BMC Medical Education

RESEARCH Open Access

A cross sectional investigation of ChatGPT-like large language models application among medical students in China

Guixia Pan[1*] and Jing Ni[1]

Abstract

Objective To investigate the level of understanding and trust of medical students towards ChatGPT-like large language models, as well as their utilization and attitudes towards these models.

Methods Data collection was concentrated from December 2023 to mid-January 2024, utilizing a self-designed questionnaire to assess the use of large language models among undergraduate medical students at Anhui Medical University. The normality of the data was confirmed with Shapiro-Wilk tests. We used Chi-square tests for comparisons of categorical variables, Mann-Whitney U tests for comparisons of ordinal variables and non-normal continuous variables between two groups, Kruskall-Wallis H tests for comparisons of ordinal variables between multiple groups, and Bonferroni tests for post hoc comparisons.

Results A total of 1774 questionnaires were distributed and 1718 valid questionnaires were collected, with an effective rate of 96.84%. Among these students, 34.5% had heard and used large language models. There were statistically significant differences in the understanding of large language models between genders ($p < 0.001$), grade levels (junior-level students and senior-level students) ($p = 0.03$), and major ($p < 0.001$). Male, junior-level students, and public health management had a higher level of understanding of these models. Genders and majors had statistically significant effects on the degree of trust in large language models ($p = 0.004$; $p = 0.02$). Male and nursing students exhibited a higher degree of trust in large language models. As for usage, Male and junior-level students showed a significantly higher proportion of using these models for assisted learning ($p < 0.001$). Neutral sentiments were held by over two-thirds of the students (66.7%) regarding large language models, with only 51(3.0%) expressing pessimism. There were significant gender-based disparities in attitudes towards large language models, and male exhibited a more optimistic attitude towards these models ($p < 0.001$). Notably, among students with different levels of knowledge and trust in large language models, statistically significant differences were observed in their perceptions of the shortcomings and benefits of these models.

Conclusion Our study identified gender, grade levels, and major as influential factors in students' understanding and utilization of large language models. This also suggested the feasibility of integrating large language models with traditional medical education to further enhance teaching effectiveness in the future.

*Correspondence:
Guixia Pan
pgxkd@163.com

Full list of author information is available at the end of the article

Keywords ChatGPT, Artificial intelligence, Large language models, Medical education

Introduction

With the continuous advancement of educational technology, traditional teaching methods are being integrated with modern information technology. Large language models, as advanced natural language processing tools, have attracted the attention of educators and researchers for their potential applications in teaching and learning [1]. ChatGPT (Chat Generative Pre-trained Transformer) is an artificial intelligence chatbot program developed by OpenAI in November 2022 [2]. ChatGPT quickly gained global attention after its introduction, leading AI to a breakthrough in technology [3]. Within just 2 months, ChatGPT amassed over 100 million active users worldwide, becoming one of the fastest-growing consumer applications in history [3, 4]. Meanwhile, in China, various types of ChatGPT-like large-scale language models are also emerging, being applied across multiple industries, prompting widespread social attention to their development.

Medicine is a constantly evolving discipline, and healthcare professionals and medical students must strive to keep up with advancements in the field of medicine. Medical students typically encounter vast amounts of knowledge and information, ranging from anatomy to pathology and clinical practice, encompassing a wide and complex array of subjects. Traditionally, medical students acquire knowledge and experience through textbooks, classroom lectures, and clinical practice. However, with the spread of the internet and smartphones, medical students have also begun to utilize large language models similar to ChatGPT to obtain information and answer questions [5, 6]. Previous studies suggest that ChatGPT may increase students' interest in learning [3]. However, researchers are concerned that students may rely too heavily on large language models as their primary source of information [7, 8].

Since ChatGPT-like models can provide information rather than in-depth knowledge, their responses to general questions may offer factual knowledge that sometimes lacks practical relevance to specific clinical scenarios [3]. Given that artificial intelligence chatbots like ChatGPT can interactively access factual knowledge on smartphones, medical students may incorporate such services into their daily routines. However, educators in medical schools are concerned that responses generated by ChatGPT-like models have not undergone extensive validation and reliability testing, which could potentially impact the learning quality of students who rely on them [9, 10].

The emergence of ChatGPT-like large language models has significantly influenced teaching practices and students' learning experiences. Whether to ban its usage or determine its future application in education poses opportunities and challenges. In this era of educational digital transformation, it is crucial to proactively address the impact of large language models like ChatGPT. Exploring how such tools can drive educational innovation is key to effectively shaping education's future. Furthermore, comprehending, explaining, and predicting the interplay between large language models such as ChatGPT and the evolution of medical education from various educational perspectives is essential. Educators must respond and adapt to these changes accordingly.

Research into large language models is expanding rapidly. Previous studies primarily aimed to validate their capabilities [11–13]. For instance, ChatGPT demonstrated proficiency in passing the United States Medical Licensing Examination [14], while the Med-PaLM 2 model approached the expertise level of human clinical professionals [15]. However, comprehensive surveys on student applications of large language models remain scarce. In South Korea, a study highlighted students' keen interest in integrating ChatGPT into classrooms but did not deeply explore their perceptions or trust in the model [16]. Similarly, another study surveyed healthcare professionals' views on ChatGPT enhancing medical knowledge training, revealing that most participants were enthusiastic about ChatGPT-assisted training, but the study was limited by a small sample size [17]. Alkhaaldi SMI et al. surveyed medical students' perceptions of ChatGPT and artificial intelligence, but the study only focused on graduating students, failing to fully capture the perspectives across different academic years [18]. Li et al. demonstrated that large language models could enhance the academic writing proficiency of non-native English students in China, with most students showing a positive attitude toward these models [19]. However, few studies have comprehensively explored the current application status of large language models among medical students in China.

Therefore, this study systematically investigates the current application status of large language models among 1,718 medical students of different academic years from five different majors in China. We evaluated their degree of understanding and trust in these models, their usage, and their perspectives on the benefits and drawbacks. Additionally, we analyzed differences based on gender, major, and academic year, aiming to provide new insights into integrating large language models with medical education.

Purpose and research question

As a newly emerging practical tool, the awareness of large language models among the population is low, and there is a lack of effective practical research to evaluate its prospect and utility in medical education. Thus, our research examines the level of understanding and trust of medical students towards ChatGPT-like large language models, as well as their utilization and attitudes towards these models, from the perspectives of medical students. It also assesses the feasibility of using such models among medical students.

Although undergraduate students already use the ChatGPT-like models, there remains a lack of extensive systematic surveys of these models among medical students. The findings of this study illuminate the present utilization of large language models among undergraduate students, offering educators propositions on integrating these models into traditional medical education. These insights provide perspectives for curriculum reform, fostering new ideas and approaches in medical education.

Method

Subjects and procedure

A cross-sectional survey of undergraduate medical students at Anhui Medical University was conducted using convenient sampling, and survey data collection concentrated from late December 2023 to mid-January 2024. All undergraduates in the second semester of the 2023–2024 academic year who were willing to participate in the study were included. Those who were graduate students and declined to participate were excluded from the study. A total of 1774 students from five majors participated in this survey, among whom 56 filled out only partial or solely the personal information section of the questionnaire, so excluding them from the analysis. Ultimately, 1718 undergraduate medical students were included in this study, consisting of 898 females and 820 males.

Data collection and questionnaire design

The data collection was completed in January 2024. A self-designed questionnaire was used to assess the application of ChatGPT-like large language models among undergraduate medical students in this study. The questionnaires were anonymously completed in the classroom. The questionnaire was developed by referencing previous literature and consisted of two main parts: the first part included demographic information about the students such as gender, age, major, year of study; the second part included questions about the degree of understanding and trust, usage and attitudes of medical students towards large language models. The investigators provided a brief description of the purpose and significance of the survey prior to distributing the questionnaire. The final version of the questionnaire included 12 items (Appendix 1).

Statistical analysis

We verified the normality of the data using the Shapiro-Wilk test. Demographic characteristics for medical students were characterized using median and interquartile range (IQR) for non-normal continuous variables, and frequencies (percentage) for categorical variables. We used Chi-square tests for comparisons of categorical variables between multiple groups. Mann-Whitney U tests were applied for comparisons of ordinal variables and non-normal continuous variables between two groups, while Kruskall-Wallis H tests were used for comparisons of ordinal variables between multiple groups. Post hoc comparisons were conducted using Bonferroni tests. All statistical analyses were performed with IBM SPSS Statistics, version 23 (IBM, Armonk, New York, USA). A level of statistical $p<0.05$ was set for all the tests and considered to be statistically significant.

Results

Population characteristics

A total of 1718 undergraduate students participated in this study, ranging from 1st to 5th year of study. The median (IQR) of age was 20.0 (19.0–21.0), with 898 (52.3%) female students. The numbers of students in the majors of preventive medicine, clinical medicine, public health management, nursing, and basic medicine were 301 (17.5%), 1228 (71.5%), 73 (4.2%), 75 (4.4%), and 41 (2.4%), respectively. Among all students, 1125 (65.5%) students reported that they did not utilize the large language models for assisted learning. Compared to females, males were more likely to use these models for assisted learning (Table 1).

Degree of understanding and trust in large language models

Among 1718 undergraduate medical students, 824 (48.0%) had heard but never used the large language models, and 593 (34.5%) had heard and used them (Fig. 1). The Mann-Whitney U test indicated that there were statistically significant differences in the understanding of large language models between genders ($p<0.001$) (Table S1), and grade levels (junior-level students and senior-level students) ($p=0.03$) (Table S1), Male and junior-level students had a higher level of understanding of large language models. Additionally, compared to other majors (preventive medicine, clinical medicine, public health management, nursing, and basic medicine), public health management has a higher level of understanding of these models, as evidenced by a Kruskal-Wallis H test (Table S1).

Table 1 Demographic data for undergraduate medical students according to the usage of large language models ($n=1718$)

Variables	Overall, $N=1718$[1]	Used large language models		*p*-value[2]
		Yes, $N=593$	No, $N=1125$	
Age, Median (IQR)	20.0 (19.0–21.0)	20.0 (19.0–21.0)	20.0 (19.0–21.0)	0.07
Gender, n(%)				< 0.001
Female	898 (52.3)	222 (37.4)	676 (60.1)	
Male	820 (47.7)	371 (62.6)	449 (39.9)	
Year of study (level), n (%)				< 0.001
1st	31 (1.8)	6 (1.0)	25 (2.2)	
2nd	541 (31.5)	220 (37.1)	321 (28.5)	
3rd	831 (48.4)	284 (47.9)	547 (48.6)	
4th	215 (12.5)	46 (7.8)	169 (15.0)	
5th	100 (5.8)	37 (6.2)	63 (5.6)	
Major (%)				0.01
Preventive medicine	301 (17.5)	114 (13.9)	187 (20.8)	
Clinical medicine	1228 (71.5)	630 (76.8)	598 (66.6)	
Public health management	73 (4.2)	29 (3.5)	44 (4.9)	
Nursing	75 (4.4)	27 (3.3)	48 (5.3)	
Basic medicine	41 (2.4)	20 (2.4)	21 (2.3)	

[1] IQR, interquartile range; Data were presented as n (%) and median (IQR)

[2] Mann-Whitney U test; Chi-squared test

In terms of the degree of trust in the information provided by large language models, 1326 students (77.2%) demonstrated a relatively high level of trust, with only 134 (7.8%) expressing strong trust (Fig. 1). Male and nursing students exhibited a higher degree of trust in large language models (Table S1), but different grade levels did not differ ($p=0.31$) (Table S1).

Usage of large language models

Of the 1718 undergraduate medical students, 651 reported using the large language models for assisted learning (Table S2). Males and junior-level students showed a significantly higher proportion of using these models for assisted learning ($p<0.001$) (Table S2). Bonferroni-corrected comparisons among majors indicated that nursing students, compared to those in preventive medicine, used large language models more frequently for assisted learning (Table S2).

Attitudes towards large language models

More than half of the undergraduate students (66.7%) were neutral about the large language models and only 51 (3.0%) were pessimistic (Fig. S1). A significant gender-based disparity was observed in attitude towards large language models ($p<0.001$) (Table 2). Male exhibit a more optimistic attitude towards these models. However, no statistically significant variances were found in attitudes towards large language models across grade levels (Table 2). There were differences between gender in their perceptions of the pros and cons of using large language models, Men were more likely to believe that the use of these models had more benefits than drawbacks (Table 2). Specifically, gender played a role in the perception of shortcomings associated with the large language models. Compared to female medical students, more male recognized the drawbacks of large language models, including offering ineffective assistance, resulting in a lack of interpretability, fabricating content haphazardly,

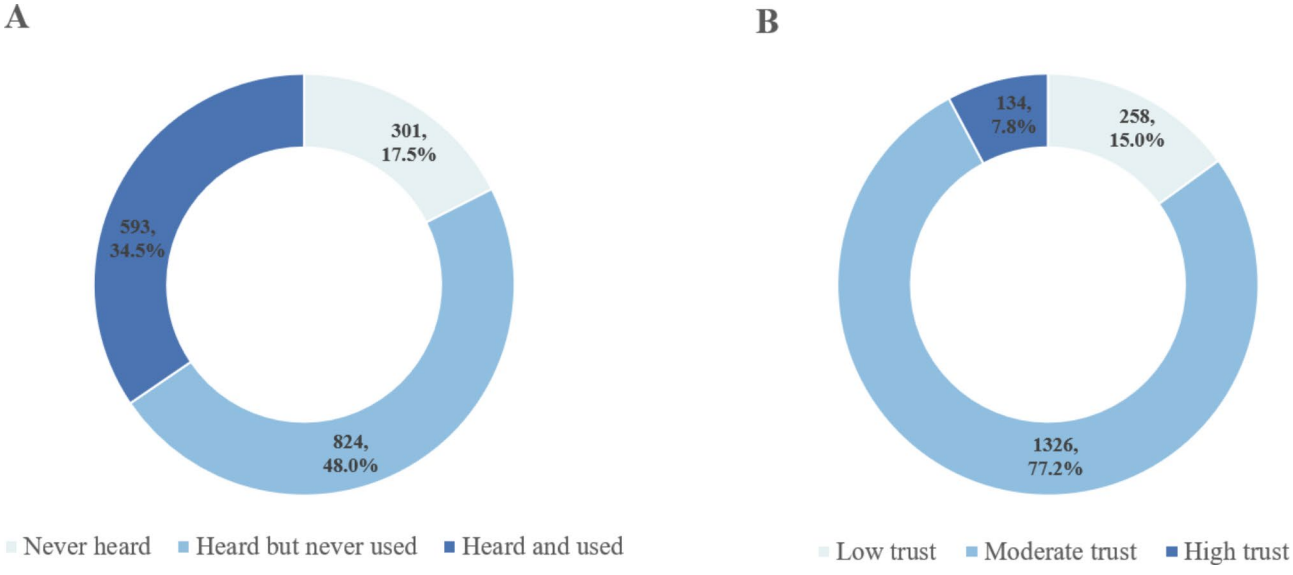

Fig. 1 Degree of understanding and trust in large language models. **A**. Degree of understanding in large language models; **B**. Degree of trust in the information provided by large language models

Table 2 Perceptions and attitudes of undergraduate medical students towards the pros and cons of using large language models

	Gender		Grade levels[2]	
	Male	Female	Junior-level students	Senior-level students
Attitude towards large language models[1]				
Pessimistic	29	22	26	25
Neutral	501	645	380	766
Optimistic	290	231	166	355
Mean Rank	898.11	824.25	839.20	869.64
Z	-3.75		-1.50	
p-value	< 0.001		0.14	
The pros and cons of using large language models[1]				
Cons outweigh pros	46	57	37	66
Equal pros and cons	409	535	308	636
Pros outweigh cons	365	306	227	444
Mean Rank	905.57	817.43	862.10	858.20
Z	-4.18		-0.20	
p-value	< 0.001		0.86	

[1] Mann-Whitney U test

[2] Junior-level students: first-year and second-year students; Senior-level students: third-year, fourth-year and fifth-year students

limited intelligence and understanding of things, superficial content with a strong sense of patchiness, and unable to grasp core issues within context (Table 3). Similarly, there were significant differences in perceptions of the shortcomings of the large language models among students with different degree of knowledge and trust in the large language models (Table 3).

Furthermore, it is notable that statistically significant disparities emerged among students with varying levels of knowledge and trust in large language models regarding their perceptions of the benefits of such models. These advantages include their versatility across diverse scenarios and exhibiting robust scalability, robust linguistic abilities, assisting in addressing everyday challenges, enhancing learning and work efficiency while alleviating burdens, and delivering superior intelligent services. Specifically, after applying Bonferroni corrections, there existed statistically significant differences in the perceptions regarding the advantages and shortcomings of large language models between students who never heard them and those who heard and used such models. Students who had used them were more capable of recognizing the advantages of large language models (Table 4). Similarly, students with high levels of trust were more likely to recognize the benefits of the models than those who did not trust these models (Table 4).

Discussion

The findings of this study shed light on the perceptions and usage of large language models among medical students. A significant proportion of students had heard but never used these models. Furthermore, gender, grade levels, and major all influenced students' understanding and utilization of large language models. In terms of trust in the information provided by large language models, a majority of students expressed relatively high trust levels, with gender and major exerting significant effects on the degree of trust. As for usage, a notable proportion of students reported utilizing large language models for assisted learning, with differences observed across gender, grade levels, and major. Notably, preventive medicine students exhibited higher usage compared to nursing students. Attitudes towards large language models were generally neutral, with a minority expressing pessimism. Gender-based disparities were evident in attitudes, particularly in perceptions of the pros and cons of using these models for learning. Specifically, the shortcomings of larger language models such as ineffective assistance and limited interpretability were more likely recognized by male students. Additionally, students' perceptions of the benefits and shortcomings of large language models varied significantly based on their degree of knowledge and trust. Those who had never heard of these models differed in their perceptions compared to those who had heard and used them, highlighting the influence of personal experience on attitudes.

In this study, we found that gender, grade levels, and major affect medical students' perceptions, usage, trust levels, and attitudes of the large language models. These differences may be due to personal experiences, educational backgrounds, and behavioral preferences. Specifically, a study of medical students showed that female students have higher critical thinking disposition than male students, including truth-seeking, open-mindedness, and maturity of judgment [20]. This could lead to female students being more cautious when using large language models, potentially contributing to the differences in perceptions, usage, and attitudes of female and male towards large language models. As for grade levels, a study showed that younger individuals exhibited greater readiness to adopt AI technologies [2]. This is consistent with our study, where junior-level students showed a higher proportion of using large language models for assisted learning, along with a higher level of understanding and trust in these models. Senior students may have more learning experiences, leading to a deeper understanding of these models which makes them adopt a more cautious attitude towards large language models. In terms of majors, students from different majors may have diverse needs and application scenarios for large language models. For instance, clinical medicine students

Table 3 Perceptions of medical students towards the shortcomings of large language models

	Gender[1]		Understanding of large language models[2]			Degree of trust in the information provided by large language models[2]		
	Male	Female	Never heard	Heard but never used	Heard and used	Low trust	Moderate trust	High trust
Offering ineffective assistance								
Agree	320 (39.0)	261 (29.1)	94	256	231	132	414	35
Disagree	500 (61.0)	637 (70.9)	207	568	362	126	912	99
$\chi 2$	18.90		10.68			42.22		
p-value	< 0.001		< 0.001			< 0.001		
Lacking the ability to reason through complex issues								
Agree	462 (56.3)	472 (52.6)	125	454	355	169	720	45
Disagree	358 (43.7)	426 (47.4)	176	370	238	89	606	89
$\chi 2$	2.47		27.40			36.23		
p-value	0.12		< 0.001			< 0.001		
Results lack interpretability								
Agree	426 (52.0)	420 (46.8)	117	426	303	167	633	46
Disagree	394 (48.0)	478 (53.2)	184	398	290	91	693	88
$\chi 2$	4.60		15.76			37.88		
p-value	0.03		< 0.001			< 0.001		
Fabricating content haphazardly								
Agree	285 (34.8)	224 (24.9)	78	227	204	122	353	34
Disagree	535 (65.2)	674 (75.1)	223	597	389	136	973	100
$\chi 2$	19.79		10.18			45.50		
p-value	< 0.001		0.006			< 0.001		
Limited intelligence and understanding of things								
Agree	484 (59.0)	456 (50.8)	112	447	381	166	719	55
Disagree	336 (41.0)	442 (49.2)	189	377	212	92	607	79
$\chi 2$	11.76		59.06			19.88		
P-value	< 0.001		< 0.001			< 0.001		
Superficial content with a strong sense of patchiness								
Agree	477 (58.2)	459 (51.1)	113	450	373	172	720	44
Disagree	343 (41.8)	439 (48.9)	188	374	220	86	606	90
$\chi 2$	8.61		51.79			40.78		
P-value	0.003		< 0.001			< 0.001		
Unable to grasp core issues within context								
Agree	525 (64.0)	507 (56.5)	131	517	383	178	797	56
Disagree	296 (36.1)	391 (43.5)	170	307	210	80	529	78
$\chi 2$	9.90		41.84			27.21		
P-value	0.002		< 0.001			< 0.001		

[1] Chi-Squared test

[2] Kruskal-Wallis H test

might prefer using models for clinical decision support [1, 21]. A recent study also indicated that healthcare workers expressed significant interest in utilizing ChatGPT for medical research, but their interest was less pronounced for patient care purposes [22]. This is consistent with our viewpoint that students from different majors may have varying intentions in using large language models, resulting in differences in their understanding, trust, and usage of such models.

Recently, there has been growing interest in utilizing large language models in medicine to enrich basic medical knowledge, facilitate clinical learning, and promote innovation [1, 23–29]. Based on the current study, using large language models in medical education is reasonable. First, these models can enrich students' medical knowledge and enhance their research capabilities. Furthermore, the majority of medical students held optimistic or neutral views and had confidence in the results of the large language models. Lastly, since large language models primarily communicate through text, they seamlessly integrate with traditional learning formats. Some research aligns with our viewpoint. Lee has suggested that ChatGPT possesses the capacity to heighten student involvement and enrich student learning experiences

Table 4 Perceptions of medical students towards the benefits of large language models

	Gender[1]		Understanding of large language models[2]			Degree of trust in the information provided by large language models[2]		
	Male	Female	Never heard	Heard but never used	Heard and used	Low trust	Moderate trust	High trust
Versatile across diverse scenarios and exhibiting robust scalability								
Agree	746 (91.0)	829 (92.3)	240	769	566	189	1257	129
Disagree	74 (9.0)	69 (7.7)	61	55	27	69	69	5
χ2	1.01		70.20			135.34		
p-value	0.32		< 0.001			< 0.001		
Robust linguistic abilities								
Agree	740 (90.2)	814 (90.6)	246	762	546	195	1228	131
Disagree	80 (9.8)	84 (9.4)	55	62	47	63	98	3
χ2	0.08		32.25			81.51		
p-value	0.77		< 0.001			< 0.001		
Assisting in addressing everyday challenges								
Agree	695 (84.8)	778 (86.6)	230	711	532	170	1176	127
Disagree	125 (15.2)	120 (13.3)	71	113	61	88	150	7
χ2	1.24		29.28			101.50		
p-value	0.27		< 0.001			< 0.001		
Enhance learning and work efficiency while alleviating burdens								
Agree	743 (90.6)	812 (90.4)	243	745	567	182	1243	130
Disagree	77 (9.4)	86 (9.6)	58	79	26	76	83	4
χ2	0.02		51.50			142.50		
p-value	0.90		< 0.001			< 0.001		
Delivering superior intelligent services								
Agree	741 (90.4)	834 (92.9)	246	768	561	189	1258	128
Disagree	79 (9.6)	64 (7.1)	55	56	32	69	68	6
χ2	3.53		48.22			135.07		
p-value	0.06		< 0.001			< 0.001		

[1] Chi-Squared test

[2] Kruskal-Wallis H test

[30]. In addition, Kung et al. have demonstrated that ChatGPT can effectively process intricate medical and clinical information with a high level of accuracy [14]. This further provides evidence for the rationality of integrating large language models into medical education. Nevertheless, it's important to consider factors such as gender, grade levels, and major when utilizing large language models for medical education. Tailored teaching strategies should be implemented for various demographic groups to enhance teaching effectiveness. Meanwhile, although the large language models have robust learning and processing capabilities, the information they offer may not always be entirely accurate. Therefore, medical education should still be rooted in classroom instruction and practical training, rather than solely relying on the large language models.

When using large language models for medical education, it's crucial to tailor their implementation based on the diverse perceptions, usage patterns, trust levels, and attitudes observed among medical students. This customization should consider factors such as gender, grade level, and major, which influence how students perceive and utilize these tools. First, recognize and address potential gender differences in how students interact with and trust large language models. Provide support for fostering critical thinking and ensuring inclusivity in educational content. Encourage students to critically evaluate the reliability, bias, and potential impact of these technologies on patient care and medical practice. Second, tailor the use of large language models to align with the educational goals of different medical majors. For clinical medicine students, emphasize applications in clinical decision support, while for research-focused majors, highlight applications in medical research and data analysis. Third, regularly assess student perceptions, understanding, and satisfaction with the use of large language models. Incorporate feedback to refine educational strategies and improve the integration of these tools into the curriculum. By adopting these strategies, medical education can effectively capitalize on the benefits of large language models while addressing the unique needs and perspectives of diverse student populations.

This study demonstrates several strengths. First, it encompassed 1,718 diverse medical students from various majors, including preventive medicine, clinical medicine, nursing, public health management, and basic medicine, thereby enhancing the applicability of its findings to the broader medical student population. Second, it contributes to the limited research by systematically and comprehensively exploring the relationship between various factors and the usage of large language models, offering new insights into their integration with medical education.

Several limitations in this study warrant attention. First, individual information, such as academic performance, urban/rural background, and family income, were not included, potentially impacting medical students' perceptions and acceptance of large language models. Second, this study employed a cross-sectional design to investigate the use of the large language models by medical students. A new medical education model incorporating the large language models should be further explored through longitudinal studies and intervention trials.

Conclusions

This study provides insights into the perceptions and usage of large language models among undergraduate medical students. Gender, grade levels, and major were found to influence students' understanding and utilization of these models. Combining large language models with traditional medical education to enhance teaching effectiveness seems feasible.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12909-024-05871-8.

Supplementary Material 1

Acknowledgements
The authors would like to thank Wenbing Fang, Yaning Sun and Ruyu Ni for their help in collecting the data. We also would like to thank all participants during the investigational period for their cooperation.

Author contributions
Guixia Pan contributed to the conception and design of the study, questionnaire design, and drafting the original manuscript. Jing Ni contributed to the data analysis and questionanaire design. Jing Ni and Guixia Pan reviewed the manuscript and approved the final draft submitted.

Funding
This work was supported by the Anhui New Era education quality engineering project (2022zyxwjxalk050) and Education Department of Anhui quality engineering project (2023zyxwjxalk056).

Data availability
The data of this study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
All participants have provided written informed consent and the Ethics Committee of Anhui Medical University has approved the study (No. 83241200).

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Epidemiology and Biostatistics, School of Public Health, Anhui Medical University, Meishan Road 81, Hefei 230032, Anhui, China

Received: 22 May 2024 / Accepted: 7 August 2024
Published online: 23 August 2024

References

1. Liu J, Wang C, Liu S. Utility of ChatGPT in Clinical Practice. J Med Internet Res. 2023;25:e48568.
2. Sharma S, Pajai S, Prasad R, Wanjari MB, Munjewar PK, Sharma R, Pathade A. A critical review of ChatGPT as a potential substitute for diabetes educators. Cureus. 2023;15(5):e38380.
3. Sabry Abdel-Messih M, Kamel Boulos MN. ChatGPT in clinical toxicology. JMIR Med Educ. 2023;9:e46876.
4. Mohammad B, Supti T, Alzubaidi M, Shah H, Alam T, Shah Z, Househ M. The pros and cons of using ChatGPT in Medical Education: a scoping review. Stud Health Technol Inf. 2023;305:644–7.
5. Currie G, Singh C, Nelson T, Nabasenja C, Al-Hayek Y, Spuur K. ChatGPT in medical imaging higher education. Radiography (Lond). 2023;29(4):792–9.
6. 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.
7. Sedaghat S. Early applications of ChatGPT in medical practice, education and research. Clin Med (Lond). 2023;23(3):278–9.
8. Fatani B. ChatGPT for Future Medical and Dental Research. Cureus. 2023;15(4):e37285.
9. Dave T, Athaluri SA, Singh S. ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. Front Artif Intell. 2023;6:1169595.
10. Lecler A, Duron L, Soyer P. Revolutionizing radiology with GPT-based models: current applications, future possibilities and limitations of ChatGPT. Diagn Interv Imaging. 2023;104(6):269–74.
11. Thirunavukarasu AJ, Hassan R, Mahmood S, Sanghera R, Barzangi K, El Mukashfi M, Shah S. 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;9:e46599.
12. Sarraju A, Bruemmer D, Van Iterson E, Cho L, Rodriguez F, Laffin L. Appropriateness of Cardiovascular Disease Prevention recommendations Obtained from a Popular Online Chat-based Artificial Intelligence Model. JAMA. 2023;329(10):842–4.
13. Nori H, King N, McKinney SM, Carignan D, Horvitz E. Capabilities of GPT-4 on Medical Challenge Problems. *ArXiv* 2023, abs/2303.13375.
14. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, Madriaga M, Aggabao R, Diaz-Candido G, Maningo J, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. PLOS Digit Health. 2023;2(2):e0000198.
15. Singhal K, Tu T, Gottweis J, Sayres R, Wulczyn E, Hou L, Clark K, Pfohl SR, Cole-Lewis HJ, Neal D et al. Towards Expert-Level Medical Question Answering with Large Language Models. *ArXiv* 2023, abs/2305.09617.
16. Park J. Medical students' patterns of using ChatGPT as a feedback tool and perceptions of ChatGPT in a Leadership and Communication course in Korea: a cross-sectional study. J Educ Eval Health Prof. 2023;20:29.
17. Hu JM, Liu FC, Chu CM, Chang YT. Health Care trainees' and professionals' perceptions of ChatGPT in improving medical knowledge training: Rapid Survey Study. J Med Internet Res. 2023;25:e49385.

18. Alkhaaldi SMI, Kassab CH, Dimassi Z, Oyoun Alsoud L, Al Fahim M, Al Hageh C, Ibrahim H. Medical student experiences and perceptions of ChatGPT and Artificial Intelligence: cross-sectional study. JMIR Med Educ. 2023;9:e51302.
19. Li J, Zong H, Wu E, Wu R, Peng Z, Zhao J, Yang L, Xie H, Shen B. Exploring the potential of artificial intelligence to enhance the writing of English academic papers by non-native English-speaking medical students - the educational application of ChatGPT. BMC Med Educ. 2024;24(1):736.
20. Zhai J, Zhang H. Critical thinking disposition of medical students in Anhui Province, China: a cross-sectional investigation. BMC Med Educ. 2023;23(1):652.
21. Liu S, Wright AP, Patterson BL, Wanderer JP, Turer RW, Nelson SD, McCoy AB, Sittig DF, Wright A. Using AI-generated suggestions from ChatGPT to optimize clinical decision support. J Am Med Inf Assoc. 2023;30(7):1237–45.
22. Temsah MH, Aljamaan F, Malki KH, Alhasan K, Altamimi I, Aljarbou R, Bazuhair F, Alsubaihin A, Abdulmajeed N, Alshahrani FS et al. ChatGPT and the Future of Digital Health: A Study on Healthcare Workers' Perceptions and Expectations. *Healthcare (Basel)* 2023, 11(13).
23. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. Nat Med. 2023;29(8):1930–40.
24. Clusmann J, Kolbinger FR, Muti HS, Carrero ZI, Eckardt JN, Laleh NG, Löffler CML, Schwarzkopf SC, Unger M, Veldhuizen GP, et al. The future landscape of large language models in medicine. Commun Med (Lond). 2023;3(1):141.
25. Shah NH, Entwistle D, Pfeffer MA. Creation and adoption of large Language models in Medicine. JAMA. 2023;330(9):866–9.
26. Betzler BK, Chen H, Cheng CY, Lee CS, Ning G, Song SJ, Lee AY, Kawasaki R, van Wijngaarden P, Grzybowski A, et al. Large language models and their impact in ophthalmology. Lancet Digit Health. 2023;5(12):e917–24.
27. Qiu J, Yuan W, Lam K. The application of multimodal large language models in medicine. Lancet Reg Health West Pac. 2024;45:101048.
28. Rengers TA, Thiels CA, Salehinejad H. Academic surgery in the era of large Language models: a review. JAMA Surg. 2024;159(4):445–50.
29. Jowsey T, Stokes-Parish J, Singleton R, Todorovic M. Medical education empowered by generative artificial intelligence large language models. Trends Mol Med. 2023;29(12):971–3.
30. Lee H. The rise of ChatGPT: exploring its potential in medical education. Anat Sci Educ 2023.

Publisher's Note