicating with LLMs, students and educators may reveal their personal information (eg, name, email, phone number, prompts, uploaded images, and generated images). OpenAI acknowledges that it may use users' personal information for several purposes, such as analyzing, maintaining, and improving its services; conducting research; preventing fraud, criminal activity, or misuse of its services; and complying with legal obligations and legal processes [57]. Moreover, OpenAI may share users' personal information with third parties without further notice to users or users' consent [57]. A recent reflection of these concerns is Italy's data protection group discontinuing access to ChatGPT while it conducts an investigation around data use and collection practices, in alignment with requirements of the General Data Protection Regulation [58]. In addition, LLM use during clerkship clinical rotations for patient care (eg, SOAP [Subjective, Objective,

Assessment, and Plan] note generation) could result in unintended patient privacy breaches. Questions surrounding how to safeguard student and patient data should be central in curricular discussions.

Copyright

LLMs may be trained on copyrighted materials (eg, books, scientific articles, and images), thereby potentially producing text that bears similarity to or even directly copies content protected by copyright, which could potentially impact downstream uses. Such a situation brings up apprehensions regarding the utilization of content created by generative AI tools (eg, educational materials, presentations, course syllabi, quizzes, and scientific papers) without appropriate acknowledgment and authorization from the copyright holder. There are ongoing discussions related to authorship rights for articles that are written by using LLMs. Although various publishers and editors do not accept listing such tools as coauthors (eg, those of *Nature*, *Jinan Journal*, and *eLife*), others do (eg, those of *Oncoscience* [59], *Nurse Education in Practice* [60], and medRxiv [61]). As this is an area likely to evolve, it raises questions regarding how students and educators should acknowledge the use of these systems while complying with professional and regulatory expectations.

Future Directions

Overview

Considering the opportunities and challenges presented by the use of LLMs and generative AI tools in medical education, we discuss future directions, targeting academic institutions, educators, students, developers, and researchers. We argue that those who embrace the use of the technology, including LLMs, will challenge the status quo and will likely be better positioned and higher performing than those who do not. Therefore, the following recommendations and future directions can be useful to all of the previously mentioned stakeholders and many others.

Academic Institutions

With the rise of generative AI tools and LLMs, there is a fear that in the future, these technologies may make the human brain dormant in nearly all tasks, including some of the basic ones. Now more than ever, medical schools and academic institutions need to consider the appropriate strategies to incorporate the use of LLMs into medical education. One possibility is to develop guidelines or best practices for the use of AI tools in their assignments. These guidelines should explain to students how to properly disclose or cite any content generated by LLMs when writing essays, research papers, and assignments. Academic institutions may also subscribe to tools that can detect AI-generated text, such as Turnitin, ZeroGPT, and Originality.AI. Academic institutions should provide training sessions and workshops to teach students and educators how to effectively and ethically use such tools in medical education. Ultimately, academic institutions should favor student-centered pedagogy that nurtures building trusting relationships that focus on *assessment for learning* and do not entirely focus on *assessment of learning* [62].

Educators

Given the rapid, explosive advances driven by the expected use of GPT-4 and other LLMs, medical educators are encouraged to embrace these technologies rather than stay away from them. With AI's rapid evolution, it is paramount for medical educators to upskill their competencies in utilizing generative AI tools effectively within medical curricula. Current medical curricula do not include education on the proper use of AI. Content covering such technologies and their application to medicine (eg, disease discovery) should be included. Medical educators should consider how LLMs can be integrated into medical education, thus requiring them to reconsider the teaching and learning process. This can be done through updating course syllabi to set and clarify the objectives of the use of LLMs (eg, GPT-4), as well as by reflecting on their use in practice and their impact on the profession.

Assignments will also have to be reconsidered, and educators should strive to assign multimodal activities that require high-order thinking, creativity, and teamwork. For example, educators could use oral exams and presentations, hands-on activities, and group projects to assess their students' analytical and critical reasoning, the soundness and precision of their arguments, and their persuasive capabilities. Educators may consider involving students in peer evaluations and exercise "teach-back."

Because health care is complex and often involves high stakes, it is paramount that educators also explain to their medical students the abovementioned limitations of LLMs. For example, educators should highlight the importance of proper citation and attribution in medical school, as well as how to avoid potential user privacy and copyright issues, misinformation, and biases. We recommend that educators discourage reliance that can lead to reduced clinical reasoning skills. Instead, educators should encourage students to check, critique, and improve responses generated by LLMs. Educators should emphasize that these technological tools should be continually monitored by human experts and that they should be used with guidance and critical thinking before acting on any of their recommendations.

Although LLMs, like GPT-4, are powerful tools capable of generating detailed, personalized study plans and learning materials, they are not infallible. They are as good as the data that they have been trained on, and there is always a risk of inaccuracies or misinterpretations, particularly when dealing with complex, nuanced fields, such as medical education. Therefore, we believe that it is crucial to incorporate human input or expert-reviewed content into the process of developing such tools. For instance, subject matter experts, such as experienced medical educators or practitioners, could review and validate the content generated by an LLM. They could provide the correct context, ensure that the material aligns with current medical standards and guidelines, and verify the content's relevance to students' specific learning needs.

Students

Students should ethically and safely use these tools and technologies in a constructive manner to thrive outside of the

classroom, in a world that is rapidly being dominated by AI. Similar to educators, medical students also need to elevate their skills and competencies in effectively leveraging and utilizing generative AI tools and LLMs in their practices. It is paramount that students acknowledge the use of LLMs in their medical and academic work and, at the same time, do so ethically and responsibly.

Developers

Developers of generative AI tools bear the responsibility of meticulously developing generative AI tools while taking into account prevalent constraints, such as inequality, privacy, impartiality, contextual understanding, human engagement, and misinformation. Although recent generative AI technologies, like GPT-4 and ChatGPT, possess the ability to communicate in various languages, their performance is notably more effective in English compared to their performance in other languages. This could be attributed to the lack of data sets and corpora in languages other than English (eg, Arabic) [63]. Developers and researchers should collaborate to build large data sets and corpora in other languages to improve the performance of LLMs when using such languages [63,64]. To tackle the challenges of fairness and equity, developers need to create generative AI technologies that can accommodate the varied requirements and backgrounds of users, particularly for underprivileged or marginalized students and educators. For example, developers ought to equip generative AI tools with the capability to interact with students and educators through voice, visuals, and videos, as well as text, to make them more humanized and accessible to those with disabilities (eg, blindness).

With some generative AI tools creating or “faking” certain articles or information, it is essential for developers to clearly state and discern facts from fiction in the outputs. Additionally, developers should also make an effort to develop more humanized LLMs that consider the virtual relationship that has been developed between humans and machines. The development of generative AI tools should rely on various theories that consider relationship formation among humans, such as social exchange theory. When developing generative AI technologies, it is also essential to adhere to user-focused design principles while taking into account the social, emotional, cognitive, and pedagogical dimensions [65]. We recommend that developers create responsible generative AI tools that correspond with core human principles and comply with our legal system.

Developers play a crucial role in integrating ChatGPT into medical education platforms, drawing inspiration from its use in popular educational platforms, such as Duolingo and Khan Academy. By examining these examples, they can design and develop innovative learning experiences for medical students who use LLMs. Duolingo and Khan Academy use ChatGPT to provide personalized learning experiences based on the individual needs and progress of each student. This approach can be adopted in medical education to create tailored study plans and learning materials that cater to the unique strengths, weaknesses, and learning styles of medical students. Both Duolingo and Khan Academy use ChatGPT to offer real-time feedback and guidance to learners as they engage with the

platform. In the context of medical education, ChatGPT could be integrated into learning management systems or virtual patient simulations to provide instant feedback on students' performance, diagnostic decisions, or treatment plans. By giving students immediate access to targeted guidance and correction, ChatGPT can facilitate continuous improvement and foster a deeper understanding of medical principles. Duolingo utilizes ChatGPT to create interactive, conversation-based lessons that help learners practice their language skills in a more engaging and natural manner. Similarly, ChatGPT can be used in medical education to develop interactive learning modules that allow students to practice clinical communication skills, such as taking patient histories, explaining diagnoses, or discussing treatment options. Khan Academy leverages ChatGPT to facilitate peer-to-peer interactions and support, enabling students to learn from each other and collaborate on problem-solving tasks. In medical education, ChatGPT could be used to create virtual study groups, in which students can discuss clinical cases, share insights, and work together to solve complex medical problems.

Researchers

There is an urgent need to conduct more empirical and evidence-based human-computer interaction and user interface design research for the use of LLMs in medical education. Researchers should explore ways to strike a balance between using these technologies and maintaining the essential human interaction and feedback in education to enhance learning and teaching experiences and outcomes [48]. Further, research is required to investigate the impact of LLMs on students' learning processes and outcomes. Lastly, there is a need to delve deeper into the possible consequences of overdependence on LLMs in medical education [48].

Conclusion

In conclusion, LLMs are double-edged swords. Specifically, LLMs have the potential to revolutionize medical education, enhance the learning experience, and improve the overall quality of medical education by offering a wide range of applications, such as acting as a virtual patient and medical tutor, generating medical case studies, and developing personalized study plans. However, LLMs do not come without challenges. Academic dishonesty, misinformation, privacy concerns, copyright issues, overreliance on AI, algorithmic bias, lack of consistency and human interaction, and inequity in access are some of the major hurdles that need to be addressed.

To overcome these challenges, a collaborative effort is required from educators, students, academic institutions, researchers, and developers of generative AI tools and LLMs. Rather than banning them, medical schools and academic institutions should embrace generative AI tools and develop clear guidelines and rules for the use of these technologies for academic activities. Institutional efforts may be required to help students and educators develop the skills necessary to incorporate the ethical use of AI into medical training. Educators should use new teaching philosophies and redesign assessments and assignments to allow students to use such technologies. Students should ethically and safely use these technologies in a constructive manner. Developers have a duty to carefully design such

technologies while considering common limitations, such as interaction, and misinformation. inequity, privacy, unbiased responses, lack of context and human

Conflicts of Interest

A Abd-alrazaq is an Associate Editor of *JMIR Nursing* at the time of this publication. The other authors have no conflicts of interest to declare.

Multimedia Appendix 1

Example of using ChatGPT (GPT-4) to create interactive case studies for medical students.

[\[DOCX File , 2248 KB-Multimedia Appendix 1\]](#)

Multimedia Appendix 2

Example of using ChatGPT (GPT-4) to provide personalized explanations of medical terminology to students at different levels.

[\[DOCX File , 1794 KB-Multimedia Appendix 2\]](#)

Multimedia Appendix 3

Example of using large language models (Petal) for document analysis.

[\[DOCX File , 677 KB-Multimedia Appendix 3\]](#)

Multimedia Appendix 4

Example of using ChatGPT (GPT-4) to provide an outline and references for research papers.

[\[DOCX File , 2362 KB-Multimedia Appendix 4\]](#)

References

1. OpenAI. GPT-4 technical report. arXiv Preprint posted online on March 27, 2023. [\[FREE Full text\]](#)
2. Ramesh A, Pavlov M, Goh G, Gray S, Voss C, Radford A, et al. Zero-shot text-to-image generation. In: Proc Mach Learn Res. 2021 Presented at: 38th International Conference on Machine Learning; July 18-24, 2021; Virtual p. 8821-8831.
3. Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, et al. Segment anything. arXiv Preprint posted online on April 5, 2023. [\[FREE Full text\]](#)
4. Touvron H, Lavril T, Izacard G, Martinet X, Lachaux MA, Lacroix T, et al. LLaMA: Open and efficient foundation language models. arXiv Preprint posted online on February 27, 2023. [\[FREE Full text\]](#)
5. Thoppilan R, De Freitas D, Hall J, Shazeer N, Kulshreshtha A, Cheng HT, et al. LaMDA: Language models for dialog applications. arXiv Preprint posted online on February 10, 2022. [\[FREE Full text\]](#)
6. Zhai X, Kolesnikov A, Houlsby N, Beyer L. Scaling vision transformers. 2022 Presented at: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); June 18-24, 2022; New Orleans, Louisiana p. 12104-12113. [doi: [10.1109/cvpr52688.2022.01179](#)]
7. OpenAI. GPT-4 is OpenAI's most advanced system, producing safer and more useful responses. OpenAI. URL: <https://openai.com/product/gpt-4> [accessed 2023-03-20]
8. Wang LKP, Paidisetty PS, Cano AM. The next paradigm shift? ChatGPT, artificial intelligence, and medical education. Med Teach. Epub ahead of print 2023 Apr 10. [doi: [10.1080/0142159X.2023.2198663](#)] [Medline: [37036176](#)]
9. Temsah O, Khan SA, Chaiah Y, Senjab A, Alhasan K, Jamal A, et al. Overview of early ChatGPT's presence in medical literature: Insights from a hybrid literature review by ChatGPT and human experts. Cureus 2023 Apr 08;15(4):e37281 [\[FREE Full text\]](#) [doi: [10.7759/cureus.37281](#)] [Medline: [37038381](#)]
10. Subramani M, Jaleel I, Mohan SK. Evaluating the performance of ChatGPT in medical physiology university examination of phase I MBBS. Adv Physiol Educ 2023 Jun 01;47(2):270-271 [\[FREE Full text\]](#) [doi: [10.1152/advan.00036.2023](#)] [Medline: [36971685](#)]
11. Strong E, DiGiammarino A, Weng Y, Basaviah P, Hosamani P, Kumar A, et al. Performance of ChatGPT on free-response, clinical reasoning exams. medRxiv Preprint posted online on March 29, 2023. [\[FREE Full text\]](#) [doi: [10.1101/2023.03.24.23287731](#)] [Medline: [37034742](#)]
12. Sallam M, Salim NA, Al-Tammemi AB, Barakat M, Fayyad D, Hallit S, et al. ChatGPT output regarding compulsory vaccination and COVID-19 vaccine conspiracy: A descriptive study at the outset of a paradigm shift in online search for information. Cureus 2023 Feb 15;15(2):e35029 [\[FREE Full text\]](#) [doi: [10.7759/cureus.35029](#)] [Medline: [36819954](#)]
13. Abdel-Messih MS, Boulos MNK. ChatGPT in clinical toxicology. JMIR Med Educ 2023 Mar 08;9:e46876 [\[FREE Full text\]](#) [doi: [10.2196/46876](#)] [Medline: [36867743](#)]

14. Mbakwe AB, Lourentzou I, Celi LA, Mechanic OJ, Dagan A. ChatGPT passing USMLE shines a spotlight on the flaws of medical education. *PLOS Digit Health* 2023 Feb 09;2(2):e0000205 [FREE Full text] [doi: [10.1371/journal.pdig.0000205](https://doi.org/10.1371/journal.pdig.0000205)] [Medline: [36812618](https://pubmed.ncbi.nlm.nih.gov/36812618/)]
15. Masters K. Ethical use of artificial intelligence in health professions education: AMEE guide No. 158. *Med Teach* 2023 Jun;45(6):574-584. [doi: [10.1080/0142159X.2023.2186203](https://doi.org/10.1080/0142159X.2023.2186203)] [Medline: [36912253](https://pubmed.ncbi.nlm.nih.gov/36912253/)]
16. Masters K. Response to: Aye, AI! ChatGPT passes multiple-choice family medicine exam. *Med Teach* 2023 Jun;45(6):666. [doi: [10.1080/0142159X.2023.2190476](https://doi.org/10.1080/0142159X.2023.2190476)] [Medline: [36940462](https://pubmed.ncbi.nlm.nih.gov/36940462/)]
17. Lee H. The rise of ChatGPT: Exploring its potential in medical education. *Anat Sci Educ*. Epub ahead of print 2023 Mar 14. [doi: [10.1002/ase.2270](https://doi.org/10.1002/ase.2270)] [Medline: [36916887](https://pubmed.ncbi.nlm.nih.gov/36916887/)]
18. 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 Feb 09;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/)]
19. 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/)]
20. Karkra R, Jain R, Shivaswamy RP. Recurrent strokes in a patient with metastatic lung cancer. *Cureus* 2023 Feb 06;15(2):e34699 [FREE Full text] [doi: [10.7759/cureus.34699](https://doi.org/10.7759/cureus.34699)] [Medline: [36909080](https://pubmed.ncbi.nlm.nih.gov/36909080/)]
21. Ide K, Hawke P, Nakayama T. Can ChatGPT be considered an author of a medical article? *J Epidemiol*. Epub ahead of print 2023 Apr 08 [FREE Full text] [doi: [10.2188/jea.JE20230030](https://doi.org/10.2188/jea.JE20230030)] [Medline: [37032109](https://pubmed.ncbi.nlm.nih.gov/37032109/)]
22. Huh S. Are ChatGPT's knowledge and interpretation ability comparable to those of medical students in Korea for taking a parasitology examination?: a descriptive study. *J Educ Eval Health Prof* 2023;20:1 [FREE Full text] [doi: [10.3352/jeehp.2023.20.1](https://doi.org/10.3352/jeehp.2023.20.1)] [Medline: [36627845](https://pubmed.ncbi.nlm.nih.gov/36627845/)]
23. Hallsworth JE, Udaondo Z, Pedrós-Alió C, Höfer J, Benison KC, Lloyd KG, et al. Scientific novelty beyond the experiment. *Microb Biotechnol*. Epub ahead of print 2023 Feb 14 [FREE Full text] [doi: [10.1111/1751-7915.14222](https://doi.org/10.1111/1751-7915.14222)] [Medline: [36786388](https://pubmed.ncbi.nlm.nih.gov/36786388/)]
24. Guo AA, Li J. Harnessing the power of ChatGPT in medical education. *Med Teach*. Epub ahead of print 2023 Apr 10. [doi: [10.1080/0142159X.2023.2198094](https://doi.org/10.1080/0142159X.2023.2198094)] [Medline: [37036161](https://pubmed.ncbi.nlm.nih.gov/37036161/)]
25. Goodman RS, Patrinely JRJ, Osterman T, Wheless L, Johnson DB. On the cusp: Considering the impact of artificial intelligence language models in healthcare. *Med* 2023 Mar 10;4(3):139-140. [doi: [10.1016/j.medj.2023.02.008](https://doi.org/10.1016/j.medj.2023.02.008)] [Medline: [36905924](https://pubmed.ncbi.nlm.nih.gov/36905924/)]
26. 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 Feb 08;9:e45312 [FREE Full text] [doi: [10.2196/45312](https://doi.org/10.2196/45312)] [Medline: [36753318](https://pubmed.ncbi.nlm.nih.gov/36753318/)]
27. Eysenbach G. The role of ChatGPT, generative language models, and artificial intelligence in medical education: A conversation with ChatGPT and a call for papers. *JMIR Med Educ* 2023 Mar 06;9:e46885 [FREE Full text] [doi: [10.2196/46885](https://doi.org/10.2196/46885)] [Medline: [36863937](https://pubmed.ncbi.nlm.nih.gov/36863937/)]
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 Dec;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. Anderson N, Belavy DL, Perle SM, Hendricks S, Hespanhol L, Verhagen E, et al. AI did not write this manuscript, or did it? Can we trick the AI text detector into generated texts? The potential future of ChatGPT and AI in sports & exercise Medicine manuscript generation. *BMJ Open Sport Exerc Med* 2023 Feb 16;9(1):e001568 [FREE Full text] [doi: [10.1136/bmjsem-2023-001568](https://doi.org/10.1136/bmjsem-2023-001568)] [Medline: [36816423](https://pubmed.ncbi.nlm.nih.gov/36816423/)]
30. Alkaissi H, McFarlane SI. Artificial hallucinations in ChatGPT: Implications in scientific writing. *Cureus* 2023 Feb 19;15(2):e35179 [FREE Full text] [doi: [10.7759/cureus.35179](https://doi.org/10.7759/cureus.35179)] [Medline: [36811129](https://pubmed.ncbi.nlm.nih.gov/36811129/)]
31. Davis MH, Harden RM. Planning and implementing an undergraduate medical curriculum: the lessons learned. *Med Teach* 2003 Nov;25(6):596-608. [doi: [10.1080/0142159032000144383](https://doi.org/10.1080/0142159032000144383)] [Medline: [15369907](https://pubmed.ncbi.nlm.nih.gov/15369907/)]
32. Ngo B, Nguyen D, vanSonnenberg E. The cases for and against artificial intelligence in the medical school curriculum. *Radiol Artif Intell* 2022 Aug 17;4(5):e220074 [FREE Full text] [doi: [10.1148/ryai.220074](https://doi.org/10.1148/ryai.220074)] [Medline: [36204540](https://pubmed.ncbi.nlm.nih.gov/36204540/)]
33. Dumić-Čule I, Orešković T, Brkljačić B, Tiljak MK, Orešković S. The importance of introducing artificial intelligence to the medical curriculum - assessing practitioners' perspectives. *Croat Med J* 2020 Oct 31;61(5):457-464 [FREE Full text] [doi: [10.3325/cmj.2020.61.457](https://doi.org/10.3325/cmj.2020.61.457)] [Medline: [33150764](https://pubmed.ncbi.nlm.nih.gov/33150764/)]
34. Çalışkan SA, Demir K, Karaca O. Artificial intelligence in medical education curriculum: An e-Delphi study for competencies. *PLoS One* 2022 Jul 21;17(7):e0271872 [FREE Full text] [doi: [10.1371/journal.pone.0271872](https://doi.org/10.1371/journal.pone.0271872)] [Medline: [35862401](https://pubmed.ncbi.nlm.nih.gov/35862401/)]
35. Maini B, Maini E. Artificial intelligence in medical education. *Indian Pediatr* 2021 May 15;58(5):496-497 [FREE Full text] [Medline: [33980744](https://pubmed.ncbi.nlm.nih.gov/33980744/)]
36. Grunhut J, Marques O, Wyatt ATM. Needs, challenges, and applications of artificial intelligence in medical education curriculum. *JMIR Med Educ* 2022 Jun 07;8(2):e35587 [FREE Full text] [doi: [10.2196/35587](https://doi.org/10.2196/35587)] [Medline: [35671077](https://pubmed.ncbi.nlm.nih.gov/35671077/)]
37. Kitamura FC. ChatGPT is shaping the future of medical writing but still requires human judgment. *Radiology* 2023 Apr;307(2):e230171. [doi: [10.1148/radiol.230171](https://doi.org/10.1148/radiol.230171)] [Medline: [36728749](https://pubmed.ncbi.nlm.nih.gov/36728749/)]

38. Frommeyer TC, Fursmidt RM, Gilbert MM, Bett ES. The desire of medical students to integrate artificial intelligence into medical education: An opinion article. *Front Digit Health* 2022 May 13;4:831123 [FREE Full text] [doi: [10.3389/fdgh.2022.831123](https://doi.org/10.3389/fdgh.2022.831123)] [Medline: [35633734](https://pubmed.ncbi.nlm.nih.gov/35633734/)]
39. Ali SR, Dobbs TD, Hutchings HA, Whitaker IS. Using ChatGPT to write patient clinic letters. *Lancet Digit Health* 2023 Apr;5(4):e179-e181 [FREE Full text] [doi: [10.1016/S2589-7500\(23\)00048-1](https://doi.org/10.1016/S2589-7500(23)00048-1)] [Medline: [36894409](https://pubmed.ncbi.nlm.nih.gov/36894409/)]
40. Biswas S. ChatGPT and the future of medical writing. *Radiology* 2023 Apr;307(2):e223312. [doi: [10.1148/radiol.223312](https://doi.org/10.1148/radiol.223312)] [Medline: [36728748](https://pubmed.ncbi.nlm.nih.gov/36728748/)]
41. Chen TJ. ChatGPT and other artificial intelligence applications speed up scientific writing. *J Chin Med Assoc* 2023 Apr 01;86(4):351-353. [doi: [10.1097/JCMA.0000000000000900](https://doi.org/10.1097/JCMA.0000000000000900)] [Medline: [36791246](https://pubmed.ncbi.nlm.nih.gov/36791246/)]
42. Koo M. The importance of proper use of ChatGPT in medical writing. *Radiology* 2023 May;307(3):e230312. [doi: [10.1148/radiol.230312](https://doi.org/10.1148/radiol.230312)] [Medline: [36880946](https://pubmed.ncbi.nlm.nih.gov/36880946/)]
43. Patel SB, Lam K. ChatGPT: the future of discharge summaries? *Lancet Digit Health* 2023 Mar;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/)]
44. Dahmen J, Kayaalp ME, Ollivier M, Pareek A, Hirschmann MT, Karlsson J, et al. Artificial intelligence bot ChatGPT in medical research: the potential game changer as a double-edged sword. *Knee Surg Sports Traumatol Arthrosc* 2023 Apr;31(4):1187-1189. [doi: [10.1007/s00167-023-07355-6](https://doi.org/10.1007/s00167-023-07355-6)] [Medline: [36809511](https://pubmed.ncbi.nlm.nih.gov/36809511/)]
45. Graf A, Bernardi RE. ChatGPT in research: Balancing ethics, transparency and advancement. *Neuroscience* 2023 Apr 01;515:71-73. [doi: [10.1016/j.neuroscience.2023.02.008](https://doi.org/10.1016/j.neuroscience.2023.02.008)] [Medline: [36813155](https://pubmed.ncbi.nlm.nih.gov/36813155/)]
46. Hill-Yardin EL, Hutchinson MR, Laycock R, Spencer SJ. A Chat(GPT) about the future of scientific publishing. *Brain Behav Immun* 2023 May;110:152-154. [doi: [10.1016/j.bbi.2023.02.022](https://doi.org/10.1016/j.bbi.2023.02.022)] [Medline: [36868432](https://pubmed.ncbi.nlm.nih.gov/36868432/)]
47. Choi EPH, Lee JJ, Ho MH, Kwok JYY, Lok KYW. Chatting or cheating? The impacts of ChatGPT and other artificial intelligence language models on nurse education. *Nurse Educ Today* 2023 Jun;125:105796. [doi: [10.1016/j.nedt.2023.105796](https://doi.org/10.1016/j.nedt.2023.105796)] [Medline: [36934624](https://pubmed.ncbi.nlm.nih.gov/36934624/)]
48. Tlili A, Shehata B, Adarkwah MA, Bozkurt A, Hickey DT, Huang R, et al. What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments* 2023 Feb 22;10(15):1-24 [FREE Full text] [doi: [10.1186/s40561-023-00237-x](https://doi.org/10.1186/s40561-023-00237-x)]
49. Bair H, Norden J. Large language models and their implications on medical education. *Acad Med*. Epub ahead of print 2023 May 10. [doi: [10.1097/ACM.00000000000005265](https://doi.org/10.1097/ACM.00000000000005265)] [Medline: [37162220](https://pubmed.ncbi.nlm.nih.gov/37162220/)]
50. Akhter HM, Cooper JS. Acute pulmonary edema after hyperbaric oxygen treatment: A case report written with ChatGPT assistance. *Cureus* 2023 Feb 07;15(2):e34752 [FREE Full text] [doi: [10.7759/cureus.34752](https://doi.org/10.7759/cureus.34752)] [Medline: [36909067](https://pubmed.ncbi.nlm.nih.gov/36909067/)]
51. Manohar N, Prasad SS. Use of ChatGPT in academic publishing: A rare case of seronegative systemic lupus erythematosus in a patient with HIV infection. *Cureus* 2023 Feb 04;15(2):e34616 [FREE Full text] [doi: [10.7759/cureus.34616](https://doi.org/10.7759/cureus.34616)] [Medline: [36895547](https://pubmed.ncbi.nlm.nih.gov/36895547/)]
52. Trust T. ChatGPT & education. University of Massachusetts Amherst. 2023. URL: <https://docs.google.com/presentation/d/1Vo9w4ftPx-rizdWyaYoB-pQ3DzK1n325OgDgXsnt0X0/edit#slide=id.p> [accessed 2023-05-23]
53. Ahn S. The impending impacts of large language models on medical education. *Korean J Med Educ* 2023 Mar;35(1):103-107 [FREE Full text] [doi: [10.3946/kjme.2023.253](https://doi.org/10.3946/kjme.2023.253)] [Medline: [36858381](https://pubmed.ncbi.nlm.nih.gov/36858381/)]
54. Bolukbasi K, Chang KW, Zou J, Saligrama V, Kalai A. Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. 2016 Presented at: 30th Conference on Neural Information Processing Systems (NIPS 2016); December 5-10, 2016; Barcelona, Spain URL: https://proceedings.neurips.cc/paper_files/paper/2016/file/a486cd07e4ac3d270571622f4f316ec5-Paper.pdf
55. D'Mello S, Lehman B, Pekrun R, Graesser A. Confusion can be beneficial for learning. *Learn Instr* 2014 Feb;29:153-170 [FREE Full text] [doi: [10.1016/j.learninstruc.2012.05.003](https://doi.org/10.1016/j.learninstruc.2012.05.003)]
56. Thirunavukarasu AJ, Hassan R, Mahmood S, Sanghera R, Barzangi K, El Mukashfi M, et al. Trialling a large language model (ChatGPT) in general practice with the applied knowledge test: Observational study demonstrating opportunities and limitations in primary care. *JMIR Med Educ* 2023 Apr 21;9:e46599 [FREE Full text] [doi: [10.2196/46599](https://doi.org/10.2196/46599)] [Medline: [37083633](https://pubmed.ncbi.nlm.nih.gov/37083633/)]
57. Markovski Y. Data usage for consumer services FAQ. OpenAI. URL: <https://help.openai.com/en/articles/7039943-data-usage-for-consumer-services-faq> [accessed 2023-03-19]
58. McCallum S. ChatGPT banned in Italy over privacy concerns. BBC. 2023 Apr 01. URL: <https://www.bbc.com/news/technology-65139406> [accessed 2023-04-03]
59. ChatGPT Generative Pre-trained Transformer, Zhavoronkov A. Rapamycin in the context of Pascal's Wager: generative pre-trained transformer perspective. *Oncoscience* 2022 Dec 21;9:82-84 [FREE Full text] [doi: [10.18632/oncoscience.571](https://doi.org/10.18632/oncoscience.571)] [Medline: [36589923](https://pubmed.ncbi.nlm.nih.gov/36589923/)]
60. O'Connor S. Open artificial intelligence platforms in nursing education: Tools for academic progress or abuse? *Nurse Educ Pract* 2023 Jan;66:103537. [doi: [10.1016/j.nepr.2022.103537](https://doi.org/10.1016/j.nepr.2022.103537)] [Medline: [36549229](https://pubmed.ncbi.nlm.nih.gov/36549229/)]
61. Kung TH, Cheatham M, ChatGPT, Medenilla A, Sillos C, De Leon L, et al. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. medRxiv Preprint posted online on December 21, 2022. [FREE Full text] [doi: [10.1101/2022.12.19.22283643](https://doi.org/10.1101/2022.12.19.22283643)]

62. Rudolph J, Tan S, Tan S. ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning and Teaching* 2023;6(1):1-22 [FREE Full text] [doi: [10.37074/jalt.2023.6.1.9](https://doi.org/10.37074/jalt.2023.6.1.9)]
63. Ahmed A, Ali N, Alzubaidi M, Zaghouni W, Abd-alrazaq AA, Househ M. Freely available Arabic corpora: A scoping review. *Comput Methods Programs Biomed Update* 2022;2:100049 [FREE Full text] [doi: [10.1016/j.cmpbup.2022.100049](https://doi.org/10.1016/j.cmpbup.2022.100049)]
64. Ahmed A, Ali N, Alzubaidi M, Zaghouni W, Abd-alrazaq A, Househ M. Arabic chatbot technologies: A scoping review. *Comput Methods Programs Biomed Update* 2022;2:100057 [FREE Full text] [doi: [10.1016/j.cmpbup.2022.100057](https://doi.org/10.1016/j.cmpbup.2022.100057)]
65. Kuhail MA, Alturki N, Alramlawi S, Alhejori K. Interacting with educational chatbots: A systematic review. *Educ Inf Technol (Dordr)* 2022 Jul 09;28:973-1018 [FREE Full text] [doi: [10.1007/s10639-022-11177-3](https://doi.org/10.1007/s10639-022-11177-3)]

Abbreviations

AI: artificial intelligence
GPT: Generative Pre-trained Transformers
LaMDA: Language Models for Dialog Applications
LLM: large language model
SAM: Segment Anything Model
SOAP: Subjective, Objective, Assessment, and Plan
ViT: Vision Transformer

Edited by K Venkatesh; submitted 19.04.23; peer-reviewed by B Chaves, A Thirunavukarasu, K Masters; comments to author 05.05.23; revised version received 15.05.23; accepted 17.05.23; published 01.06.23

Please cite as:

Abd-alrazaq A, AlSaad R, Alhuwail D, Ahmed A, Healy PM, Latifi S, Aziz S, Damseh R, Alabed Alrazak S, Sheikh J

Large Language Models in Medical Education: Opportunities, Challenges, and Future Directions

JMIR Med Educ 2023;9:e48291

URL: <https://mededu.jmir.org/2023/1/e48291>

doi: [10.2196/48291](https://doi.org/10.2196/48291)

PMID:

©Alaa Abd-alrazaq, Rawan AlSaad, Dari Alhuwail, Arfan Ahmed, Pdraig Mark Healy, Syed Latifi, Sarah Aziz, Rafat Damseh, Sadam Alabed Alrazak, Javaid Sheikh. Originally published in *JMIR Medical Education* (<https://mededu.jmir.org>), 01.06.2023. 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 Medical Education*, is properly cited. The complete bibliographic information, a link to the original publication on <https://mededu.jmir.org/>, as well as this copyright and license information must be included.