THE EUROPEAN SOCIETY OF MEDICINE
Medical Research Archives, Volume 14 Issue 2

RESEARCH ARTICLE

Evaluating the Use of Large Language Models vs. Traditional Textbooks in Physiology Education: Insights into Generation Z Medical Students' Preferences

Dr. L Sundareswaran Ph.D [1]; Dr. Abhishek Sinha MD [2]; Dr. Krishnan S MD, Ph.D [1]; Dr. Manasi Bhattacharjee MD [3]; Dr. Naveen P, MD [4]; Dr. Anindita Mahanta MD [1]

1 Associate professor, Department of Physiology, AIIMS, Guwahati, Assam, India

2 Additional Professor, Department of Physiology, AIIMS, Guwahati, Assam, India

3 Professor & Head, Department of Physiology, AIIMS, Guwahati, Assam, India

4 Professor, Department of Physiology, AIIMS, Guwahati, Assam, India

OPEN ACCESS

PUBLISHED
28 February 2026

CITATION
Sundareswaran, L., Sinha, A., et al., 2026. Evaluating the Use of Large Language Models vs. Traditional Textbooks in Physiology Education: Insights into Generation Z Medical Students Preferences. Medical Research Archives, [online] 14(2).

ISSN
2375-1924

ABSTRACT

Background: Medical education is rapidly evolving with the integration of artificial intelligence [AI] and digital technology. Large language models [LLMs] such as ChatGPT, Bing, Bard, Gemini, Claude, and Perplexity are increasingly used by medical students primarily for their accessibility and ease of use. However, the abundance of information generated by these tools may be overwhelming and can complicate the independent assimilation of knowledge.

Objectives: This experimental study aimed to evaluate the effectiveness and acceptability of LLM based AI tools compared with traditional textbooks for learning Physiology among Generation Z medical students.

Methods: First-year medical students were randomized in to 2 groups: 1- standard Physiology textbooks and 2- Large language models based AI tools. After one hour of self directed study, students completed a 15 item multiple choice questions [topic specific] via Google form. The Medical Artificial Intelligence Readiness Scale for Medical Students was used to evaluate AI readiness and perceptions towards AI-assisted learning.

Results: Analysis revealed no statistically significant difference in assessment scores between students who used traditional textbooks and those who used large language model based artificial intelligence tools across the selected Physiology topics. Qualitative feedback indicated that many students favored combining both resources rather than relying exclusively on a single method.

Conclusion: Although large language models are gaining prominence in medical education, this study did not demonstrate a performance advantage over traditional textbooks in learning Physiology among Generation Z medical students highlighting a shift in learning preferences among Generation Z medical students. Future research is warranted to explore the role of large language models based artificial intelligence tools roles in medical education, with particular attention to global content validation, learning retention and ethical consideration.

Keywords: AI readiness, Medical Education, Generation Z, learning preferences.

Introduction

Medical education is already undergoing a generational transformation, particularly with increasing integration of artificial intelligence and digital technologies. Among these innovations, large language models, such as Chat GPT, Bing, Bard, Gemini, Claude, Perplexity, etc, have emerged as supplementary educational tools, functioning as virtual tutors for medical students education as tutors [1]. These tools may serve as adjuncts to conventional learning methods, standard curriculum, and educational practices [2,3]. Undergraduate medical education represents a critical phase in the development of future healthcare professionals. Although teacher-led instructions remain a cornerstone of medical curriculum, student centered led learning approached are gaining prominence due to their flexibility and higher engagement [4]. Medical students are increasingly inclined to use AI based tools to acquire knowledge and explore the educational contents rather than relying solely on traditional textbooks. This preference may be attributed to widespread availability, accessibility, rapid generation of responses, ready-to-use answers, related explanations, and existing resources on the internet with less effort offered by LLMs. AI driven tools & algorithms can help students save time, provide access to extensive information in the online world, and enable independent learning without the direct involvement of instructors [5]. However, these platforms often provides abundant data of information from the available resources, which can be overwhelming and challenging for students to critically evaluate, synthesize and retain. Consequently, the overall impact of AI assisted tools on student comprehension, learning outcomes and long term attention remains insufficiently understood and the entire learning and retention process using these tools is less known [6,7]. Artificial intelligence tools have established their important role across multiple academic disciplines [8] The application of artificial intelligence in medical education and training offers numerous potential, particularly in facilitation personalized learning experiences [9]. Research indicates that when used thoughtfully and appropriate prompting, AI tools can support students in studying and preparing for examinations. Additionally, AI can provide personalized academic support while adhering to educational standards [10]. AI is also extensively utilized in many clinical research and Physiology systems, such as an ECG analysis and evaluation from a large number of data sets to anticipate and predict disease conditions [11].

The scientific research studies directly comparing the freely available large language models with reference standard textbooks in Physiology learning remain limited, particularly in context of undergraduate Physiology education. Although LLMs shows promise in reinforcing clinical knowledge reinforcement, their effectiveness may be enhanced by rethinking AI development to ensure easier, safer, and more equitable to use. Nevertheless, medicine remain a particularly intricate domain for the application of LLMs [12]. Several studies have demonstrated that chat GPT can successfully pass medical examinations; a recent study by Subramani M et al., [13] reported that ChatGPT scored more than 75% marks in an University examination that consists of MCQs from Physiology. While AI has the potential to reshape and improve medical systems, there remains a risk of AI generated content and automation bias if not adequately validated by clinicians and academicians.

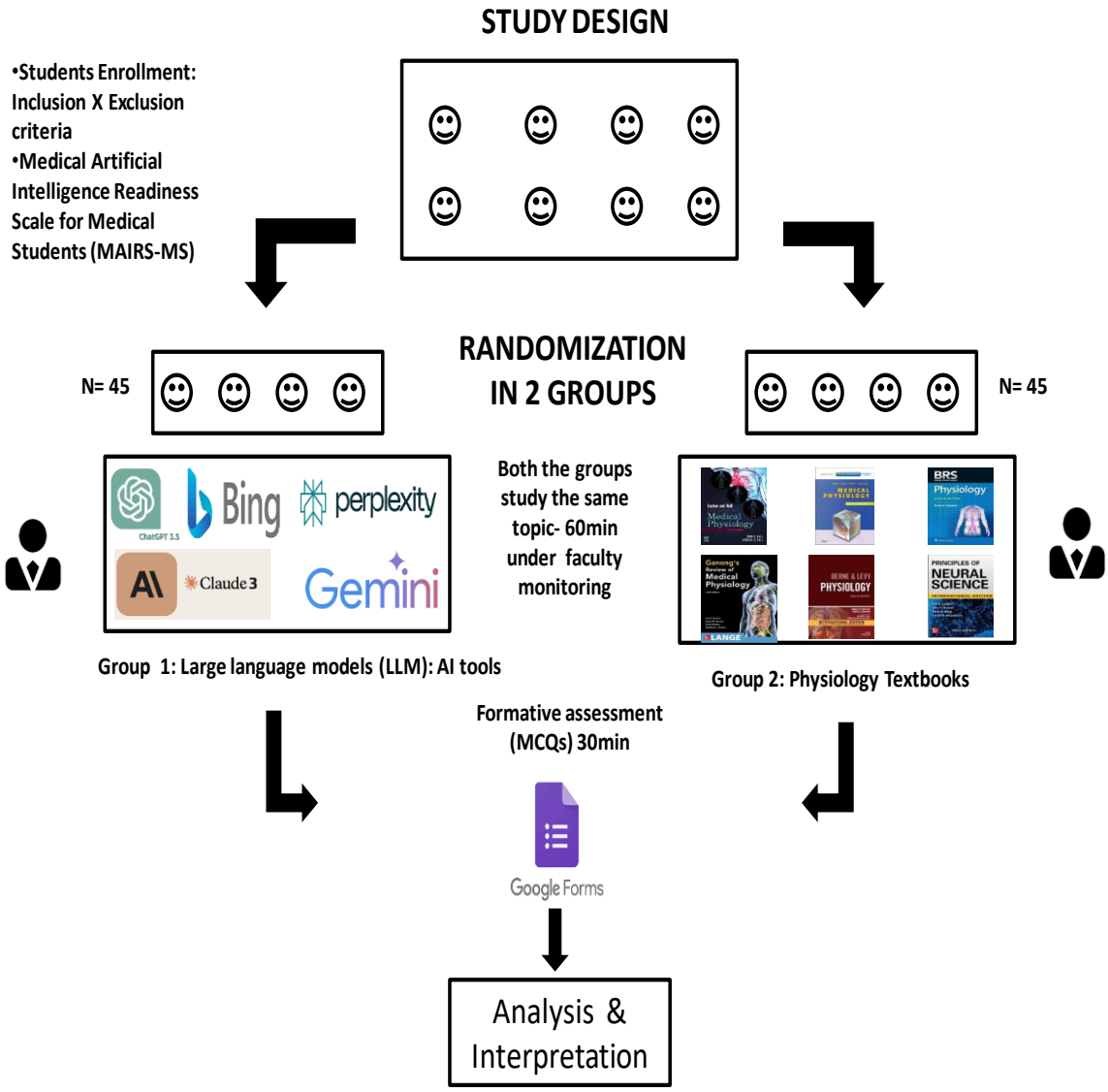
Traditional textbook based learning has been the cornerstone of medical education for decades. Textbooks offer a comprehensive and structured approach to knowledge acquisition and promote deep learning, they often lack interactivity and real-time updates. Medical students continue to rely on textbooks for foundational knowledge, but may perceive them monotonous and static compared to dynamic and interactive nature of LLM based AI tools. Studies have shown that although medical students possess relatively strong technical skills and their practical usability related to AI are significantly lower [14]. Despite the availability of educational AI resources, most users report having little or no formal AI training and rely primarily on self-directed learning. Incorporating LLM based AI tools awareness programs and certificate courses, including training on Large language models, in to medical education may help to bridge this gap [15]. Recent advancements in AI driven educational technologies and tools encompassing various technological innovations to improve Physiology education have been utilized in medical education for current-generation students. However, the adoption of LLMs requires valid and essential background scientific evidence to recognize and address the concerns related to accuracy, bias, ethical implications and educations effectiveness of AI based LLMs [16]. To fully realize the potential of LLM based AI tools, it is necessary to ensure and understand the students and educators perceptions of their effectiveness. In addition, issues related to accessibility, effectiveness usability, and cultural acceptance must be examined to support widespread recommendation and adoption. Therefore, this study was designed to evaluate and understand the medical students and doctors perceptions and attitudes towards integration of LLM based AI tools in medicine is crucial and need of an hour [17]. With this background, the present study aimed to explore the effectiveness of large language models in Physiology education compared with traditional textbook learning among 1st MBBS students.

Materials and methods:

The outline of the study design is given in the study design flowchart [Figure:1]. This study was conducted in the Department of Physiology after obtaining institutional ethical approval. [M5/F142/2024] Medical students were enrolled in the study after detailed informed consent and scrutiny based on the inclusion and exclusion criteria. Students who joined 1st MBBS - Bachelor of Medicine, Bachelor of Surgery after their higher secondary schools were included in the study. Students who already had a graduate degree in Physiology or completed any course containing Physiology before joining MBBS were excluded. Students who did not participate in the questionnaire were also excluded. Upon enrollment, students meeting

the inclusion criteria were allocated to two groups through randomization using their roll numbers. The sample size was calculated using G power computer software. A total sample size of 45 in each group was calculated to detect a significant effect [d=0.45]. The power of the test was fixed at 90% and alpha at 0.05 [1].

Figure 1: Flow chart- Study Design:



Group 1- This group of students used standard textbooks to read Physiology [Authors by: Guyton, Ganong, Boron, Costanzo, Berne & Levy, Principles of neural Science by Eric Kandel], Group 2- This group of students used LLM [ChatGPT, Google Gemini, Microsoft Bing, Claude AI, Perplexity] to study different topics of Physiology. Two main aspects, cognitive and affective, were assessed in this study. The cognitive part included studying the topic, followed by the MCQ test. The practical part included an assessment questionnaire and interview questions in google forms regarding the study and its methodology Table 3 & 4. Important topics related to each physiological system were chosen for the study. The topic was disclosed to the participants only on the day of the study. Moreover, only topics that were new or not already covered during routine lectures were selected for this study. The study was conducted in a practical lab with seating arrangements. The Textbook group and the LLM group were placed in separate rooms, and both were supervised by faculty and staff. The students were given 60 minutes to read about the topic, and immediate recall was assessed using 15 MCQs using google forms, and the time given was 30 minutes. During the entire session, the departmental faculty and residents monitored the students. The experts in Physiology internally and externally validated the MCQs used for the assessment. The marks obtained in the MCQs were announced to the students, and a review session was conducted before the next topic. The marks were not considered for their internal assessment.

The medical artificial intelligence readiness scale questionnaire for medical students [MAIRS-MS] was administered to evaluate and monitor perceived readiness levels among medical students on AI technologies and applications. The MAIRS-MS consists of four main aspects- cognition [1-8 questions], vision [17-19 questions], ability [9 -16 questions], and ethics [20-22 questions] [18]. Following the MCQ test, the student participants were asked five questions, with responses rated on a 5-point Likert-type scale: 1—strongly disagree, 2—disagree, 3—neutral, 4—agree, and 5—strongly agree.

QUESTIONS:

1. How do you rate the effectiveness of the above study on your learning outcome?
2. How do you rate textbooks as a source of study for acquiring knowledge?
3. How do you rate AI tools as a source of study for acquiring knowledge?
4. Do you think this AI system supports learner–instructor interaction?
5. Which of the two learning tools, in your opinion, is likely to enhance your critical thinking skills? AI or Textbooks?

DATA ANALYSIS:

The responses were exported to Excel from the Google form and questionnaires, and a systematic analysis using SPSS was done. Statistical significance was determined with a threshold [alpha] level set at 0.05. The quantitative data obtained from surveys and questionnaires were analyzed using statistical tools such as descriptive statistics and independent *t*-tests. The mean average scores obtained from all the variables were calculated and compared statistically.

Results:

A total of 90 first-year medical students participated in the study and were equally allocated to the large language model group and the textbook group. Comparison of post-learning assessment scores revealed no statistically significant difference between the two groups across the evaluated Physiology topics, including respiratory, cardiovascular, gastrointestinal, renal, special senses, and general Physiology systems ($p > 0.05$). The corresponding results are presented in Tables 1 to 4. Student feedback regarding the use of large language model–based artificial intelligence tools and textbooks is illustrated in Figures 2 and 3.

Figure 2: Students feedback about LLM: AI tools [AI] and Test book [T] usage.

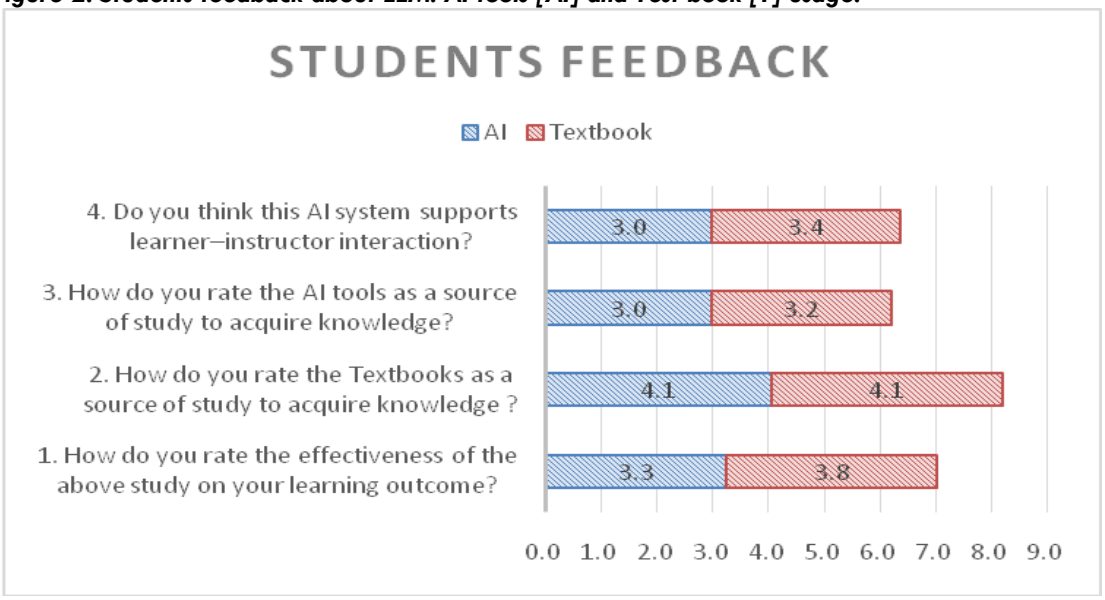
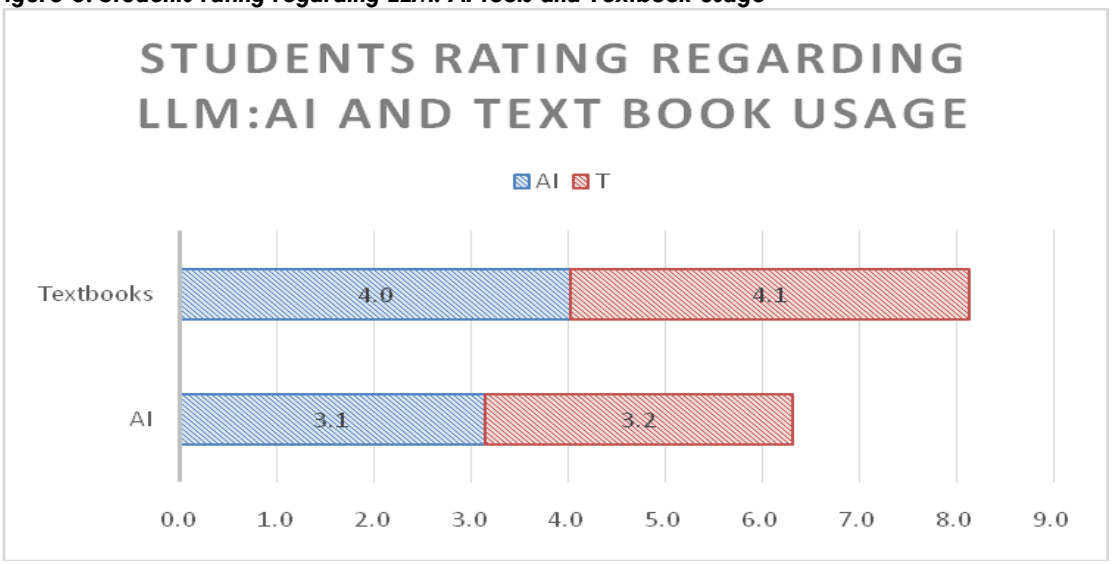


Figure 3: Students rating regarding LLM: AI tools and Textbook usage



Figures 2 and 3 represent the feedback and ratings regarding using LLM tools [AI] versus traditional Textbook [T] methods for learning Physiology compilation among Generation Z students.

Table 1: Comparison of scores between the AI group and text group on different topics

S. No	Topic/System	Group	Score [Mean ± SD]	p-value
1	Respiratory System	LLM: AI tools	3.1±1.3	0.06
		Textbook	3.1±1.2	
2	Special senses	LLM: AI tools	2.01±2.52	0.58
		Textbook	1.53±1.99	
3	Cardiovascular System	LLM: AI tools	2.21±2.74	0.86
		Textbook	1.34±1.82	
4	General Physiology	LLM: AI tools	1.42±1.67	0.29
		Textbook	1.65±2	
5	Gastrointestinal System	LLM: AI tools	1.66±2.04	0.10
		Textbook	1.33±1.92	
6	Renal System	LLM: AI tools	1.98±2.38	0.12
		Textbook	1.83±2.2	

Table 1. Comparison of mean ± SD scores between the large language model group and the textbook group across selected Physiology topics. Data were analyzed using an independent t-test, with $p < 0.05$ considered statistically significant. No statistically significant differences were observed between groups.

Table 2: Comparison of overall mean scores between the AI group and Textbook group

S. No	Group	Score [Mean ± SD]	p-value
1	LLM: AI tools	6.3±2.9	0.81
	Textbook	6.7±2.6	

Table 2 provides the overall mean scores between the LLM tools and the textbook group, which showed no significant distinctions.

Table 3: Distribution of responses to the Medical Artificial Intelligence Readiness Scale (MAIRS) across four domains (ability, cognition, vision, and ethics)

Tables 3 and 4. Medical Artificial Intelligence Readiness Scale for Medical Students scores across four domains. MAIRS is a validated instrument with a Cronbach's alpha of 0.87 used to assess readiness for artificial intelligence among medical students. Most responses fell within the neutral and agree categories, with higher scores observed for ability and cognition compared with vision and ethics.

S. No	Parameters	Strongly Disagree n [%]	Disagree n [%]	Neutral n [%]	Agree n [%]	Strongly Agree n [%]
1. 1	I can define the basic concepts of data science	4[4.7]	15 [17.4]	35[40.7]	30.2[26]	8 [7]
2. 2	I can define the basic concepts of statistics	5 [5.8]	11 [12.8]	34 [39.5]	28 [32.6]	8 [9.3]
3. 3	I can explain how AI systems are trained	11 [12.8]	31 [36]	25 [29.1]	15 [17.4]	4 [4.7]
4. 4	I can define the basic concepts and terminology of AI	5 [5.8]	16 [18.6]	44 [51.2]	19 [22.1]	2 [2.3]
5. 5	I can properly analyze the data obtained by AI in healthcare.	3 [3.5]	18 [20.9]	33 [38.4]	28 [32.6]	4 [4.7]
6. 6	I can differentiate the functions and features of AI-related tools and applications.	4 [4.7]	15 [17.4]	39 [45.3]	23 [26.7]	5 [5.8]
7. 7	I can organize workflows compatible with AI.	3 [3.5]	20 [23.3]	28 [32.6]	29 [33.7]	6 [7]
8. 8	I can express the importance of data collection, analysis, evaluation, and safety in developing AI in healthcare.	3 [3.5]	7 [8.1]	35 [40.7]	28 [32.6]	13 [15.1]
9. 9	I can harness AI-based information combined with my professional knowledge.	2 [2.3]	6 [7]	33 [38.4]	36 [41.9]	9 [10.5]
0. 10	I can use AI technologies effectively and efficiently in healthcare delivery.	1 [1.2]	7 [8.1]	38 [44.2]	28 [32.6]	12 [14]

S. No	Parameters	Strongly Disagree n [%]	Disagree n [%]	Neutral n [%]	Agree n [%]	Strongly Agree n [%]
1. 11	I can use artificial intelligence applications for their purpose.	1 [1.2]	2 [2.3]	34 [39.5]	40 [46.5]	9 [10.5]
2. 12	I can access, evaluate, use, share and create new knowledge using information and communication technologies.	2 [2.3]	8 [9.3]	33 [38.4]	35 [40.7]	8 [9.3]
3. 13	I can explain how AI applications offer a solution to which problem in healthcare.	2 [1.2]	12 [14]	37 [43]	29 [33.7]	7 [8.1]
4. 14	It is valuable to use AI for education, service and research purposes.	1 [1.2]	2 [2.3]	26 [30.2]	43 [50]	14 [16.3]
5. 15	I can explain the AI applications used in healthcare services to the patient.	1 [1.2]	8 [9.3]	35 [40.7]	33 [38.4]	9 [10.5]
6. 16	I can choose proper AI applications for the problems encountered in healthcare.	4 [4.7]	9 [10.5]	41 [47.7]	25 [29.1]	7 [8.1]
7. 17	I can explain the limitations of AI technology.	1 [1.2]	7 [8.1]	34 [39.5]	34 [39.5]	10 [11.6]
8. 18	I can explain the strengths and weaknesses of AI technology.	2 [2.3]	5 [5.8]	38 [44.2]	32 [37.2]	9 [10.5]
9. 19	I can foresee the opportunities and threats that AI technology can create.	2 [2.3]	7 [8.1]	34 [39.5]	32 [37.2]	11 [12.8]
20. 20	I can use health data by legal and ethical norms.	2 [2.3]	4 [4.7]	44 [51.2]	29 [33.7]	7 [8.1]
21. 21	I can conduct my work under ethical principles while using AI technologies.	2 [2.3]	7 [8.1]	37 [43]	36 [41.9]	4 [4.7]
22. 22	I can follow legal regulations regarding the use of AI technologies in healthcare.	2 [2.3]	6 [7]	38 [44.2]	27 [31.4]	13 [15.1]

Table 4 represents the overall mean score and the proportion of the study population who scored more than 50 %.

S. No	Parameter	Score mean±SD	n [%]
1.	Cognition score	24.9±5.5	73 [85]
2.	Ability factor	27.9±5.1	84 [98]
3.	Vision factor	10.5±2.2	80 [93]
4.	Ethics factor	10.2±2.1	80 [93]
5.	MAIRS points	73.6±12.9	83 [97]

Following are the few comments from the students after participation in this study:

1. To enhance our understanding, we must combine the tools and textbooks.
2. Textbook is better
3. AI is a great way to facilitate the process of learning and memorizing,
4. Both textbooks and AI should be used together to gain a better understanding of the subject
5. A teacher's explanation is always better than reading books or AI.
6. AI can't be a primary source for acquiring knowledge
7. Integrated use of AI and books is a good way to learn.
8. Combining AI and textbooks is a better option than AI or textbooks alone.
9. Some more time should have been provided for Studying.
10. Textbooks give us limited and essential information that is vital
11. Textbooks have more authentic and relevant content than AI
12. AI cannot be the primary source of knowledge acquisition. However, it can be a supplementary resource along with the Textbook...
13. Both are important for gaining the required knowledge and information. Anyone should replace none.
14. Textbooks have complicated language and are very difficult to understand, but AI provides the easiest way to learn
15. AI tends to mislead information and not necessarily understand the concept, but standard books generally lack significant errors.
16. Both are effective for students.
17. AI tools are not for beginners, but the Textbook provides information from the basics.
18. Books are better for acquiring knowledge and cannot depend on AI
19. I felt There was a shortage of resources and study materials while using the AI mode of studying
20. Ai is better
21. AI can complement the traditional learning method. However, AI cannot be the sole means of learning, especially if the student is involved in a professional

course. Indeed, under the proper guidance of the instructors, AI will benefit the learning process.
22. Textbooks are effective but require more time to cover such vast topics
23. I do not know how to use AI, so it may not be fruitful to use AI for me
24. Textbooks are better than AI tools in matters of study purpose
25. Textbooks are better than AI

Discussion:

This study explored the effectiveness of conventional learning. Using textbook versus large language models: AI tools among generation Z first year medical students. Previous data on this topic is limited, especially the original articles involving medical students are scarce. Most of the studies have explored the role of large language models, mainly Chatgpt. These tools were mainly used to generate and solve the MCQs-based exams/tests, and the performance was compared with actual students [19–22]. A recent study states that ChatGPT could potentially revolutionize medical education settings by assisting medical students and faculty, offering innovation within the medical sciences and education framework. This interesting opinion could be due to the intriguing nature of ChatGPT or any large language models that create rapid responses and conversations, such as interactions within seconds. In another study, it was reported that Claude-2 showed better proficiency than ChatGPT-3.5 in conceptual MCQ explanation [23,24]. The novelty of the present study is its unique study design, which did not assess the language model's utility to prepare MCQs or question papers; instead, this study was used to evaluate the efficacy of LLM: AI tools in transferring knowledge to medical students. Two groups were assigned to study Physiology topics, one of the significant subjects in 1st year MBBS, using either textbooks or LLMs and their performance was assessed. The findings shed light on several critical aspects of this comparative educational approach in Physiology learning and the concerns regarding using LLM as an educational tool [25]. ·The study also assessed the students attitudes towards these tools.

Despite the predominant learning methodology among Generation Z students centered around technology and videos [26]. Gen Z participants being optimistic about LLM: AI's benefits for enhanced productivity and personalized learning, the results show no significant difference among the scores between LLM and textbook group. This indicates that there was no added benefit while learning using either tool. This is even though language models provide topic-specific searching options, explanations and suggestions, allowing students to acquire the required information in less time and effort [27] compared to conventional textbooks. It is also worth noting that the feedback and ratings compiled from Generation Z students regarding using LLM tools versus traditional textbook methods for learning Physiology credited both approaches equally [Table 1 & 2].

A recent study says that ChatGPT and Bard are extremely useful because they can imitate human conversation, engage students in interactive discussions, and offer helpful information [28,29]. The other possible reason for the students who preferred LLM could be the complex languages and new words used in medical textbooks that demand a high level of comprehensive understanding and critical thinking [30].

It has been found that the perceived effectiveness of tools employed by textbooks remains high among Generation Z students in medical education [31]. This could be the reason for the finding in most of the student's feedback stating that textbook content is validated, reliable and better for beginners when compared with LLM: AI tools. This perception may vary based on demographic factors and settings, with graphics often proving particularly efficacious. In medical education, the preference for traditional Textbook learning over newer technologies like large language models such as Chat GPT is often influenced by established academic norms, familiarity with traditional resources, and the perceived reliability of textbook information [32]. Various factors have been postulated to potentially impact medical students textbook preferences, including text clarity, language complexity, directness to the intended concept, flow of ideas, clinical/multidisciplinary integration, and effectiveness of aiding tools utilized within textbooks [30,33].

Regarding preference for various learning language models, Chat GPT received the highest rating, followed by laude two and perplexity and Gemini/Bard. Any students did not use Microsoft Bing. Chat GPT has also been reported to have a higher score in solving hematology cases than Google Bard and Microsoft Bing [34]. The LLMs, namely, Chat GPT, Google Bard, Microsoft Bing, and Perplexity, all showed >75% concordance in suggested statistical tests for research case vignettes, with all having acceptance of >95%. Another study regarding using AI language models for medical decision-making on complex scenarios found that Claude AI performs better than ChatGPT, Google Bard, and Perplexity AI [35]. This suggests a wide variation in responses regarding preference and efficacy among different AI language models with no consistent finding. Interestingly, based on the feedback comments following the current study, most students reported that combining LLM: AI and textbook approach can enhance learning and understanding of any topic more effectively.

Gen X students and Gen Y teachers acknowledge these benefits but, at the same time, voice concerns about overreliance, ethical issues, and the need for guidelines to ensure the responsible use of AI in education [36]. Excessive LLM dependency on answers and explanations may put the student's capacity for exploration and analysis at risk, which may reduce their innate nature [25]. Additionally, medical students utilize such language tools to generate answers, and for the assignment, seminars may lead to misconduct [37] and require immediate attention. Further, students use it for written assignments and examinations, leading to concerns about AI-assisted cheating [38]. Moreover, such electronic resources like LLM: AI tools may attenuate the student's exposure to critical thinking, which is crucial for a deep understanding of concepts in medicine.

In terms of AI acceptance and readiness among medical students, overall readiness scored above 50% across all domains of the MAIRS assessment. The highest mean scores were recorded in the areas of ability [27.9±5.1] and cognition [24.9±5.5], compared to lower scores in vision [10.5±2.2] and ethics [10.2±2.1]. These scores align with the majority of students selecting "agree" or "neutral" in their MAIRS responses [Table 3 &4]. The ability domain evaluates a student's preparedness to choose, implement, and articulate the use of AI applications within the medical field. Vision assesses their capacity to analyze and communicate AI technologies limitations, advantages, and prospects. Ethics determines their commitment to adhering to legal and ethical guidelines in the application of AI [39]. AI readiness and acceptance showed varied effects across domains among medical students from various countries, according to the existing literature on MAIRS scores. However, the MAIRS scale's cut-off point has not yet been determined, and further research in diverse countries is needed for more effective cross-cultural comparisons [40].

Large language models could have yet another application in the sense that these can be incorporated with machine learning tools that sync and automatically perceive or extract the student's stress data from wearables [41]. This can assist in the prediction of anxiety and depression during exams [42]. Language models can also be tuned to analyze individual student performance and generate personalized learning materials that address their specific areas of weakness [43]. Strict regulations, guidelines, and supervision must be formulated to prevent misinformation on LLM: AI tools in medical education and ensure optimum use that will benefit students, teachers, and society. [2]

Limitations

This study involves data exclusively from first-year students at a single central institution, with efforts to ensure uniformity in student motivation and involvement across all groups. It is important to note that textbooks may not always provide the most up-to-date information on specific topics. The time allotted for learning was the same for students with different levels of prior knowledge. Future research should involve collaboration with multiple medical institutions and compare study methods across developed and developing countries to fully understand the role of large language models [LLMs] in medical education. The study did not validate the content or address data privacy concerns related to using LLMs for Physiology, which included information from textbooks, articles, news, blogs, and other verified online sources. It is also crucial to recognize that while LLMs often provide highly accurate responses, they are unreliable. Recent advancements in LLMs and generative AI tools may produce different outputs for the same prompt, and information from ChatGPT may not be fully up-to-date beyond 2021, with several associated limitations [17].

Conclusion:

The study compared the effectiveness of traditional textbook learning vs large language models [LLMs: AI tools] in teaching Physiology to first-year Generation Z medical students. Most existing research focuses on LLM tools for generating multiple-choice questions [MCQs] and exam preparation rather than their broader use in medical education. Among various LLMs, ChatGPT shows the potential to revolutionize medical education with its fast, interactive responses, surpassing other models like Claude, Perplexity, Gemini, and Bing. However, the study found no significant preference for LLM over textbooks among students studying and learning physiology. Most students favoured a blended approach, combining both LLM and textbooks. Future research must focus on collaborative studies, both nationally and globally, to thoroughly explore the role of LLM tools in medical education. These studies should ensure robust content validation and address the ethical and educational challenges of LLM usage among medical students. Continuous research and validation of LLM tools will be essential for understanding their long-term impact and acceptance among medical students, faculty, and clinicians. It will provide valuable insights for effective resource allocation in education shortly.

Acknowledgments: The authors sincerely acknowledge the administrative support extended by the departmental staff and the All India Institute of Medical Sciences (AIIMS), Guwahati, Assam, India.

Conflicts of interest: Nil

Data availability: The data that support the findings of this study are not openly available due to reasons of sensitivity and are available from the corresponding author upon reasonable request.

References:

1. Mahat RK, Jantikar AM, Rathore V, Panda S. Assessing the performance of ChatGPT to solve biochemistry question papers of university examination. *Adv Physiol Educ*. 2023;47(3):528-529. doi:10.1152/advan.00076.2023
2. Zarei M, Mamaghani HE, Abbasi A, Hosseini MS. Application of artificial intelligence in medical education: A review of benefits, challenges, and solutions. *Med Clínica Práctica*. 2024;7(2):100422.
3. Biri SK, Kumar S, Panigrahi M, Mondal S, Behera JK, Mondal H. Assessing the Utilization of Large Language Models in Medical Education: Insights From Undergraduate Medical Students. *Cureus*. 15(10):e47468. doi:10.7759/cureus.47468
4. Jagzape A, Jagzape TB. Teacher-led versus student-led seminar blended with portfolio for "assessment of learning": An interventional study. *J Educ Health Promot*. 2022;11:339. doi:10.4103/jehp.jehp_165_22
5. Chu ETT, Chu ECP. The potential of chatbot in chiropractic practice. *J Contemp Chiropr*. 2023;6(1):72-78.
6. Seo K, Tang J, Roll I, Fels S, Yoon D. The impact of artificial intelligence on learner–instructor interaction in online learning. *Int J Educ Technol High Educ*. 2021;18(1):54. doi:10.1186/s41239-021-00292-9
7. Guilherme A. AI and education: the importance of teacher and student relations. *AI Soc*. 2019;34(1):47-54. doi:10.1007/s00146-017-0693-8
8. Gusev AV, Vladzymyrskyy AV, Sharova DE, Arzamasov KM, Khramov AE. Evolution of research and development in the field of artificial intelligence technologies for healthcare in the Russian Federation: results of 2021. *Digit Diagn*. 2022;3(3):178-194. doi:10.17816/DD107367
9. Murtaza M, Ahmed Y, Shamsi J, Sherwani F, Usman M. AI-Based Personalized E-Learning Systems: Issues, Challenges, and Solutions. *IEEE Access*. 2022;10:1-1. doi:10.1109/ACCESS.2022.3193938
10. Favero TG. Using artificial intelligence platforms to support student learning in physiology. *Adv Physiol Educ*. 2024;48(2):193-199. doi:10.1152/advan.00213.2023
11. Shutov DV, Валериевич ШД, Shutov DV, et al. Artificial intelligence in clinical physiology: How to improve learning agility. *Digit Diagn*. 2023;4(1):1. doi:10.17816/DD123559
12. Singh A, Dangat DP, Shenoy S, Sandhu JS. Body Composition Analysis using Karada Scan in Indian Collegiate Population: An Observational Study. *Med J Dr Patil Univ*. 2023;16(4):551. doi:10.4103/mjdrdypu.mjdrdypu_248_21
13. Subramani M, Jaleel I, Krishna Mohan S. Evaluating the performance of ChatGPT in medical physiology university examination of phase I MBBS. *Adv Physiol Educ*. 2023;47(2):270-271. doi:10.1152/advan.00036.2023
14. Laupichler MC, Aster A, Meyerheim M, Raupach T, Mergen M. Medical students' AI literacy and attitudes towards AI: a cross-sectional two-center study using pre-validated assessment instruments. *BMC Med Educ*. 2024;24(1):401. doi:10.1186/s12909-024-05400-7
15. Basu K, Sinha R, Ong A, Basu T. Artificial intelligence: How is it changing medical sciences and its future? *Indian J Dermatol*. 2020;65(5):365-370.
16. Abd-alrazaq A, AlSaad R, Alhuwail D, et al. Large Language Models in Medical Education: Opportunities, Challenges, and Future Directions. *JMIR Med Educ*. 2023;9:e48291.` ` doi:10.2196/48291
17. Kansal R, Bawa A, Trehan S, Goyal K, Goyal N, Malhotra K. Differences in Knowledge and Perspectives on the Usage of Artificial Intelligence Among Doctors and Medical Students of a Developing Country: A Cross-Sectional Study. *Cureus*. 2022;14. doi:10.7759/cureus.21434
18. EmİR B, Yurdem T, Ozel T, et al. Artificial Intelligence Readiness Status of Medical Faculty Students. *Konuralp Tıp Derg*. 2024;16(1):88-95. doi:10.18521/ktd.1387826
19. Bharatha A, Ojeh N, Rabbi AMF, et al. Comparing the Performance of ChatGPT-4 and Medical Students on MCQs at Varied Levels of Bloom’s Taxonomy. *Adv Med Educ Pract*. 2024;15:393-400. doi:10.2147/AMEP.S457408
20. Brin D, Sorin V, Vaid A, et al. Comparing ChatGPT and GPT-4 performance in USMLE soft skill assessments. *Sci Rep*. 2023;13(1):16492.
21. Friederichs H, Friederichs WJ, März M. ChatGPT in medical school: how successful is AI in progress testing? *Med Educ Online*. 2023;28(1):2220920. doi:10.1080/10872981.2023.2220920
22. Banerjee A, Ahmad A, Bhalla P, Goyal K. Assessing the Efficacy of ChatGPT in Solving Questions Based on the Core Concepts in Physiology. *Cureus*. Published online August 10, 2023. doi:10.7759/cureus.43314
23. Roumeliotis KI, Tselikas ND. Chatgpt and open-ai models: A preliminary review. *Future Internet*. 2023;15(6):192.
24. Agarwal M, Goswami A, Sharma P, Agarwal M, Goswami A, Sharma P. Evaluating ChatGPT-3.5 and Claude-2 in Answering and Explaining Conceptual Medical Physiology Multiple-Choice Questions. *Cureus*. 2023;15. doi:10.7759/cureus.46222
25. Hasanein AM, Sobaih AEE. Drivers and Consequences of ChatGPT Use in Higher Education: Key Stakeholder Perspectives. *Eur J Investig Health Psychol Educ*. 2023;13(11):2599-2614. doi:10.3390/ejihpe13110181
26. Granitz N, Kohli C, Lancellotti MP. Textbooks for the YouTube generation? A case study on the shift from text to video. *J Educ Bus*. 2021;96(5):299-307. doi:10.1080/08832323.2020.1828791
27. Gaitsgory O, Burgess A, Mellis C. Opinion piece:'medical students-learning from textbooks or electronic media?'. *J Paediatr Child Health*. 2013;49(9). Accessed June 15, 2024. https://search.ebscohost.com/login.aspx?direct=true&profile=ehost&scope=site&authtype=crawler&jrnl=10344810&asa=N&AN=90096518&h=W8Aq29iAqmCJFT7J7YBDp5ew6uT%2B00SENdcjO6u6Tm5

QaBLI%2BDi8y%2F2PjkgyWMJ8alHcLpATYxyrGwHcltd5jw%3D%3D&crl=c
28. Ahmed Y. Utilization of ChatGPT in Medical Education: Applications and Implications for Curriculum Enhancement. *Acta Inform Medica*. 2023;31(4):300-305. doi:10.5455/aim.2023.31.300-305
29. Mahdavi A, Amanzadeh M, Hamedan M. The Role of Large Language Models in Modern Medical Education: Opportunities and Challenges. *Shiraz E-Med J*. 2024;25(5):5. doi:10.5812/semj-144847
30. Baker LM, Gollop CJ. Medical textbooks: Can lay people read and understand them? Published online 2004. Accessed June 15, 2024. https://www.ideals.illinois.edu/items/1824
31. Eldeib A, Eldeib O, Mohammed A, Eldeib A, Eldeib O, Mohammed A. Examining the Effectiveness of Educational Tools and Elements Used in Medical Textbooks: Students' Perspective. *Cureus*. 2024;16. doi:10.7759/cureus.59264
32. Smith LC. *Reference and Information Services: An Introduction*. Bloomsbury Publishing USA; 2020. Accessed June 26, 2024. https://books.google.com/books?hl=en&lr=&id=lhTHEAAAQBAJ&oi=fnd&pg=PP1&dq=of+textbook+information+(Smith,+2020%5C&ots=wx5MYTOOn2&sig=7J9_7mYxMRDEjODpteueKR6PJUY
33. Wilson FL, Mood DW, Risk J, Kershaw T. Evaluation of Education Materials Using Orem's Self-Care Deficit Theory. *Nurs Sci Q*. 2003;16(1):68-76. doi:10.1177/0894318402239069
34. Kumari A, Kumari A, Singh A, et al. Large language models in hematology case solving: a comparative study of ChatGPT-3.5, Google Bard, and Microsoft Bing. *Cureus*. 2023;15(8). Accessed June 15, 2024. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10511207/
35. Uppalapati VK, Nag DS. A Comparative Analysis of AI Models in Complex Medical Decision-Making Scenarios: Evaluating ChatGPT, Claude AI, Bard, and Perplexity. *Cureus*. 2024;16(1):e52485. doi:10.7759/cureus.52485
36. Chan CKY, Lee KKW. The AI generation gap: Are Gen Z students more interested in adopting generative AI such as ChatGPT in teaching and learning than their Gen X and millennial generation teachers? *Smart Learn Environ*. 2023;10(1):60. doi:10.1186/s40561-023-00269-3
37. Cotton DRE, Cotton PA, Shipway JR. Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innov Educ Teach Int*. Published online March 13, 2023:1-12. doi:10.1080/14703297.2023.2190148
38. Lo CK. What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature. *Educ Sci*. 2023;13(4):4. doi:10.3390/educsci13040410
39. Tung AYZ, Dong LW. Malaysian Medical Students' Attitudes and Readiness Toward AI (Artificial Intelligence): A Cross-Sectional Study. *J Med Educ Curric Dev*. 2023;10:23821205231201164. doi:10.1177/23821205231201164
40. Aboalshamat K, Alhuzali R, Alalyani A, et al. Medical and dental professionals readiness for artificial intelligence for Saudi Arabia vision 2030. *Int J Pharm Res Allied Sci*. 2022;11(4-2022):52-59.
41. Dongre P. Physiology-Driven Empathic Large Language Models (EmLLMs) for Mental Health Support. In: *Extended Abstracts of the 2024 CHI Conference on Human Factors in Computing Systems*. CHI EA '24. Association for Computing Machinery; 2024:1-5. doi:10.1145/3613905.3651132
42. Akbar BMB, Prawesti DRD, Perbani WSA. Big Picture Mental Health of Generation Z in The World. *J Kesehat Komunitas Indones*. 2024;4(1):1-20. doi:10.58545/jkki.v4i1.223
43. 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;9(1):e46885. doi:10.2196/46885

