

The Comprehensive Impact of Large Language Models on Medical Students' Research, Learning, and Stress: A Cross-Sectional Study

Weitong Zhao

First Hospital of China Medical University

Xiang Li

First Hospital of China Medical University

Zhengtao Gao

China Medical University

Zhiguang Chen

First Hospital of China Medical University

Liang Sang

13889167622@163.com

First Hospital of China Medical University

Research Article

Keywords: Large Language Models, Medical Students, Medical Education, Cross-Sectional Study

Posted Date: September 22nd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10590073/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

Large language models (LLMs) have rapidly expanded across medical education and research. However, systematic quantitative evidence on their usage patterns, perceived benefits, risk awareness, and their potential integration into medical education across different stages of medical training remains limited, particularly for emerging domestic models.

Objective

This study aimed to evaluate the usage characteristics of LLMs among medical students at different educational levels in China and to examine their implications for medical education and academic research.

Methods

A cross-sectional survey was conducted among four groups—undergraduates, academic master’s students, professional master’s students, and doctoral students—using a structured questionnaire. The survey assessed LLM usage patterns, application scenarios, risk awareness, and attitudes toward potential medical education reform. Data were analyzed using descriptive statistics and between-group comparisons.

Results

A total of 334 valid questionnaires were obtained, and 98.2% of respondents reported prior use of LLMs. DeepSeek and Doubao were the most frequently used models, with usage rates of 45.4% and 31.1%, respectively. Significant differences across educational levels were observed in usage frequency, duration, model preference, and application scenarios. Research and innovation-related use increased with higher educational levels, whereas exam-oriented and clinically focused learning declined. Risk awareness and the ability to critically evaluate generated information also increased with higher educational attainment. Research-related stress showed a structural divergence: output-related pressure increased with academic level, whereas skill-related pressure decreased.

Conclusion

LLMs have become integral tools for medical students’ learning and research, with distinct functional roles across stages of medical training. Future medical education should move beyond the question of

whether to use LLMs and focus instead on establishing standardized and effective strategies for their use.

Introduction

In recent years, LLMs, exemplified by ChatGPT and DeepSeek, have rapidly expanded across higher education, stimulating extensive debate regarding educational models and research paradigms [1]. Owing to their advanced natural language processing capabilities, LLMs have shown considerable potential in medical knowledge retrieval, clinical decision support, and medical education [2]. Medicine, a discipline that closely integrates theoretical knowledge with clinical practice, is at the forefront of this technological shift [3]. Medical students must acquire extensive theoretical knowledge while developing clinical decision-making and research-oriented thinking in complex clinical environments. The strong semantic processing and reasoning capabilities of LLMs offer new opportunities to support these diverse learning demands [4].

Within the medical domain, LLMs are increasingly used not only to support professional knowledge acquisition [5] and simulate doctor–patient communication scenarios [6, 7], but also to assist medical students in academic writing and research design [8]. However, the growing adoption of these technologies is accompanied by important risks. Issues such as model “hallucinations” [9, 10], data privacy concerns [11], and potential academic misconduct [12] raise significant ethical challenges for medical education and research.

Although preliminary studies have explored the use of LLMs among medical students, existing research has largely focused on Western contexts or early models such as ChatGPT [13, 14]. In the context of rapid technological iteration and the emergence of high-performance domestic models, these findings may no longer adequately capture current usage patterns. In recent years, the increasing adoption of domestic LLMs in China such as DeepSeek, has substantially reshaped how Chinese medical students engage with these technologies [15,16]. Moreover, many previous studies have treated medical students as a homogeneous population [14], overlooking educational level as a key analytical variable. Medical education is characterized by clear stage-specific training objectives: undergraduate students primarily focus on acquiring foundational knowledge; professional master’s students emphasize clinical skill development; and academic master’s and doctoral students place greater emphasis on research and innovation [17]. These differences in training priorities are likely to result in substantial variation across educational levels in model preference, usage frequency, application scenarios, and risk perception regarding LLMs.

Moreover, although LLMs have become important tools for medical students’ learning and research, their role within formal medical education remains unclear [18]. Medical curricula often lag behind technological developments, leaving many students to explore the use of LLMs independently without systematic training in LLM literacy. In the absence of structured training and ethical guidance, the use of these technologies may lead to overreliance, reduced critical evaluation, and potential risks to academic

integrity. Therefore, determining whether medical students support the integration of artificial intelligence (AI) into medical education, identifying the stages at which it should be incorporated, and exploring strategies for its responsible and standardized use represent key challenges for contemporary medical education reform. A comprehensive understanding of students' perceptions of AI in medical education is essential for developing evidence-informed educational reforms.

In this context, the present study conducted a cross-sectional survey among undergraduate, master's, and doctoral medical students at China Medical University with three objectives: (1) to compare LLM usage behaviors across different educational levels; (2) to examine differences in risk perception, information evaluation ability, and research-related pressure associated with LLM use; and (3) to investigate students' attitudes toward integrating AI into formal medical education and to identify the stages of training at which they believe AI should be introduced.

Method

Participants and procedure

This study employed a cross-sectional questionnaire survey design conducted at China Medical University between January and March 2026. A stratified cluster sampling strategy was used, with educational level as the stratification criterion. A total of 357 questionnaires were collected. After screening for completeness and logical consistency, 334 responses were deemed valid and included in the final analysis (valid response rate: 93.6%). The final sample comprised four educational groups: undergraduate students ($n = 106$), professional master's students ($n = 84$), academic master's students ($n = 99$), and doctoral students ($n = 45$).

Data collection and questionnaire development

Data collection was completed in March 2026. The questionnaire was developed based on existing literature and adapted to the context of Chinese medical education. A small-scale pilot test was conducted prior to the formal survey to assess item clarity and internal consistency. The questionnaire consisted of eight sections (see Supplementary Materials for details).

1. Basic Information and Usage Overview;
2. Application of LLMs in Academic Research and Writing Support;
3. Application of LLMs in Learning, Exams, and Clinical Practice;
4. Risk Awareness and Information Evaluation Ability of LLMs;
5. Information Evaluation and Critical Thinking Assessment;
6. Research-Related Pressure Assessment;
7. LLM's Contribution to Research Output;
8. Future Application of LLMs in Medical Education.

Except for objective multiple-choice questions, all scale items were assessed using a five-point Likert scale (1 = strongly disagree, 5 = strongly agree). Several items were reverse-scored to evaluate response consistency. All subscales demonstrated acceptable internal consistency (Cronbach's $\alpha > 0.70$).

Inclusion and exclusion criteria

Clear inclusion and exclusion criteria were defined for participant selection. The inclusion criteria were as follows: (1) full-time registered medical students at China Medical University during the survey period (January-March 2026); (2) students enrolled in undergraduate, master's, or doctoral programs; and (3) individuals who understood the study purpose and provided voluntary consent to participate. The exclusion criteria were: (1) students from non-medical majors, visiting students, or exchange students; (2) questionnaires with incomplete responses or clear logical inconsistencies; and (3) respondents who failed the reverse-scored validity checks embedded in the questionnaire.

Statistical analysis

All data were analyzed using SPSS software. Categorical variables (e.g., model preference) were analyzed using the chi-square test or Fisher's exact test when appropriate. Continuous variables were analyzed using one-way analysis of variance (ANOVA), followed by appropriate post hoc comparisons. The internal consistency of the scales was assessed using Cronbach's α coefficients. Differences in scores among items within the same scale dimension were examined using the Friedman test. Differences in selection proportions among response options in multiple-choice items were analyzed using Cochran's Q test. A two-sided p value < 0.05 was considered statistically significant.

Result

1. High Prevalence and Model Preferences of LLMs

Among the 334 respondents, 328 (98.2%) reported having used LLMs. Subsequent analyses of model preference were therefore restricted to these 328 users. This high prevalence did not differ significantly across educational levels, indicating that LLMs have been widely adopted among medical students. In terms of model preference, DeepSeek (45.4%) and Doubao (31.1%) were the most frequently used platforms, with significantly higher usage rates than ChatGPT (15.2%) and Gemini (5.5%) ($p < 0.001$).

Between-group comparisons revealed that doctoral students showed significantly higher preference rates for ChatGPT (31.8%) and Gemini (15.9%) than the other educational groups (both $p < 0.05$). In contrast, undergraduates demonstrated a significantly stronger preference for Doubao than the other three groups (all $p < 0.001$). Meanwhile, the preference rate for DeepSeek did not differ significantly across educational levels ($p = 0.52$), with all groups maintaining consistently high usage rates of approximately 40%–50% (**Table 1 and Figure 1a**).

Table 1.Distribution of LLM model preferences among medical students across educational levels and results of between-group comparisons (N = 328).

Education	ChatGPT	DeepSeek	Gemini	Doubao	Kimi	Others	Total
Undergraduate	8 (7.8%)	42 (40.8%)	1 (1.0%)	51 (49.5%)	1 (1.0%)	0	103
Professional Master	13 (15.7%)	43 (51.8%)	4 (4.8%)	20 (24.1%)	3 (3.6%)	0	83
Academic Master	15 (15.3%)	46 (46.9%)	6 (6.1%)	26 (26.5%)	4 (4.1%)	1 (1.0%)	98
PhD	14 (31.8%)	18 (40.9%)	7 (15.9%)	5 (11.4%)	0	0	44
p-value	0.003	0.52	0.004	< 0.001	0.72	1.000	
Total	50 (15.2%)	149 (45.4%)	18 (5.5%)	102 (31.1%)	8 (2.8%)	1 (0.3%)	328

2. Differences in LLM Usage Patterns and Application Scenarios

Significant differences in usage frequency and duration were observed across educational levels (p < 0.001). Doctoral students and academic master’s students reported significantly higher usage frequency than undergraduates and professional master’s students.

Regarding application scenarios, clear stratification was evident across educational levels (**Figure 1b**). Scenarios related to research and innovation showed a clear increasing trend with higher educational levels: "Stimulating research ideas and insights" increased from 28.2% among undergraduates to 65.9% among doctoral students (p<0.001); "Assisting in research design and proposal writing" increased from 30.1% to 56.8% (p=0.002); and "Programming and data analysis" increased from 9.7% to 34.1% (p=0.005). In contrast, scenarios related to examinations and clinical learning showed a decreasing trend with higher educational levels: "Preparing for exams and tests" decreased from 46.6% in undergraduates to 11.4% in doctoral students (p<0.001), and "Analyzing and discussing clinical cases" decreased from 29.1% to 4.5% (p=0.001). No significant differences across educational levels were observed for foundational learning and general tool scenarios, indicating that the use of LLMs for basic learning and general purposes represents a common need among medical students (**Table S1**).

3. Widespread Use of LLMs in Academic Research and Writing Support

Among the 328 students who reported using LLMs, 255 (77.7%) indicated that they used LLMs for research-related activities. The proportion increased significantly with educational level: 52.4% among undergraduates, 85.5% among professional master’s students, 89.8% among academic master’s students, and 95.5% among doctoral students (p < 0.001).

Further analysis of these 255 students regarding academic research and writing support (**Section 2**) showed no significant differences in overall scores across educational levels ($p = 0.205$), indicating a relatively consistent perception of LLMs’ research support functions among students at different educational levels.

Moreover, the mean scores for individual items ranged from 3.71 to 4.19, indicating that students generally hold positive attitudes toward the use of LLMs in research activities. However, the Friedman test revealed significant differences among items ($p<0.001$). The highest-scoring item was “LLMs help me clarify research ideas and select study methods” ($M = 4.19$), followed by “LLMs provide new research ideas or directions through interaction.” ($M = 4.01$). These findings suggest that LLMs provide greater benefits in idea generation than in actual writing or data analysis tasks (Table 2), primarily supporting medical students during the conceptualization stage of research.

Table 2. Evaluation of LLM Applications in Academic Research and Writing Support Among Medical Students (N = 255)

Evaluation item	Mean	SD
Clarifying research ideas and selecting research method	4.19	0.68
Providing new research ideas or directions	4.01	0.84
Polishing and revising text in academic papers or reports	3.89	0.91
Serving as an efficient tool for literature search and summarization	3.81	1.03
Assisting in data organization, coding, and preliminary analysis	3.75	1.14
Assisting in drafting research-related manuscripts	3.71	1.15

4. Educational Level Differences in LLM Support for Learning, Examinations, and Clinical Practice

Section 3 examined the support provided by LLMs for professional knowledge learning, examinations, and clinical practice. In contrast to Section 2, significant differences across educational levels were observed ($p < 0.001$). Undergraduates had the highest score ($M = 4.05$), followed by professional master’s students ($M = 3.70$) and academic master’s students ($M = 3.70$), while doctoral students had the lowest score ($M = 3.50$).

Furthermore, the Friedman test revealed significant differences among items ($p<0.001$). The item “Simulating doctor–patient communication scenarios” scored significantly lower than all other items ($p < 0.001$) (Table 3). This finding suggests that LLMs have limited effectiveness in highly specialized clinical skill training, whereas they are more widely recognized for their role in knowledge acquisition and broadening perspectives.

Table 3. Evaluation of LLM Applications in Medical Knowledge Acquisition and Clinical Skills Training (N=328)

Evaluation item	Mean	SD
Establishing interdisciplinary knowledge connections and broadening research perspectives	4.13	0.80
Facilitating the comprehension and mastery of medical knowledge	3.96	0.86
Providing substantial support for examinations and clinical practice	3.77	0.99
Assisting in examination preparation by generating practice questions and explaining errors	3.40	1.36
Simulating doctor-patient communication scenarios	3.36	1.04

5. Risk Awareness and Information Evaluation Ability: Increasing with Educational Level

Section 4 assessed students' awareness of LLM limitations and risks from two aspects: the accuracy of LLM-generated content and data privacy, while Section 5 evaluated their information evaluation ability. The results showed clear differences across educational groups, with both risk awareness and information evaluation ability increasing with higher educational levels ($p < 0.001$) (**Table S2**).

The Friedman test revealed no significant differences among individual items, suggesting relatively balanced development across different dimensions of risk awareness and information evaluation ability.

6. Structural Differentiation of Research-Related Pressure

Section 6 assessed research-related pressure among medical students across three dimensions: time and energy pressure, research skill pressure, and research output pressure, revealing a differentiated pattern of research pressure across educational levels (**Table 4**).

Table 4. Comparison of Mean Scores Across Dimensions of Research-Related Pressure Among Medical Students at Different Educational Levels (N=328)

Education	Time and energy pressure	Research skill pressure	Research output pressure	Total stress score
Undergraduate	2.97	4.37	3.19	3.51
Professional Master	3.80	4.38	3.71	3.96
Academic Master	3.20	3.27	2.85	3.10
PhD	3.56	3.12	4.46	3.71
p-value	< 0.001	< 0.001	< 0.001	< 0.001

Time and energy pressure: Professional master's students exhibited the highest score (3.80), significantly higher than undergraduates (2.97, $p < 0.001$) and academic master's students (3.20, $p =$

0.002), whereas doctoral students scored intermediate values (3.56). This pattern aligns with the professional master's training model, in which students must balance clinical rotations with research responsibilities.

Research skill pressure: Undergraduates and professional master's students scored significantly higher than academic master's and doctoral students ($p < 0.001$). Undergraduates likely experience substantial pressure due to limited familiarity with statistical methods and academic writing, whereas professional master's students face skill-related challenges stemming from intensive clinical duties and limited research training. In contrast, academic master's and doctoral students undergo systematic research training and consequently experience relatively lower skill-related pressure.

Research output pressure: Doctoral students exhibited the highest score (4.46), significantly higher than the other three groups ($p < 0.001$). Professional master's students ranked second (3.71), significantly higher than academic master's students (2.85, $p < 0.001$) and undergraduates (3.19, $p = 0.019$). This pattern reflects the substantial publication pressure faced by doctoral students due to graduation and career requirements, whereas professional master's students must complete their thesis within a constrained timeframe.

7. General Benefits of LLMs for Research Output

Section 7 assessed the benefits of LLMs for research output. One-way ANOVA revealed no significant differences among educational groups ($p = 0.564$), indicating that students across different educational levels provided relatively consistent evaluations of LLM effectiveness in supporting research output.

The Friedman test revealed no significant differences among individual item scores ($p = 0.749$), suggesting that LLMs confer relatively balanced benefits across different aspects of research output.

8. Broad Consensus and Rational Consideration of Integrating LLMs into Medical Education

Regarding medical education, 82.6% of students indicated that LLMs should be incorporated to assist teaching, and 88.1% stated that students should receive formal instruction on using LLMs for research. No significant differences were observed across educational levels ($p > 0.05$), reflecting broad consensus among medical students on the integration of LLMs into medical curricula.

Perceptions of educational challenges are summarized in Supplementary **Table S3**. "Inaccurate or erroneous generated content" was the most frequently selected challenge (291 respondents, 88.7%), followed by "Difficulty in verifying information sources" (241 respondents, 73.5%) and "Overreliance" (221 respondents, 67.4%). By contrast, "Privacy and data security concerns" (156 respondents, 47.6%) and "Inability to keep up with the latest research" (119 respondents, 36.3%) received moderate attention, while "Lack of formal training" had the lowest selection rate (102 respondents, 31.1%). Cochran's Q test indicated statistically significant differences in selection proportions among the options ($p < 0.001$), demonstrating that students' concerns regarding different challenges were unevenly distributed. Students prioritized the accuracy and verifiability of LLM-generated content, followed by the

risk of overreliance, whereas external factors, such as insufficient training, attracted comparatively less attention. These findings suggest that current medical students' perceptions of LLM-related challenges primarily focus on content quality rather than on training in LLM use.

Perceptions of integrating LLMs into educational processes are summarized in Supplementary Table S4. "Instructional support" (86.6%) and "Research guidance" (71.0%) received the highest levels of endorsement, followed by "Clinical training" (45.4%) and "Assessment and evaluation" (28.7%). Cochran's Q test revealed highly significant differences among the options ($p < 0.001$), indicating that students consider LLMs more suitable for supporting teaching and research, while adopting a more cautious stance toward their application in clinical training and assessment.

Between-group analyses further confirmed heterogeneity in educational needs. Academic master's and doctoral students reported higher expectations for LLM use in "research guidance" ($p = 0.032$), whereas undergraduates and professional master's students assigned greater importance to "clinical training" ($p = 0.002$). This pattern is highly consistent with the analysis of application scenarios: students at higher educational levels tend to recognize the research value of LLMs, while those at lower levels place greater emphasis on their utility in clinical training.

Discussion

This study demonstrates that LLMs have achieved a 98.2% penetration rate among Chinese medical students, with widespread adoption across all educational levels. This rate substantially exceeds the 28.7% reported in a nationwide survey of Chinese medical students in early 2024 [19] and the 62.9% reported in 2025 [20], as well as contemporaneous rates in Germany (76.2%) [21] and the United States (48.9%–52%) [13,14]. The rapid growth and high adoption may be attributed to the recent development of domestic LLMs, which have lowered barriers to use [15], coupled with persistent academic and research pressures across different training stages. Collectively, these findings suggest that LLM usage among medical students has progressed from an initial exploratory phase to a structurally integrated component of daily learning and research.

In terms of model selection, DeepSeek (45.4%) and Doubao (31.1%) were the most frequently used models, markedly exceeding the proportion of ChatGPT users (15.2%). This pattern indicates a clear preference for locally developed LLMs among Chinese medical students. Such a preference may be attributed to the contextual adaptability and accessibility of domestic models. Compared with Western general-purpose models such as ChatGPT, domestic models tend to align more closely with Chinese medical terminology and clinical practice, while also providing more contextually relevant interpretations of local policies, examination systems, and clinical guidelines [22]. This reduces the need for language translation and makes these models more suitable for use in contexts such as postgraduate entrance examinations, standardized residency training, and medical licensing examinations. In addition, domestic models often provide advantages in network accessibility and lower usage costs [22], which

further reduce barriers to adoption and facilitate broader use among medical students across diverse regions and resource settings.

The study further identified model preference patterns based on educational level. Doctoral students exhibited higher selection rates for ChatGPT (31.8%) and Gemini (15.9%) than other groups, reflecting the differentiated demands associated with advanced medical training. One possible explanation is that doctoral students face stronger pressures for internationally oriented research output. Models such as ChatGPT and Gemini perform well in tasks relevant to this stage, including multilingual SCI manuscript editing and translation, complex information processing, and cross-disciplinary reasoning [2]. These capabilities align more closely with the research-oriented tasks that characterize doctoral training. In contrast, undergraduates more commonly use LLMs for knowledge acquisition and examination preparation, whereas doctoral students are more likely to employ them for research-related tasks. This functional differentiation partly explains the divergence in model selection across educational levels.

Medical education is characterized by a stage-specific training structure [17]. In this study, LLM application scenarios differed significantly across educational levels. Undergraduates primarily used LLMs for examination preparation and support in learning professional knowledge. At this stage, students typically face intensive coursework and frequent assessments. LLMs can provide logically coherent and well-structured explanatory feedback, which may enhance knowledge comprehension and learning efficiency, thereby helping students consolidate foundational knowledge and improve examination performance [23]. Previous research by Christy K. Boscardin et al. in 2024 reported that LLMs can facilitate instant information integration and structured feedback for medical students, supporting self-directed learning, knowledge framework construction, and clinical reasoning training, ultimately contributing to the development of self-regulated learning skills and deeper cognitive processing [2].

In contrast, doctoral and academic master's students primarily use LLMs for research-related activities. The data suggest that LLMs have become integrated into multiple stages of the research process, with the highest scores observed for "clarifying research ideas" and "providing research ideas", rather than for manuscript writing. This pattern indicates that LLMs are increasingly functioning not only as tools for language polishing or text generation, but also as cognitive aids that support problem formulation and the expansion of research thinking. Consistent with this observation, recent reviews report that advanced medical trainees increasingly use LLMs for tasks such as preliminary screening of large bodies of literature, generating complex statistical code (e.g., in R or Python), and exploring multidimensional research hypotheses [24].

Among doctoral students, high research output pressure ($M = 4.46$) may be an important factor associated with frequent LLM use. By helping clarify research ideas, LLMs can shorten the interval between research conception and initial draft development, thereby improving the efficiency of the overall research workflow. As a result, they are increasingly functioning as a support tool in doctoral-level academic research.

Notably, LLMs received relatively low scores for patient–physician communication and real-time clinical decision support ($M = 3.36$) across all educational levels. This finding is consistent with the results reported by Rincón et al. [25]. Although students recognize the capacity of LLMs to process medical knowledge, they remain cautious about their use in clinical contexts characterized by uncertainty and ethical complexity. These perceptions underscore the continuing centrality of human clinical judgment in human–AI collaboration.

This study indicates that risk perception and information appraisal abilities increase with educational level. Previous studies have noted that LLMs are associated with risks such as hallucinated outputs [9,10], data privacy concerns [11], and potential academic misconduct [12]. Doctoral students demonstrated the highest levels of risk perception ($M = 4.55$) and information appraisal ability ($M = 4.30$), exceeding those of undergraduates and professional master’s students. This heightened awareness may be associated with the systematic research training and disciplinary knowledge accumulated during doctoral education. In contrast, the relatively low risk perception observed among undergraduates ($M = 2.92$) suggests that using LLMs without sufficient domain knowledge or critical evaluation may introduce potential risks. With limited capacity for independent information appraisal, early-stage students may be more susceptible to overreliance on AI-generated outputs, which could hinder the development of independent thinking and clinical reasoning skills. Within the context of medical education reform, integrating AI literacy education and structured training on LLM use into curricula—emphasizing risk identification, information verification, and ethical responsibility—may therefore be warranted. Such measures can support more effective use of LLMs while contributing to the development of independent judgment and professional responsibility among future physicians.

This study found that research pressure exhibits distinct structural differences across educational levels. Among undergraduates and professional master’s students, the primary sources of pressure arise from limited research competencies, including insufficient training in research design, statistical analysis, and academic writing. In these groups, anxiety is therefore primarily manifested as skill-related stress. In addition, professional master’s students reported the highest level of time and effort pressure ($M = 3.80$), which is closely associated with the stage-specific structure of medical education. These students are typically required to complete intensive clinical rotations while simultaneously meeting academic output requirements within a limited timeframe [17]. In this context, LLMs possess strong capabilities in knowledge retrieval and content generation [2], which may help mitigate skill-related uncertainties and enhance the efficiency of completing research tasks.

In contrast, doctoral students experience the highest pressure from research output and achievement expectations ($M = 4.46$), significantly exceeding that of other groups. Within a training system that prioritizes publication quantity and quality as core evaluation metrics, the combination of publication requirements, graduation standards, and employment competition positions research output as the primary source of anxiety [25]. While LLMs provide cognitive support and enhance task completion efficiency, their effect is confined to the tool and skill level and cannot fundamentally alleviate performance pressure imposed by the evaluation system.

The data indicate that LLMs are more effective at mitigating skill-related pressure than at reducing output-related pressure. Despite high frequency and extensive use of LLMs among doctoral students across the entire research process, overall output pressure remains at peak levels. This suggests that when pressure is rooted in institutionalized academic evaluation systems, improvements in technological efficiency alone cannot fundamentally alleviate output-related anxiety. If evaluation standards remain unchanged, widespread adoption of technology, such as the 98.2% usage rate observed in this study, may fail to reduce pressure and could even intensify competition, as all researchers simultaneously enhance output efficiency. These findings highlight that discussions of AI-enabled medical research should also address the structural effects of evaluation systems to prevent technological tools from inadvertently becoming catalysts for heightened academic competition.

In medical education, 82.6% of students believe that LLMs should be integrated into teaching, and 88.1% believe that students should be trained to use AI for research. Support rates did not differ significantly across educational levels, indicating broad consensus. This suggests that medical education has reached a critical stage in which proactive adoption of technological advancements is essential. As emphasized by Faiza Alam et al., given the rapid development of generative AI, medical education should actively integrate relevant technologies to systematically cultivate future physicians' critical thinking and ethical awareness [11]. However, despite strong student receptivity, the absence of systematic course design and clear guidelines may lead to overreliance and blind trust during self-directed exploration [26]. Therefore, medical curriculum reform should incorporate LLMs as teaching aids, for example by generating case scenarios and summarizing assessment feedback [27], while also embedding instruction on the proper use of AI as a core component to balance technological empowerment with risk management.

Further subgroup analysis indicated that undergraduates and professional master's students are more likely to integrate LLMs into clinical training, whereas academic master's and doctoral students prioritize their use in research guidance. These findings suggest that medical education reform should avoid a uniform approach and instead implement tiered, stage-specific designs. For undergraduates and professional master's students, reforms should emphasize clinical skill development. Medical schools can employ LLM-based virtual standardized patients to provide a safe clinical simulation environment that enhances clinical reasoning. Early integration of AI literacy and standardized usage training is also essential to prevent overreliance, data misuse, or ethical violations.

For academic master's and doctoral students, the educational emphasis should focus on AI-driven research guidance. Institutions can instruct students on leveraging models to generate research ideas, identify innovative directions, and assist in programming complex code and managing bioinformatics data. Such training can substantially improve research efficiency and reduce project timelines. Crucially, clear distinctions between legitimate academic assistance and academic misconduct must be established to ensure that students uphold research integrity while enhancing productivity.

This study provides systematic insights but has several limitations. First, it is a single-center, cross-sectional survey conducted at China Medical University. Although the sample includes students from undergraduate to doctoral levels, the findings may not be generalizable to medical schools nationwide. Second, data were collected through student self-reports, which may introduce subjective bias, and the doctoral sample size (N=45) is relatively small, potentially limiting the robustness of some between-group comparisons. Third, LLM technology evolves rapidly; despite including domestic models such as DeepSeek, model preferences may change quickly with software updates. Finally, this study focuses on subjective perceptions and lacks objective quantitative assessments of students' actual academic performance or research output following LLM use.

Conclusion

This study demonstrates that LLMs are extensively adopted in medical education, with usage patterns varying across different training stages. Risk awareness and information evaluation capabilities increase with educational level. While LLMs can mitigate skill-related pressure, they are less effective in reducing output pressure imposed by institutional evaluation systems. To maximize the benefits of AI, medical education should actively integrate LLMs, establish a standardized AI literacy curriculum, define clear boundaries and ethical guidelines, and implement stage-specific, layered reforms. Such a structured approach is essential for AI to meaningfully enhance the quality of medical education.

Abbreviations

LLM

Large Language Models

ANOVA

One-way Analysis of Variance

AI

artificial intelligence

Declarations

Ethical Approval and consent to participate:

The study was approved by the Ethics Committee of the First Affiliated Hospital of China Medical University, Liaoning, China. Informed consent to participate was obtained from all participants before completing the questionnaire. All procedures involving human participants were conducted in accordance with the ethical standards of the institutional and/or national research committee and with the principles of the Declaration of Helsinki.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Funding

This study has received funding by China Postdoctoral Science Foundation (Certificate Number: 2025M771979).

Author Contribution

Conceptualization, C.Z. and S.L.; Methodology, Z.W.; Data collection, Z.W. and L.X. ; Formal analysis, Z.W.; Writing—original draft preparation, Z.W.; Writing—review and editing, C.Z., S.L. and G.Z.; Funding acquisition, C.Z. All authors have read and agreed to the published version of the manuscript.

Acknowledgements

Not applicable.

Data Availability

Data generated or analyzed during the study are available from the corresponding author by request.

References

1. Liu J, Wang C, Liu S. Utility of ChatGPT in Clinical Practice. *J Med Internet Res.* 2023;25:e48568.
2. Boscardin CK, Gin B, Golde PB, Hauer KE. ChatGPT and Generative Artificial Intelligence for Medical Education: Potential Impact and Opportunity. *Acad Med.* 2024;99(1):22–7.
3. Wijnen-Meijer M, Gartmeier M, Berberat PO. Übersicht über die Forschung im Bereich der medizinischen Ausbildung [Overview on research in the field of medical education]. *HNO.* 2020;68(4):231–7.
4. Huang J, Chen-Chuan K. Chang. Towards reasoning in large language models: A survey. Findings of the association for computational linguistics: ACL 2023. 2023.
5. Cheng Y, Zhu L. A review of ChatGPT in medical education: exploring advantages and limitations. *Int J Surg.* 2025;111(7):4586–602.
6. Safranek CW, Sidamon-Eristoff AE, Gilson A, Chartash D. The Role of Large Language Models in Medical Education: Applications and Implications. *JMIR Med Educ.* 2023;9:e50945.

7. Holderried F, Stegemann-Philipps C, Herrmann-Werner A, Festl-Wietek T, Holderried M, Eickhoff C, Mahling M. A Language Model-Powered Simulated Patient With Automated Feedback for History Taking: Prospective Study. *JMIR Med Educ.* 2024;10:e59213.
8. Kim TW. Application of artificial intelligence chatbots, including ChatGPT, in education, scholarly work, programming, and content generation and its prospects: a narrative review. *J Educ Eval Health Prof.* 2023;20:38.
9. Alkaisi H, McFarlane SI. Artificial hallucinations in ChatGPT: implications in scientific writing. *Cureus.* 2023;15(2):e35179.
10. Jiawen Deng A, Zubair Y-J, Park. Limitations of large language models in medical applications. *Postgrad Med J.* December 2023;99:1298–9.
11. Alam F, Lim MA, Zulkipli IN. Integrating AI in medical education: embracing ethical usage and critical understanding. *Front Med (Lausanne).* 2023;10:1279707.
12. Sahu PK, Benjamin LA, Singh Aswal G, Williams-Persad A. ChatGPT in research and health professions education: challenges, opportunities, and future directions. *Postgrad Med J.* 2023;100(1179):50–5.
13. Ganjavi C, Eppler M, O'Brien D, Ramacciotti LS, Ghauri MS, Anderson I, Choi J, Dwyer D, Stephens C, Shi V, Ebert M, Derby M, Yazdi B, Cacciamani GE. ChatGPT and large language models (LLMs) awareness and use. A prospective cross-sectional survey of U.S. medical students. *PLOS Digit Health.* 2024;3(9):e0000596.
14. Zhang JS, Yoon C, Williams DKA, Pinkas A. Exploring the Usage of ChatGPT Among Medical Students in the United States. *J Med Educ Curric Dev.* 2024;11:23821205241264695.
15. Reflections on DeepSeek's breakthrough. *Natl Sci Rev.* 2025;12(3):nwaf044. 16. Gibney E. China's cheap, open AI model DeepSeek thrills scientists. *Nature.* 2025;638(8049):13–14.
16. Wijnen-Meijer M, van den Broek S, Koens F, et al. Vertical integration in medical education: the broader perspective. *BMC Med Educ.* 2020;20:509.
17. Ahsan Z. Integrating artificial intelligence into medical education: a narrative systematic review of current applications, challenges, and future directions. *BMC Med Educ.* 2025;25(1):1187.
18. Tao W, Yang J, Qu X. Utilization of, Perceptions on, and Intention to Use AI Chatbots Among Medical Students in China: National Cross-Sectional Study. *JMIR Med Educ.* 2024;10:e57132.
19. Hu N, Jiang XQ, Wang YD, Kang YM, Xia Z, Chen HH, Duan SN, Chen DX. Status and perceptions of ChatGPT utilization among medical students: a survey-based study. *BMC Med Educ.* 2025;25(1):831.
20. Fußhöller A, Lechner F, Schlicker N, Muehlensiepen F, Mayr A, Kuhn S, Hirsch MC, Knitza J. Perceptions, Usage, and Educational Impact of ChatGPT Among Medical Students in Germany: Cross-Sectional Mixed Methods Survey. *JMIR Form Res.* 2025;9:e81484.
21. Peng Y, Malin BA, Rousseau JF, Wang Y, Xu Z, Xu X, Weng C, Bian J. From GPT to DeepSeek: Significant gaps remain in realizing AI in healthcare. *J Biomed Inf.* 2025;163:104791.

22. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, Madriaga M, Aggabao R, Diaz-Candido G, Maningo J, Tseng V. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. PLOS Digit Health. 2023;2(2):e0000198.
23. Sallam M. ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. Healthc (Basel). 2023;11(6):887.
24. Rincón EHH, Jimenez D, Aguilar LAC, et al. Mapping the use of artificial intelligence in medical education: a scoping review. BMC Med Educ. 2025;25:526.
25. Benítez TM, Xu Y, Boudreau JD, Kow AWC, Bello F, Van Phuoc L, Wang X, Sun X, Leung GK, Lan Y, Wang Y, Cheng D, Tham YC, Wong TY, Chung KC. Harnessing the potential of large language models in medical education: promise and pitfalls. J Am Med Inf Assoc. 2024;31(3):776–83.
26. Rodman A, Mark NM, Artino AR Jr, Lessing JN. Using Generative Artificial Intelligence in Medical Education. Acad Med. 2025;100(2):250.

Figures

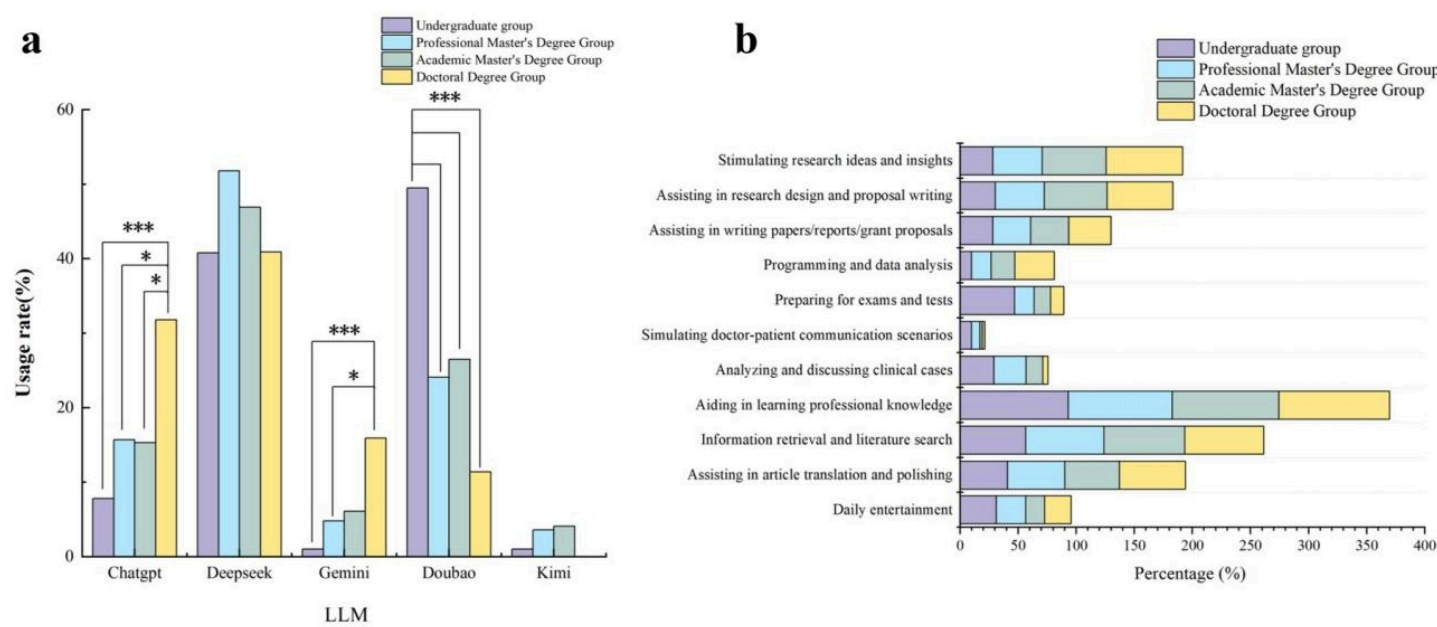


Figure 1

a) Distribution of LLM Model Preferences Among Medical Students of Different Educational Levels. b) LLM Application Scenarios among Medical Students across Different Educational Levels

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryDocument.docx](#)