

The use of GenAI in medical education in China and Indonesia: A comparative literature review

Prattama Santoso Utomo¹ and Jialin Yan²

¹MD, MHPE, Assistant Professor, Department of Medical Education and Bioethics, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia

²Med, PhD Candidate, Warner School of Education and Human Development, University of Rochester, Rochester, USA

Abstract

Background: The integration of artificial intelligence, particularly generative AI (GenAI), in learning is transforming medical education from undergraduate to subspecialty levels, promising enhanced diagnostic accuracy, personalized learning, and real-time feedback. However, concerns persist regarding reliability, accuracy, and ethical implications, including the potential for "hallucinations" and an erosion of students' critical thinking skills. The evolving regulatory landscape also highlights challenges in ensuring safety, fairness, and accountability as GenAI systems gain autonomy in clinical decision-making. This narrative literature review examines the current landscape of GenAI applications in medical education within China and Indonesia, aiming to identify trends, gaps, and strategies for developing GenAI competencies among future healthcare professionals. **Methods:** A systematic search of PubMed, MedLine, Google Scholar, Scopus, and Chinese CNKI databases, along with targeted hand searches, identified 12 empirical studies published between January 2022 and December 2024. Studies were categorized based on Kirkpatrick's evaluation levels. **Results:** We found that GenAI has been used across various medical topics and learning activities, including scientific presentations, small-

group discussions, and simulations. Benefits included increased confidence in GenAI use, improved student engagement, and enhanced practical skill development. Conversely, challenges included sporadic adoption, lack of training, concerns about misinformation, technical limitations (e.g., accuracy in non-English contexts), and the high cost of implementation. Most evaluated outcomes were at Kirkpatrick's Level 1 (satisfaction) and Level 2 (knowledge/skills), with a notable absence of Level 3 (behavioral changes) or Level 4 (healthcare outcomes), suggesting the field is in early evaluative stages. **Discussion and Conclusion:** Our review recommends comprehensive training for faculty and students, curriculum integration, and robust evaluation systems to address accuracy and ethical concerns. Future research should focus on longer-term impacts on behavioral changes and patient outcomes, utilizing multi-methodological approaches and fostering interdisciplinary collaboration. This will ensure GenAI effectively complements, rather than replaces, human expertise in preparing competent and ethical healthcare professionals.

Keywords: artificial intelligence (AI), generative AI, medical education, Indonesia, China

Date submitted: 26-July-2025

Email: Prattama Santoso Utomo (prattama.santoso.utomo@ugm.ac.id)

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-Non Commercial-Share Alike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

Background

The integration of artificial intelligence (AI) in medical education is reshaping how students acquire clinical knowledge and develop practical skills at various levels, such as undergraduate, postgraduate, specialty, and subspecialty education. AI-driven tools, particularly generative AI (GenAI) have been increasingly incorporated into medical training to enhance diagnostic accuracy, improve

Citation: Utomo PS and Yan J. The use of GenAI in medical education in China and Indonesia: A comparative literature review. *Educ Health* 2025;38:347-360

Online access: www.educationforhealthjournal.org

DOI: 10.62694/efh.2025.404

Published by The Network: Towards Unity for Health

personalized learning, solve healthcare problems, increase patient safety, enhance effective communication, and provide real-time feedback on clinical decision-making.¹⁻³

However, the application of GenAI within higher education contexts, including medical education, introduces a more complex dynamic. While GenAI-driven chatbots have been shown to support learners

by facilitating problem-solving and providing supplemental learning opportunities outside the classroom,^{4,5} there are emerging concerns regarding the epistemological implications of such tools. Kwong et al.⁶ reported a potential global “feedback loop,” wherein dependence on GenAI may reinforce dominant knowledge structures at the expense of transformative, inquiry-based learning. This concern is particularly relevant in health education, where accurate, evidence-based knowledge and the ability to engage in critical thinking are essential.

Despite growing enthusiasm surrounding GenAI, both educators and students remain skeptical about its reliability and educational impact. Health professions educators have raised concerns about the accuracy of GenAI-generated content, particularly when outputs fail to reflect current clinical guidelines or lack domain-specific rigor.^{7,8} From the learner’s perspective, anxieties tend to focus on the risk of over-reliance, with many students expressing concern that habitual dependence on GenAI may erode their critical thinking ability.⁹ Furthermore, both groups highlighted ethical issues, such as plagiarism and the difficulty of verifying GenAI-generated content, as persistent challenges.⁷

Additionally, a major technical concern is the phenomenon of “hallucination” in large language models (LLMs), wherein the model produces seemingly plausible but factually incorrect or fabricated content,¹⁰ leading users to trust incorrect responses. Many models that are researched are never deployed in practice, and those that are implemented often lack rigorous evaluation.¹¹ GenAI-driven demand forecasting tools, while promising, also need to adapt to fluctuations in market dynamics and public health trends to maintain relevance.^{12,13} Moreover, even minor GenAI-related errors in clinical contexts can result in significant consequences, including incorrect medical prescriptions and medical malpractice.^{2,14}

Finally, the governance and regulatory landscape for GenAI in the healthcare education field is still evolving. Issues such as ensuring safety, fairness, transparency, and accountability continue to pose significant challenges.¹⁵ As GenAI systems begin to make more autonomous decisions, regulatory bodies must establish robust validation frameworks, mitigate algorithmic bias, and uphold data privacy standards. A particularly pressing concern is assigning responsibility when AI-generated recommendations lead to adverse outcomes. As

Rajpurkar et al.¹⁶ caution, clinicians may become over-reliant on GenAI systems, potentially leading to a decline in their own diagnostic ability and decision-making autonomy.

Although certain literature reviews, systematic reviews and meta-analyses have examined the application of artificial intelligence (AI), particularly generative AI (GenAI), in medical diagnostics, such as radiological image analysis, predictive modelling, and clinical decision support, there is a lack of comprehensive reviews that address how GenAI can be effectively utilized to prepare future healthcare professionals for ethical, competent, and context-sensitive use of AI tools in clinical settings. Therefore, this literature review aims to map the current landscape of GenAI applications in medical education across two highly populated Asian countries: China and Indonesia. The review seeks to identify trends, gaps, pedagogical models, and institutional strategies that support the development of GenAI competencies among medical students and trainees.

Methods

We conducted a literature review using a narrative review approach,¹⁷ but used a systematic search, borrowing the PRISMA-P’s literature search guideline. The narrative review elucidates a comparison of what has been studied on how GenAI was used in medical education in the Indonesian and Chinese medical/health science education institutions. This literature review focused on synthesizing effective practices and a critical evaluation of the use of GenAI in medical education.

Literature Search

The literature search was conducted in major and common healthcare and education databases, for instance, PubMed, MedLine, GoogleScholar, Scopus, and Chinese CNKI. We also conducted a targeted hand search on some key medical/health professions education journals, backtracking from the included articles’ reference lists. We used the following Boolean operation for the systematic literature keyword, with some adjustments to each database’s keyword styles:

- Indonesian studies: ("artificial intelligence" OR AI OR GenAI) AND ("medical education" OR "health professions education") AND Indonesia*
- Chinese studies: ("artificial intelligence" OR AI OR GenAI) AND ("medical education" OR "medical training") in China*

Inclusion and Exclusion Criteria or Article Selection

We included empirical studies from China and Indonesia, published between January 2022 and December 2024 (i.e., OpenAI was launched in 2022, so our literature search starts in this year). We included studies in English, Chinese, and Bahasa Indonesia. As this review aimed to present effective practices in GenAI use for medical education, we only included empirical studies that report how GenAI was used in the institution and the evaluation results of the respective innovation or program. Non-empirical studies (e.g., case studies, innovation/perspective papers, editorials, grey papers), articles focusing on the use of GenAI for clinical diagnosis without training purposes, and papers without evaluation results were excluded from this literature review. We applied the PRISMA-P guideline for article screening and selection (see Figure 1) and extracted and synthesized the included literature based on a narrative review approach.

Data Extraction and Synthesis

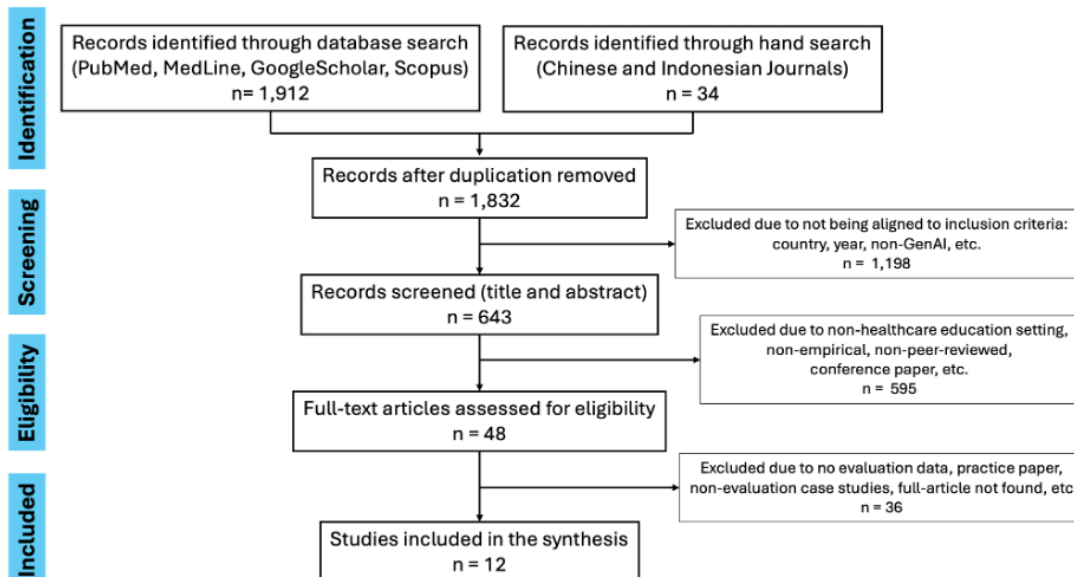
We extracted characteristics of the included studies, for instance, year, country of the study, study aims, design of the study, and data collection and evaluation approaches. Studies were classified based on Kirkpatrick’s evaluation levels to stratify the level of impact of the included studies, e.g., Level 1 – Reaction and Satisfaction, Level 2 – Learning, Level 3 – Behavioral Changes, and Level

4 – Healthcare Outcomes Changes.¹⁸ To answer the review questions, we extracted and synthesized findings of the study, categorized and reported GenAI usage in the study, effective aspects of GenAI use, ineffective aspects/challenges of using GenAI, and also suggestions for future improvements.

Results

The systematic database literature search retrieved 1,912 records, and the targeted hand-search retrieved 34 additional records. After removing duplicates, a total of 1,832 records were set for the screening phase. In the screening phase, 1,198 articles were excluded due to not aligning to the inclusion criteria (e.g., studies not from Indonesia and China, studies before 2022 when GenAI launched, articles reporting the use of AI for diagnosis but not for training education purposes, and other reasons), and 643 articles remained for title and abstract screening. A total of 595 articles were excluded after title and abstract screening due to various reasons (e.g., non-healthcare education setting, non-empirical studies, non-peer-reviewed articles, conference papers, etc.), leaving 48 articles for full-text eligibility screening. Subsequently, we excluded 36 articles after full-text screening due to a range of reasons (e.g., no evaluation data, practice papers, case presentation without evaluation data, full article not found, etc), resulting in a total of 12 studies included in the synthesis of this literature review (Figure 1).

Figure 1. Article Screening and Selection Flowchart



Study Characteristics

Table 1 depicts characteristics of the included studies, which consisted of 9 studies from China,^{19–27} and 3 studies from Indonesia.^{28–30} Across these two countries, GenAI was applied in diverse ways: in Indonesia, for scientific presentations³⁰ and small-group discussions;²⁸ in China, for case-based learning, simulations, and imaging tasks,²⁶ for curricular integration, survey-based evaluation, or interprofessional training.^{19,21} This diversity illustrates how AI has been embedded into varied educational fields and learning activities.

All Indonesian studies were cross-sectional, while approximately half of the included Chinese studies were experimental or quasi-experimental. The context of the included studies ranged from undergraduate medical education, nursing education, residency training, and interprofessional education. All included studies were conducted by students or trainees from different levels (undergraduate and residency), with only one study from Indonesia including faculty/consultants as participants.

The Chinese studies reflect a range of quantitative and experimental designs, each with distinct limitations. Wang²⁰ and Wang et al.²¹ both employed cross-sectional survey designs. However, Wang²⁰ included only 105 participants, which limits generalizability, compared with 605 participants in Wang et al.²¹ Yan et al.²⁶ conducted an experimental case-comparison study, but the small sample of 40 participants restricts the strength of its conclusions. Moreover, participants vary considerably across studies: some focus on undergraduate medical students,²² while others involve medical imaging interns.²⁷

The Indonesian studies add an important regional perspective, presenting readiness, perceptions, and curricular integration of GenAI in medical education. Pudjiadi and colleagues³⁰ surveyed a large sample of pediatric teaching staff (n = 196) and residents (n = 728) across 15 public universities. However, as a self-report survey, the findings are constrained by subjective perceptions and the lack of observed practice change. Subsequently, Busch and colleagues²⁸ conducted a global-scope cross-sectional survey of 4,596 medical, dental, and veterinary students across 48 countries, including 122 from Indonesia. While the global comparative lens is a strength, the relatively

small Indonesian subsample limits representativeness.

In terms of Kirkpatrick's level of impact evaluation, Indonesian studies were mostly at level 1 (satisfaction) and 2a (perceived increase in knowledge and skills), whereas most Chinese studies were higher, at level 2b (objectively measured increase in knowledge and skills).

Field and Learning Activities of GenAI Use

Included studies utilized GenAI for a range of medical and health professions education and training at different topic areas and instructional activities. GenAI has been used in undergraduate medical education in both Indonesia and China, and used to facilitate a range of topics, such as readiness in using GenAI in education, medical imaging, oncology, dietary patient education, and clinical decision-making training (see Table 1).

Studies in Indonesia showed the use of GenAI in scientific presentations in a pediatric residency program,³⁰ and small-group discussions in biomedical and clinical learning modules in an undergraduate medical training program.²⁹ A study also reported that GenAI was used integrally in the curriculum of various medical, dental, and veterinary education programs, but did not specify which learning activities were used in each program, although the overall evaluation results showed good satisfaction with the integration of GenAI into the programs.²⁸

In Chinese studies, GenAI tools such as ChatGPT and AlphaFold can enhance the visualization of complex biomedical content, support interactive case-based learning, and provide personalized feedback, thereby promoting student engagement and deepening conceptual understanding.^{22,27} In vocational and undergraduate settings, GenAI-powered simulations have been shown to improve practical skill acquisition, clinical reasoning, and diagnostic accuracy in disciplines such as radiology and oncology.^{19,26} Moreover, the use of GenAI to generate individualized learning paths and real-time performance analytics can help standardize education quality across diverse institutions and learner profiles.²⁰

Benefits of Using GenAI

In Indonesian studies, the implementation of GenAI into educational programs has increased faculty and trainees' confidence and readiness in using GenAI

Table 1: Characteristics of Included Chinese and Indonesian Studies (N = 12)

Characteristics	Chinese Studies	Indonesian Studies
n (%)	9 (75%)	3 (25%)
Design	<ul style="list-style-type: none"> • 5 cross-sectional¹⁹⁻²³ • 4 experimental/quasi-experimental²⁴⁻²⁷ 	All cross-sectional ²⁸⁻³⁰
Context	<ul style="list-style-type: none"> • Interprofessional education^{21,24} • Nursing education²⁵ • Medical imaging training and residency programs^{26,27} • Undergraduate medical education^{19,20,22,23} 	<ul style="list-style-type: none"> • Pediatric residency³⁰ • Interprofessional education²⁸ • Undergraduate medical education²⁹
Participants	<ul style="list-style-type: none"> • Medical students¹⁹⁻²³ • Nursing students^{24,25} • Medical residents^{26,27} 	<ul style="list-style-type: none"> • Medical students³⁰ • Dentistry students²⁸ • Veterinary students²⁸ • Medical residents³⁰ • Faculty members/teaching staffs³⁰
GenAI application	<ul style="list-style-type: none"> • Readiness of GenAI application in education²⁵ • National licensing examination accuracy/test preparation²³ • Virtual simulation training in medical imaging²⁶ and oncology training¹⁹ • Team-based learning/case-based learning/problem-based learning in dietary education²⁴, medical imaging²⁷ and medical oncology¹⁹ • Clinical decision making training in medical oncology¹⁹ • Overall learning course in microbiology²² and medical terminology²⁵ 	<ul style="list-style-type: none"> • Readiness of GenAI application in undergraduate medical/health education^{28,29} • Integration of GenAI in the pediatric residency curriculum³⁰
Data collection	<ul style="list-style-type: none"> • Survey (quantitative and qualitative) • Tests, accuracy/reliability analysis • Log and student feedback 	All surveys
Highest Kirkpatrick ¹ level of evaluation	<ul style="list-style-type: none"> • Level 1^{20,21,23} • Level 2a¹⁹ • Level 2b^{22,24-27} 	<ul style="list-style-type: none"> • Level 1^{28,30} • Level 2a²⁹

Note. Kirkpatrick's levels of evaluation: 1 (satisfaction), 2a (perceived increase in knowledge or skills), 2b (objectively measured increase in knowledge or skills)

in teaching and future clinical practice.^{29,30} Integration of GenAI in the curriculum also improved faculty and students' attitudes towards using GenAI, from rejecting to accepting, with more awareness of potential ethical issues and considerations.²⁸ Moreover, GenAI use in medical education has been perceived as positively contributing towards students' learning abilities,²⁹ and may assist in the accuracy of implementing knowledge and skills in providing patient care.³⁰

Learners also reported that the introduction of GenAI during training can validate any adverse GenAI use displayed in social media and on the internet.²⁸

Across various contexts in China, GenAI-enhanced instruction has consistently improved student engagement, learning satisfaction, and practical skill development, particularly in complex subjects such as medical imaging, oncology, and

Table 2: Comparison of GenAI Challenges, Opportunities, and Future Potentials in Chinese and Indonesian Studies

Countries	Challenges	Opportunities	Future Potentials of GenAI in Medical Education
China	<ol style="list-style-type: none"> Underperformance in licensing exams: Large language models (e.g., ChatGPT) showed weak results in high-stakes medical licensing exams.^{23,26} Critical knowledge gaps: Deficits are particularly pronounced in legal/regulatory knowledge, numerical calculations, dosage computations, and case-based clinical reasoning.^{23,26} Language limitations: ChatGPT's English-dominant training data reduced accuracy in Chinese-language medical contexts.²² Risk of over-reliance: Students relying too heavily on GenAI outputs may weaken their critical thinking skills.^{19,27} Simulation imbalance: In simulation-heavy environments, learners risk insufficient exposure to real clinical cases, limiting authentic diagnostic and interpersonal skill development.^{19,27} High infrastructure costs: Significant upfront investment is required to implement AI-based systems.^{20,21} Faculty shortages: Lack of trained instructors hinders effective integration.^{20,21} Systemic barriers: Absence of standardized curricula and low programming literacy among students constrain scalability and sustainability.^{20,21} 	<ol style="list-style-type: none"> Improved engagement and satisfaction: AI-enhanced instruction consistently boosted student engagement, learning satisfaction, and practical skill development, especially in challenging areas like medical imaging, oncology, and microbiology.^{19,22,26} Strengthened conceptual understanding: Tools such as ChatGPT, AlphaFold, and smart simulation platforms deepened students' grasp of complex biomedical concepts and improved critical thinking.^{19,27} Personalized learning experiences: GenAI applications provided individualized feedback and adaptive learning pathways, tailoring instruction to diverse student needs.^{19,27} Resource efficiency: In vocational and clinical settings, GenAI reduced reliance on costly equipment by offering scalable virtual training environments.²⁶ High student demand: Surveys showed strong interest among medical students, with more than 90% expressing a desire to learn GenAI applications relevant to healthcare.^{20,21} 	<ol style="list-style-type: none"> Development of intelligent tutoring systems that adapt to individual learning trajectories, offering tailored content, formative assessments, and continuous feedback.^{22,23} Refinement of tools like ChatGPT as virtual teaching assistants to simulate patient interactions, answer clinical queries, and support multilingual education.^{22,23} Integration into competency-based education frameworks to enable real-time performance tracking, responsive remediation, and progression pathways.^{21,27} Use of GenAI to automate and support faculty workload, such as drafting case studies, generating formative exam items, and providing engagement analytics.^{21,27}
Indonesia	<ol style="list-style-type: none"> Sporadic and interest-based adoption: Use of GenAI has been inconsistent, often depending on individual interest rather than structured integration, likely due to limited training opportunities and lack of knowledge.^{28,30} Concerns about misinformation: Students expressed worries that GenAI might generate false or misleading information, raising risks for patient care.²⁸ Ethical considerations: Learners highlighted potential ethical issues related to the responsible use of GenAI in clinical contexts.²⁸ Limitations in communication training: Current GenAI applications were found to be inadequate for teaching doctor-patient communication, particularly in providing emotional support, which remains a critical gap.²⁹ 	<ol style="list-style-type: none"> Increased confidence and readiness: Faculty and trainees reported greater preparedness to use GenAI in both teaching and future clinical practice.^{29,30} Improved attitudes toward GenAI: Integration into the curriculum shifted perceptions from rejection to acceptance, with stronger recognition of ethical issues and considerations.²⁸ Enhanced student learning abilities: Students perceived GenAI as positively supporting their ability to learn and apply complex concepts.²⁹ Support for patient care accuracy: GenAI was seen as assisting learners in applying knowledge and skills more accurately when providing patient care.³⁰ Validation of ethical concerns: Exposure to GenAI in training helped learners critically evaluate and validate concerns about its misuse, especially as portrayed in social media and online spaces.²⁸ 	<ol style="list-style-type: none"> Provide adequate training for faculty, staff, and students to ensure appropriate GenAI use and maximize benefits for learning, productivity, and patient care.²⁸⁻³⁰ Curriculum integration is needed to improve long-term acceptance and skills development.^{29,30} GenAI should complement, not replace, the human component in medical education and patient care, supporting but not substituting educators and clinicians.³⁰

microbiology.^{19,22,26} GenAI tools like ChatGPT, AlphaFold, and smart simulation platforms have been shown to improve understanding of concepts, critical thinking, and personalized learning experiences.^{19,27} In vocational and clinical education settings, GenAI has alleviated resource constraints by reducing reliance on expensive equipment and providing scalable virtual training environments.²⁶ Moreover, studies indicate a strong demand and positive perception of GenAI among medical students, with over 90% expressing interest in learning GenAI applications relevant to healthcare.^{20,21} Table 2 below highlights the comparison of benefits/opportunities, challenges, and future potentials of GenAI use in medical education based on the included Indonesian and Chinese literature.

Challenges of GenAI Use

Most studies in Indonesia reported that the use of GenAI in medical education has been sporadic and interest-based, which might be due to a lack of training opportunities and little knowledge on GenAI use.^{28,30} Students are also concerned with the potential of GenAI providing false information, and ethical issues in using GenAI for patient care.²⁸ Furthermore, Lugito and colleagues²⁹ also suggested that the current state of GenAI still possesses major limitations to educate learners in doctor-patient communication, given that emotional support has not been a strength of the current GenAI applications (see Table 2).

In China, large language models/GenAI (i.e., ChatGPT) have demonstrated underwhelming performance on high-stakes medical licensing examinations.^{23,26} These deficits are particularly acute in domains involving legal and regulatory knowledge, numerical calculations, dosage computations, and case-based clinical reasoning. Moreover, ChatGPT's predominantly English-language training corpus has contributed to lower accuracy in Chinese-language medical contexts.²² Beyond technical limitations, pedagogical concerns arise around the risk of reduced critical thinking due to students' potential over-reliance on GenAI-generated outputs. For example, in simulation-heavy learning environments, students often lack sufficient exposure to real-life clinical cases, compromising the development of authentic diagnostic and interpersonal skills.^{19,27} The high initial cost of implementing GenAI-based infrastructure, coupled with a shortage of trained faculty, further impedes effective integration. Systemically, the absence of standardized GenAI

curricula and widespread student deficits in programming literacy also constrain the scalability and sustainability of GenAI-enhanced instruction (see Table 2).^{20,21}

Future Potentials of GenAI

Studies in Indonesia suggested that medical education institutions provide adequate training for faculty, staff, and students in using GenAI appropriately to boost the potential benefits of using GenAI to improve learning, productivity, and patient care.²⁸⁻³⁰ GenAI use should be incorporated into the curriculum to ensure improvement of acceptance and skills in the future.^{29,30} However, despite the promising potential of GenAI use, Pudjiadi and colleagues³⁰ emphasized that GenAI could not replace the human component in medical education and patient care, as each system should complement and assist the other (see Table 2).

As GenAI technologies continue to evolve in China, their integration could enable intelligent tutoring systems that adapt to individual learning trajectories, offering tailored content, formative assessments, and feedback loops that reinforce student progress (see Table 2). Tools like ChatGPT could be refined to serve as virtual teaching assistants, capable of simulating patient interactions, answering clinical queries, and supporting multilingual education, especially if future models overcome current limitations related to language bias, legal reasoning, and quantitative problem-solving.^{22,23} Integration of GenAI into competency-based education frameworks may also allow for real-time performance tracking, enabling more responsive remediation and progression pathways. Furthermore, GenAI tools could be leveraged to automate and enhance faculty workload, such as drafting case studies, generating formative exam items, and providing analytics on student engagement.^{21,27}

Discussion and Implications

Our literature review found that GenAI has been used across various learning topics, indicating the versatility of GenAI applications in medical education. The finding is highly relevant to the nature of GenAI, which can be prompted to adapt to certain scenarios or play the role of indicated health professionals or patients.³¹ In this sense, GenAI has the ability to simulate the real condition, which may, in terms of simulation-based learning design, have commendable fidelity to mimic the real condition.³²

The integration of GenAI into medical education presents transformative implications for both students and faculty, for instance, for a more precise assessment of procedural skill acquisition.³³ Additionally, GenAI plays a role in enhancing the quality of teaching feedback by automating evaluation metrics and identifying instructional gaps.³⁴ Faculty have also begun experimenting with generative GenAI tools like ChatGPT to create clinical scenarios, review assessments, and develop examination content, particularly in bedside teaching.³⁵ Nevertheless, medical schools should be aware that GenAI might provide inaccurate responses in some instances, particularly when the prompts are ambiguous, as reported in a study in nursing education.³⁶ When this erroneous condition occurs, the fidelity of the GenAI may be compromised.

In addition, this literature synthesis also indicated that Indonesian and Chinese literature found different advantages in implementing GenAI in medical education. Indonesian literature mostly suggested that GenAI has assisted trainees and faculty in changing their attitudes towards GenAI use, and also to improve learning ability; while Chinese literature pinpointed improvements in student engagement and critical thinking in complex medical cases. These findings relate to the level of acceptance of GenAI in academia in both countries. For instance, Indonesia's Ministry of Higher Education still provides strict regulations on the use of GenAI in higher education,³⁷ while the Chinese counterpart has a more accommodating and progressive policy on the use of GenAI.³⁸ The regulatory differences between the two countries illustrate how governance approaches can accelerate or hinder the adoption of educational innovation. Indonesia's Ministry of Higher Education promotes with caution, but it seems to slow down the innovation and limit discipline-specific applications. In contrast, China's policy framework accelerates large-scale GenAI adoption in the medical sector by aligning integration with national strategic objectives, though it may risk neglecting local variation and ethical safeguards such as data privacy, patient safety, and responsible clinical use.

Nevertheless, the World Federation for Medical Education does not specifically ban the use of GenAI in medical education and instead recommends the use of this new technology to improve educational and patient care outcomes.³⁹ Hence, medical education institutions in all

countries should explore the benefits of GenAI use in their respective educational and patient care contexts.

As we predicted, most included studies, both from China and Indonesia, expressed apprehension about the accuracy of information and interaction generated by GenAI. Some of the included studies also reported the lack of faculty and student training, which negatively affects the quality of GenAI use. These findings resonate with the concerns of GenAI use in other education fields, such as K-12 education,⁴⁰ higher education,⁴¹ business-marketing education,⁴² and engineering education.⁴³ Faculty/teacher training is pivotal towards the effective and creative use of GenAI in teaching, learning, and assessment.⁴⁴

Untrained GenAI users have reported negative experiences in using GenAI, which can impact their low acceptance and high resistance to using GenAI.⁴⁵ Currie et al.⁴⁶ also highlight how reliance on GenAI-generated content can compromise students' learning outcomes by limiting their ability to meet assessment expectations. Not providing appropriate GenAI ethical training and precautions to learners may also pose a risk to academic dishonesty and even plagiarism.⁴⁷ Hence, medical schools should incorporate training in ethical reasoning and patient-centered care when integrating GenAI into their curricula.⁴⁸

Interestingly, studies included in this review were mostly at level 1 (i.e., satisfaction/reaction) and 2 (i.e., knowledge and skills learning) of Kirkpatrick's level of educational impact evaluation. None of the studies achieved Level 3 (i.e., behavior changes) or Level 4 (i.e., health outcomes) outcomes.¹⁸ It is possible that current studies have not yet been able to evaluate behavioral changes and health outcomes due to time limitations, as these higher impact levels may need more years of program implementation to show intended outcomes.⁴⁹ Meanwhile, the trend of GenAI use, for instance, OpenAI/ChatGPT, has only started at the end of 2022.⁵⁰ Hence, the medical education realm is currently deep in a learning process and will need time to change behavior and healthcare outcomes. This paves the way for upcoming studies to evaluate future healthcare outcomes and healthcare professionals' behavior changes in using GenAI.

Limitations

This study presents several limitations that should be acknowledged. First, the number of empirical studies from China and Indonesia is limited to only

12, and most studies, particularly Indonesian studies, were cross-sectional. In terms of research origin, our review only included Chinese and Indonesian studies. The limitation of methodologies and country settings may constrain the comprehensiveness of our literature review and may affect the generalizability of the findings to be implemented in other contexts. Moreover, some of the included studies also had non-medical (i.e., nursing, veterinarian) trainees in addition to medical trainees as participants, resulting in heterogeneity of context which made the synthesis of findings challenging. Notably, there is a lack of higher-level evaluation outcomes corresponding to levels 3 and 4 of Kirkpatrick's model, which assess behavioral change and organizational results. This suggests that the field of medical education is still in the early stages of integrating generative GenAI and has yet to produce substantial evidence on its impact on clinical behavior and healthcare outcomes.

Future research is needed to evaluate how GenAI influences long-term changes in healthcare professionals' practices and patient outcomes. Moreover, challenges such as AI reliability, adequate training, and ethical considerations remain significant barriers to responsible and effective implementation. The interdisciplinary nature of GenAI and healthcare also adds complexity, necessitating diverse methodological approaches and cross-sector collaboration.

Conclusion and Recommendations

In conclusion, this study identified the current uses of GenAI in medical education in both Indonesia and China. The applications span clinical teaching, curriculum, assessment, and the enhancement of students' learning across different domains, with most empirical research focusing on improving student learning outcomes. Although GenAI brings many opportunities and transformative potential to medical education, addressing the challenges of integration requires a multifaceted and collaborative approach.¹ These efforts show the growing role of GenAI in supporting educational innovation while also revealing areas that remain underexplored.

Hence, institutions should invest in robust evaluation systems to ensure the accuracy and reliability of GenAI-generated content, particularly in high-stakes contexts such as licensing examinations.⁵¹ Faculty development is also essential to equip educators with both the technical and ethical understanding necessary for effective GenAI use,⁵² ensuring transparency, mitigating bias,

and safeguarding both patients and students.⁵²⁻⁵⁴ At the same time, promoting interdisciplinary research and implementing systematic feedback mechanisms can help align GenAI tools with real-world clinical needs, thereby fostering sustainable and equitable implementation across diverse medical education settings.⁵⁵

More importantly, national policies and funding structures must continue to evolve to support curriculum reform, technological infrastructure, and regulatory safeguards.⁵⁶ The contrasting approaches of Indonesia and China demonstrate how governance models directly shape the adoption of GenAI in higher education, particularly within medical education. These differences highlight a broader lesson for global policy: regulatory environments must balance flexibility with accountability. Future policies should encourage discipline-specific strategies in Indonesia, strengthen ethical safeguards in China, and, more broadly, adopt frameworks that ensure GenAI integration is equitable, sustainable, and responsive to local educational and clinical needs as suggested by OECD and UNESCO guidelines.^{57,58}

Our review suggests that medical education institutions should provide appropriate training to students and faculty to use GenAI in education, to leverage the benefits of GenAI, and mitigate the risks of academic dishonesty. Prior evaluations have demonstrated that GenAI tools like ChatGPT can provide useful explanations and even support assessment tasks, yet their outputs often contain inaccuracies, biases, or "hallucinations," especially in non-English medical contexts.^{23,33} Beyond accuracy, upcoming investigations need to move past short-term measures of knowledge gain to explore the longer-term impacts of GenAI use on trainees' behaviors, particularly how reliance on GenAI may influence critical thinking, diagnostic reasoning, and ethical judgment in advanced training or healthcare practice.^{53,59} Critically, almost no studies to date have examined whether GenAI integration improves or potentially undermines the quality of patient care and outcomes, a gap also noted in a recent systematic review.⁶⁰

To address this gap, there is a need to promote collaborative, multi-method research aiming at higher Kirkpatrick evaluation levels (e.g., Level 3 and 4) and to create continuous feedback for improving the quality of GenAI use in medical education. Finally, future studies or reviews on the application of GenAI in medical and health

professions education should include wider country scopes, including lower-, middle-, and high-income

countries, to allow better generalizability of the recommendations for best practice.

References

1. Chan KS, Zary N. Applications and challenges of implementing artificial intelligence in medical education: Integrative review. *JMIR Medical Education*. 2019;5(1):e13930. <https://doi.org/10.2196/13930>
2. Choudhury A, Asan O. Role of Artificial Intelligence in Patient Safety Outcomes: Systematic Literature Review. *JMIR Medical Informatics*. 2020;8(7):e18599. <https://doi.org/10.2196/18599>
3. Zarei M, Eftekhari Mamaghani H, Abbasi A, Hosseini MS. Application of artificial intelligence in medical education: A review of benefits, challenges, and solutions. *Medicina Clínica Práctica*. 2024;7(2):100422. <https://doi.org/10.1016/j.mcpsp.2023.100422>
4. Fryer LK, Nakao K, Thompson A. Chatbot learning partners: Connecting learning experiences, interest and competence. *Computers in Human Behavior*. 2019;93:279–289. <https://doi.org/10.1016/j.chb.2018.12.023>
5. Gignac GE, Szodorai ET. Defining intelligence: Bridging the gap between human and artificial perspectives. *Intelligence*. 2024;104:101832. <https://doi.org/10.1016/j.intell.2024.101832>
6. Kwong JCC, Nickel GC, Wang SCY, Kvedar JC. Integrating artificial intelligence into healthcare systems: More than just the algorithm. *npj Digital Medicine*. 2024;7(1). <https://doi.org/10.1038/s41746-024-01066-z>
7. Dong C, Aw DCW, Lee DWC, Low SC, Yan CC. Generative AI in healthcare: Insights from health professions educators and students. *International Medical Education*. 2025;4(2):11. <https://doi.org/10.3390/ime4020011>
8. Patino GA, Amiel JM, Brown M, Lypson ML, Chan TM. The promise and perils of artificial intelligence in health professions education practice and scholarship. *Academic Medicine*. 2024;99(5):477–481. <https://doi.org/10.1097/acm.0000000000005636>
9. Li F, Yan X, Su H, Shen R, Mao G. An Assessment of Human–AI Interaction Capability in the Generative AI Era: The Influence of Critical Thinking. *Journal of Intelligence*. 2025;13(6):62. <https://doi.org/10.3390/jintelligence13060062>
10. Waldo J, Boussard S. GPTs and hallucination. *Communications of the ACM*. 2025;68(1):40–45. <https://doi.org/10.1145/3703757>
11. Plana D, Shung DL, Grimshaw AA, Saraf A, Sung JJY, Kann BH. Randomized clinical trials of machine learning interventions in health care: A systematic review. *JAMA Network Open*. 2022;5(9):e2233946. <https://doi.org/10.1001/jamanetworkopen.2022.33946>
12. Chatterjee S, Fruhling A, Kotiadis K, Gartner D. Towards new frontiers of healthcare systems research using artificial intelligence and generative AI. *Health Systems*. 2024;13(4):263–273. <https://doi.org/10.1080/20476965.2024.2402128>
13. Kumar A, Mani V, Jain V, Gupta H, Venkatesh VG. Managing healthcare supply chain through artificial intelligence (AI): A study of critical success factors. *Computers & Industrial Engineering*. 2023;175:108815. <https://doi.org/10.1016/j.cie.2022.108815>

14. Panagiotti M, Khan K, Keers RN, Abuzour A, Phipps D, Kontopantelis E, Bower P, Campbell S, Haneef R, Avery AJ, Ashcroft DM. Prevalence, severity, and nature of preventable patient harm across medical care settings: Systematic review and meta-analysis. *BMJ*. 2019;366:l4185. <https://doi.org/10.1136/bmj.l4185>
15. Reddy S, Allan S, Coghlan S, Cooper P. A governance model for the application of AI in health care. *Journal of the American Medical Informatics Association*. 2020;27(3):491–497. <https://doi.org/10.1093/jamia/ocz192>
16. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. *Nature Medicine*. 2022;28(1):31–38. <https://doi.org/10.1038/s41591-021-01614-0>
17. Snyder H. Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*. 2019;104:333–339. <https://doi.org/10.1016/j.jbusres.2019.07.039>
18. Yardley S, Dornan T. Kirkpatrick’s levels and education ‘evidence.’ *Medical Education*. 2012;46(1):97–106. <https://doi.org/10.1111/j.1365-2923.2011.04076.x>
19. Liu P. AI-enabled problem-based learning fused with case-based learning: Application in oncology teaching. *Tumor*. 2024;44(8):868–877. <https://doi.org/10.3781/j.issn.1000-7431.2024.2407-0458>
20. Wang J. An investigation on needs for the artificial intelligence courses in medical undergraduates under the background of new medicine. *Journal of Science and Education*. 2024;615:90–94. <https://doi.org/10.16871/j.cnki.kjwh.2024.15.021>
21. Wang Y, Peng J, Yan Y, Li S, Zhang L, Li Q. Medical students’ knowledge and perception on artificial intelligence and its medical applications. *Chinese Higher Medical Education*. 2023;2(3):25–27. <https://doi.org/10.3969/j.issn.1002-1701.2023.02.012>
22. Xu K, Zhang C, He M. AI-assisted integration of macromolecular structures into medical microbiology teaching practices. *Basic Medical Education*. 2024;26(12):1075–1080. <https://doi.org/10.13754/j.isn2095-1450.2024.12.13>
23. Zong H, Li J, Wu E, Wu R, Lu J, Shen B. Performance of ChatGPT on Chinese national medical licensing examinations: A five-year examination evaluation study for physicians, pharmacists and nurses. *BMC Medical Education*. 2024;24(1):143. <https://doi.org/10.1186/s12909-024-05125-7>
24. Chang CY, Yang CL, Jen HJ, Ogata H, Hwang GH. Facilitating nursing and health education by incorporating ChatGPT into learning designs. *Educational Technology & Society*. 2024;27(1):215–230. [https://doi.org/10.30191/ETS.202401_27\(1\).TP02](https://doi.org/10.30191/ETS.202401_27(1).TP02)
25. Hsu MH. Mastering medical terminology with ChatGPT and Termbot. *Health Education Journal*. 2024;83(4):352–358. <https://doi.org/10.1177/00178969231197371>
26. Yan Y, Liu H, Gao X, Tu J. AI + Education empowering higher vocational medical imaging courses: A case study on X-ray photography examination technology. *Journal of Vocational Medical Education*. 2024;23(98):98–104.
27. Zhang H, Xu H, Yang M, Wang K, Zhang D, Zhou F. Application of CBL combined with TBL teaching method in medical imaging technology teaching under the background of AI. *China Continuing Medical Education*. 2025;17(2):102–107. <https://doi.org/10.3969/j.issn.1674-9308.2025.02.022>
28. Busch F, Hoffmann L, Truhn D, Ortiz-Prado E, Makowski MR, Bressemer KK, Adams LC, COMFORT Consortium. Global cross-sectional student survey on AI in medical, dental, and veterinary education and practice at 192 faculties. *BMC Medical Education*. 2024;24(1):1066. <https://doi.org/10.1186/s12909-024-06035-4>

29. Lugito NPH, Cucunawangsih C, Suryadinata N, Kurniawan A, Wijayanto R, Sungono V, Sabran MZ, Albert N, Budianto, CJ, Rubismo KY, Purushotama NB, Zebua A. Readiness, knowledge, and perception towards artificial intelligence of medical students at faculty of medicine, Pelita Harapan University, Indonesia: A cross sectional study. *BMC Medical Education*. 2024;24(1):1044. <https://doi.org/10.1186/s12909-024-06058-x>
30. Pudjiadi AH, Alatas FS, Faizi M, Rusdi, Sulistijono E, Nancy YM, Julia M, Baso AJA, Hartoyo E, Susannah S, Wilar R, Nugroho HW, Indrayadi I, Lubis BM, Haris S, Suparyatha IBG, Amarassaphira D, Monika E, Ongko L. Integration of artificial intelligence in pediatric education: Perspectives from pediatric medical educators and residents. *Healthcare Informatics Research*. 2024;30(3):244–252. <https://doi.org/10.4258/hir.2024.30.3.244>
31. Barra FL, Rodella G, Costa A, Scalogna A, Carezzo L, Monzani A, Corte FD. From prompt to platform: An agentic AI workflow for healthcare simulation scenario design. *Advances in Simulation*. 2025;10(1):29. <https://doi.org/10.1186/s41077-025-00357-z>
32. Komasaawa N, Yokohira M. Simulation-based education in the artificial intelligence era. *Cureus*. 2023;15(6):e40940 <https://doi.org/10.7759/cureus.40940>
33. Zhang W, Cai M, Lee HJ, Evans R, Zhu C, Ming C. AI in medical education: Global situation, effects and challenges. *Education and Information Technologies*. 2024;29(4):4611–4633. <https://doi.org/10.1007/s10639-023-12009-8>
34. Sousa MJ, Dal Mas F, Pesqueira A, Lemos C, Manuel Verde J, Cobianchi L. The potential of AI in health higher education to increase the students' learning outcomes. *TEM Journal*. 2021;10(2):488–497. <https://doi.org/10.18421/tem102-02>
35. Abouammoh N, Alhasan K, Aljamaan F, Raina R, Malki KH, Altamimi I, muaygil R, Wahabi H, Jamar A, Alhaboob A, Assiri RA, Al-Tawfiq JA, Al-Eyadhy A, Soliman M, Temsah MH. Perceptions and earliest experiences of medical students and faculty with ChatGPT in medical education: Qualitative study. *JMIR Medical Education*. 2025;11:e63400. <https://doi.org/10.2196/63400>
36. Shen M, Shen Y, Liu F, Jin J. Prompts, privacy, and personalized learning: Integrating AI into nursing education—a qualitative study. *BMC Nursing*. 2025;24(1):470. <https://doi.org/10.1186/s12912-025-03115-8>
37. Ministry of Education, Culture, Research and Technology of the Republic of Indonesia. Panduan penggunaan Generative Artificial Intelligence (GenAI) pada pembelajaran di perguruan tinggi [Internet]. *Ministry of Education, Culture, Research and Technology of the Republic of Indonesia*; 2024 [cited 2025 Jun 10]. Available from: <https://simbelmawa.kemdikbud.go.id/portal/unduh/buku-panduan-penggunaan-generative-artificial-intelligence-pada-pembelajaran-di-perguruan-tinggi/>
38. China State Council. Implementation plan for the digital and intelligent transformation of the pharmaceutical industry (2025–2030) [Internet]. *Ministry of Industry and Information Technology (MIIT), Ministry of Commerce, National Health Commission, National Healthcare Security Administration, National Data Administration, National Administration of Traditional Chinese Medicine, & National Medical Products Administration*; 2025 [cited 2025 Jun 10]. Available from: https://www.gov.cn/zhengce/zhengceku/202504/content_7020857.htm
39. World Federation For Medical Education. Training and research: Role of artificial intelligence and the WFME standards [Internet]. *World Federation For Medical Education*; 2024 [cited 2025 Jun 10]. Available from: <https://amcoa.org/wp-content/uploads/2024/08/Training-and-Research-Role-of-Artificial-Intelligence-and-the-WFME-Standards-%E2%80%93Dr.-Ricardo-Leon-Borquez.pdf>

40. Cheah YH, Lu J, Kim J. Integrating generative artificial intelligence in K-12 education: Examining teachers' preparedness, practices, and barriers. *Computers and Education: Artificial Intelligence*. 2025;8:100363. <https://doi.org/10.1016/j.caeai.2025.100363>
41. Meakin L. Exploring the impact of generative artificial intelligence on higher education students' utilization of library resources: A critical examination. *Information Technology and Libraries*. 2024;43(3). <https://doi.org/10.5860/ital.v43i3.17246>
42. Acar OA. Reimagining marketing education in the age of generative AI. *International Journal of Research in Marketing*. 2024;41(3):489–495. <https://doi.org/10.1016/j.ijresmar.2024.06.004>
43. Honig C, Rios S, Desu A. Generative AI in engineering education: Understanding acceptance and use of new GPT teaching tools within a UTAUT framework. *Australasian Journal of Engineering Education*. 2025;30(1):80–92. <https://doi.org/10.1080/22054952.2025.2467500>
44. Zhang X, Zhang P, Shen Y, Liu M, Wang Q, Gašević D, et al. A systematic literature review of empirical research on applying generative artificial intelligence in education. *Frontiers of Digital Education*. 2024;1(3):223–245. <https://doi.org/10.1007/s44366-024-0028-5>
45. Watermeyer R, Lanclos D, Phipps L, Shapiro H, Guizzo D, Knight C. Academics' weak(ening) resistance to generative AI: The cause and cost of prestige? *Postdigital Science and Education*. 2024; <https://doi.org/10.1007/s42438-024-00524-x>
46. Currie G, Singh C, Nelson T, Nabasenja C, Al-Hayek Y, Spuur K. ChatGPT in medical imaging higher education. *Radiography*. 2023;29(4):792–799. <https://doi.org/10.1016/j.radi.2023.05.011>
47. Yusuf A, Pervin N, Román-González M. Generative AI and the future of higher education: A threat to academic integrity or reformation? Evidence from multicultural perspectives. *International Journal of Educational Technology in Higher Education*. 2024;21(1):21. <https://doi.org/10.1186/s41239-024-00453-6>
48. Lo B. Resolving ethical dilemmas: *A guide for clinicians*. Baltimore, MD: Lippincott Williams & Wilkins; 2012.
49. Heydari MR, Taghva F, Amimi M, Delavari S. Using Kirkpatrick's model to measure the effect of a new teaching and learning methods workshop for health care staff. *BMC Research Notes*. 2019;12(1):388. <https://doi.org/10.1186/s13104-019-4421-y>
50. OpenAI. Introducing ChatGPT [Internet]. *OpenAI*; 2022 [cited 2025 Jun 10]. Available from: <https://openai.com/index/chatgpt/>
51. McDonald N, Johri A, Ali A, Collier AH. Generative artificial intelligence in higher education: Evidence from an analysis of institutional policies and guidelines. *Computers in Human Behavior: Artificial Humans*. 2025;3:100121. <https://doi.org/10.1016/j.chbah.2025.100121>
52. Khakpaki A. Advancements in artificial intelligence transforming medical education: A comprehensive overview. *Medical Education Online*. 2025;30(1):2542807. <https://doi.org/10.1080/10872981.2025.2542807>
53. Shaw K, Henning MA, Webster CS. Artificial Intelligence in medical education: A scoping review of the evidence for efficacy and future directions. *Medical Science Educator*. 2025;35(3):1803–1816. <https://doi.org/10.1007/s40670-025-02373-0>
54. Preiksaitis C, Rose C. Opportunities, challenges, and future directions of Generative Artificial Intelligence in medical education: Scoping review. *JMIR Medical Education*. 2023;9:e48785. <https://doi.org/10.2196/48785>

55. Ahsan Z. Integrating artificial intelligence into medical education: A narrative systematic review of current applications, challenges, and future directions. *BMC Medical Education*. 2025;25(1):1187. <https://doi.org/10.1186/s12909-025-07744-0>
56. Thimbleby H. Technology and the future of healthcare. *Journal of Public Health Research*. 2013;2(3):e28. <https://doi.org/10.4081/jphr.2013.e28>
57. OECD. OECD principles on artificial intelligence. *OECD Publishing*; 2021 [cited 2025 Sep 1]. <https://doi.org/10.1787/eedfee77-en>
58. UNESCO. Recommendation on the ethics of artificial intelligence [Internet]. *United Nations Educational, Scientific and Cultural Organization*; 2021 [cited 2025 Sep 2]. Available from: <https://unesdoc.unesco.org/ark:/48223/pf0000381137>
59. Janumpally R, Nanua S, Ngo A, Youens K. Generative artificial intelligence in graduate medical education. *Frontiers in Medicine*. 2025;11:1525604. <https://doi.org/10.3389/fmed.2024.1525604>
60. Feigerlova E, Hani H, Hothersall-Davies E. A systematic review of the impact of artificial intelligence on educational outcomes in health professions education. *BMC Medical Education*. 2025;25(1):129. <https://doi.org/10.1186/s12909-025-06719-5>