


# AI Use for Medical Students: Impact on Clinical Skill Acquisition and Retention. A Systematic Review

Jonathan Turney <sup>1</sup>, Tim M Young<sup>2</sup>, Dhyana R Chauhan <sup>1</sup>, Roshni Beeharry<sup>2</sup>,  
Mohammad Mahmud <sup>2</sup>

<sup>1</sup>UCL Medical School, University College London, London, UK; <sup>2</sup>The Queen Square Institute of Neurology, University College London, London, UK

Correspondence: Jonathan Turney, UCL Medical School, University College London, London, WC1N 3BG, UK, Tel +447729186903,  
Email zchaurn@ucl.ac.uk

**Purpose:** Artificial Intelligence (AI) is increasingly used in undergraduate medical education but has a potentially negative impact on clinical reasoning development. Specifically, the use of AI in medical student education may lead to deskilling and upskilling inhibition - where automation reduces practice or limits skill development - potentially impairing clinical reasoning. This systematic review aimed to synthesise evidence regarding AI-supported learning effects on acquisition and retention of clinical skills in medical students to assess its potential negative impact in medical education.

**Methods:** A systematic search was conducted on 21 October 2025 across PubMed, Scopus, and Embase using structured Boolean queries restricted to titles and abstracts. Inclusion criteria targeted published studies involving medical students exposed to AI tools in clinical learning, reporting outcomes related to skill acquisition, reasoning, or overreliance on AI. Exclusions included non-AI digital tools, administrative AI applications, and studies without clear educational outcomes. Screening followed PRISMA guidelines.

**Results:** From 420 records, 255 were screened. Four studies met the screening criteria, incorporating a total of 408 medical students. Across included studies, AI exposure was associated with improved efficiency and improved basic knowledge acquisition. When higher-order clinical reasoning and complex decision-making were assessed, findings were mixed: one study reported no overall difference, while others suggested weaker performance or reduced engagement when AI-supported approaches were used.

**Conclusion:** Current evidence suggests that AI-supported learning may be associated with improved efficiency and basic knowledge acquisition in undergraduate medical education. Findings were less consistently supportive of higher-order reasoning outcomes compared with traditional teaching approaches, although the evidence base was limited. Potential risks of deskilling and upskilling inhibition warrant attention as medical schools increasingly integrate AI tools into their curricula. A striking finding of our systematic review was the very low number of existing studies identified in this important field. Further research should explore the long-term impacts of AI on medical students' independent clinical judgement and consider strategies to mitigate overreliance on AI given the profound potential impact on future patient care.

**Keywords:** clinical reasoning, deskilling, upskilling inhibition

## Introduction

Artificial Intelligence (AI) is rapidly becoming embedded within medical education, with tools such as adaptive learning platforms, AI-assisted simulation, automated clinical reasoning support, and machine-learning-driven feedback now increasingly integrated into undergraduate medical curricula. These developments may aid medical student learning, personalised instruction, and also better prepare future clinicians for an AI-enabled healthcare environment. However, as AI-supported tools become increasingly incorporated in medical students' learning experiences, questions have emerged about how these systems may influence the development and retention of foundational clinical skills among these future doctors.<sup>1</sup>

Rather than functioning solely as isolated educational tools, AI systems are increasingly part of a systemic shift in medical education, altering the ways learners access knowledge, approach clinical problems, and demonstrate

competence. As generative and decision support systems become embedded across healthcare and training environments, medical students may encounter AI-mediated reasoning not as an optional supplement but as a routine part of learning and future clinical practice. This shift raises urgent questions about how to best integrate AI while preserving opportunities for independent clinical reasoning and skill development.

Work by Thurzo has highlighted how AI is reshaping medical research, education, and clinical practice simultaneously, enabling immersive simulation, personalised learning pathways, and rapid AI-supported diagnostic assistance, whilst also raising ethical and practical concerns related to over-reliance. To help counter such threats, Thurzo argues that human oversight remains essential and that future doctors must continue to develop the critical thinking and contextual judgement required to interpret AI outputs appropriately and to transcend AI's algorithmic limitations.<sup>2</sup> Such concerns are particularly important at the undergraduate level when the development of critical thinking typically begins. The development of critical thinking in medical students has been associated with improvements in academic performance, diagnostic reasoning, and preparedness for clinical decision-making, with implications for future patient care.<sup>3</sup> However, overdependence on AI by such undergraduates may risk impeding the development of critical thinking skills in addition to potentially eroding such skills in those medical learners who have already acquired them.<sup>2</sup>

Clinical reasoning is often identified as a core component of physician competence, yet is challenging to define, teach, and assess, and remains inconsistently emphasised within undergraduate curricula.<sup>4</sup> AI applications in clinical decision-making are supported by a well-established evidence base, with multiple reviews demonstrating enhanced performance across diagnostic, triage, and imaging tasks.<sup>5,6</sup> However, the educational implications of exposing medical students to learning through AI systems remain under-examined. Existing research provides limited insight into how such tools influence skill acquisition, diagnostic reasoning, or long-term learning trajectories. Concerns about how AI may shape clinical judgement are also beginning to appear in policy-level commentary. For instance, a recent review by the World Health Organization noted that while large language models may enhance personalised learning and simulated patient interactions, they also risk encouraging premature reliance on AI-generated outputs and reducing opportunities for independent clinical reasoning.<sup>7</sup> Such observations reflect a growing concern about how the use of such autonomous systems may alter the development of foundational skills during undergraduate medical training.

Evidence is emerging that over-reliance on AI tools may lead to declines in core cognitive skills. A recent systematic review found that when students across different disciplines routinely accept AI-generated outputs without verification, their critical thinking, analytical reasoning, and decision-making abilities become impaired.<sup>8</sup> However, few studies to date have assessed AI's impact on critical thinking or clinical reasoning specifically in medical students. Over-dependence on AI in medical students could arise if learners favour quick, AI-generated shortcuts over slower, independent reasoning processes required to evaluate information for themselves. This could be compared to a traditional setting where a medical student might rely on a senior clinician to make seemingly instant diagnoses on seeing a patient. Such heuristic, intuitive, clinical thinking (the System 1 of dual-process theory) by the senior clinician however would be built on many years of practice initially using the slower systematic and analytical System 2 process.<sup>9</sup> Rather than viewing System 1 as being "better" because it is faster, the combination of both systems (for example cross-checking intuitive thinking with systematic analysis) may produce more accurate results and allow for senior clinicians to work productively together with medical students on clinical problems with each contributing.<sup>10</sup> Reducing the need for medical students to evaluate evidence or justify decisions might result in erosion of such independent judgement, diminished information retention, and a weakening of higher-order skills essential for clinical reasoning in medical education.<sup>8</sup>

Recent literature has highlighted two mechanisms through which AI automation may reshape human expertise: deskilling, in which previously acquired skills erode due to reduced practice, and upskilling inhibition, which refers to missed opportunities to develop new competencies in the first place when tasks are delegated to automated systems.<sup>1,11</sup> While deskilling has traditionally been discussed in relation to practising clinicians or senior medical students, upskilling inhibition is particularly salient in undergraduate education, where core competencies are still forming. Students develop clinical capability through repeated engagement with uncertain and cognitively demanding tasks such as generating differential diagnoses, interpreting ambiguous findings, and integrating information during patient encounters.<sup>12,13</sup> If AI tools provide these outputs too early or too consistently, learners may engage less deeply with the underlying reasoning processes, reducing opportunities to practise and consolidate the skills required for independent clinical judgement, procedural proficiency, and diagnostic problem-solving. This framework

therefore offers a useful lens for examining how AI-enabled learning environments might alter the development and trajectory of foundational clinical skills, and it underpins the focus of the present review.

Given the rapid expansion of AI in medical curricula,<sup>14</sup> it is essential to understand how these technologies may adversely affect the acquisition and retention of clinical skills during the formative stages of medical training. This systematic review therefore examined current evidence on the potential risks associated with AI-supported learning in undergraduate medicine, with particular attention to mechanisms such as upskilling inhibition and early-stage deskilling. By examining where and how AI integration may affect the development of independent clinical judgement, procedural competence, and diagnostic reasoning, this review offers a counterbalance to the largely benefit-focused accounts in the current literature and supports more considered, evidence-informed approaches to integrating AI in medical education.

## Methods

Searching for relevant studies was performed using the terms shown in [Table 1](#) and was completed on the 21<sup>st</sup> of October 2025. The following bibliographic databases were sourced for the search: PubMed, Scopus, and Embase. PubMed and Embase were selected based on strong evidence for use in biomedical literature reviews.<sup>15</sup> While Bramer et al refer to MEDLINE, PubMed was used in this present systematic review as it included MEDLINE and additional records before being indexed or sourced from other biomedical journals.<sup>16</sup> Scopus was additionally included to expand the search into additional scientific fields and to include interdisciplinary journals that were not fully covered by biomedical-focused databases such as Embase.<sup>16</sup>

**Table 1** Search Terms Used for PubMed, Scopus, and Embase

Database	Search Query
PubMed	("artificial intelligence"[Title/Abstract] OR ai[Title/Abstract] OR "machine learning"[Title/Abstract] OR "deep learning"[Title/Abstract]) AND ("deskilling"[Title/Abstract] OR "de-skilling"[Title/Abstract] OR "loss of skills"[Title/Abstract] OR "skills degradation"[Title/Abstract] OR "skills erosion"[Title/Abstract] OR "skills decline"[Title/Abstract] OR "skill atrophy"[Title/Abstract] OR "skill loss"[Title/Abstract] OR "skill erosion"[Title/Abstract] OR "skill decay"[Title/Abstract] OR "skill fade"[Title/Abstract] OR "upskilling inhibition"[Title/Abstract] OR "suppressed skill acquisition"[Title/Abstract] OR "skill acquisition"[Title/Abstract] OR "learning inhibition"[Title/Abstract] OR "competency development"[Title/Abstract] OR "clinical reasoning"[Title/Abstract] OR "critical thinking reduction"[Title/Abstract] OR "reduced critical thinking"[Title/Abstract] OR "critical thinking decline"[Title/Abstract] OR "critical thinking erosion"[Title/Abstract] OR "loss of critical thinking"[Title/Abstract] OR "impaired critical thinking"[Title/Abstract] OR "critical reasoning decline"[Title/Abstract] OR "reduced clinical reasoning"[Title/Abstract]) AND ("medical education"[Title/Abstract] OR "medical students"[Title/Abstract] OR "clinical education"[Title/Abstract] OR "undergraduate medical education"[Title/Abstract] OR "medical training"[Title/Abstract] OR "clinical training"[Title/Abstract] OR "medical curriculum"[Title/Abstract] OR "medical learning"[Title/Abstract] OR "clinical reasoning education"[Title/Abstract] OR "medical teaching"[Title/Abstract] OR "medical school"[Title/Abstract] OR "health professions education"[Title/Abstract] OR "healthcare education"[Title/Abstract] OR "medical trainees"[Title/Abstract] OR "medical undergraduates" [Title/Abstract])
Scopus	TITLE-ABS ( ("artificial intelligence" OR ai OR "machine learning" OR "deep learning") AND ("deskilling" OR "de-skilling" OR "loss of skills" OR "skills degradation" OR "skills erosion" OR "skills decline" OR "skill atrophy" OR "skill loss" OR "skill erosion" OR "skill decay" OR "skill fade" OR "upskilling inhibition" OR "suppressed skill acquisition" OR "skill acquisition" OR "learning inhibition" OR "competency development" OR "clinical reasoning" OR "critical thinking reduction" OR "reduced critical thinking" OR "critical thinking decline" OR "critical thinking erosion" OR "loss of critical thinking" OR "impaired critical thinking" OR "critical reasoning decline" OR "reduced clinical reasoning") AND ("medical education" OR "medical students" OR "clinical education" OR "undergraduate medical education" OR "medical training" OR "clinical training" OR "medical curriculum" OR "medical learning" OR "clinical reasoning education" OR "medical teaching" OR "medical school" OR "health professions education" OR "healthcare education" OR "medical trainees" OR "medical undergraduates"))

(Continued)

**Table 1** (Continued).

Database	Search Query
Embase	((“artificial intelligence” or AI or “machine learning” or “deep learning”).ti,ab. and (“deskilling” or “de-skilling” or “loss of skills” or “skills degradation” or “skills erosion” or “skills decline” or “skill atrophy” or “skill loss” or “skill erosion” or “skill decay” or “skill fade” or “upskilling inhibition” or “suppressed skill acquisition” or “skill acquisition” or “learning inhibition” or “competency development” or “clinical reasoning” or “critical thinking reduction” or “reduced critical thinking” or “critical thinking decline” or “critical thinking erosion” or “loss of critical thinking” or “impaired critical thinking” or “critical reasoning decline” or “reduced clinical reasoning”).ti,ab. and (“medical education” or “medical students” or “clinical education” or “undergraduate medical education” or “medical training” or “clinical training” or “medical curriculum” or “medical learning” or “clinical reasoning education” or “medical teaching” or “medical school” or “health professions education” or “healthcare education” or “medical trainees” or “medical undergraduates”).ti,ab.)

Boolean operators (AND, OR) and parentheses were used to structure the search queries and ensure logical grouping of terms. In PubMed, the “Title/Abstract” field was selected directly through the advanced search interface. In Scopus, the default field “Article title, Abstract, Keywords” was used, and in the advanced search, the KEY field was manually removed to restrict the search to titles and abstracts. In Embase, the same was achieved by manually typing “ti,ab” at the end of the query to limit the search to title and abstract fields. Full search queries are provided in [Table 1](#).

Restricting searches to the Title and Abstract fields was adopted to improve precision and ensure alignment with the systematic review’s inclusion criteria. Terms such as artificial intelligence, deskilling, and skill acquisition are used widely across multiple disciplines, and unrestricted searches produced several thousand records, the majority unrelated to the impact of AI on clinical skill development. This restriction enabled the retrieval of papers in which these topics formed a primary focus rather than a secondary or incidental mention. This approach reduced irrelevant returns while maintaining sensitivity to studies directly addressing the mechanisms of interest.

The inclusion and exclusion criteria ([Table 2](#)) were organised using an adapted PIRT framework: Population, Intervention, Result, and Type of study.<sup>16</sup> This structure was selected to ensure alignment with the review’s focus on

**Table 2** Inclusion and Exclusion Criteria

	Inclusion	Exclusion
Population	Studies involving medical students	Studies that focus primarily on nursing, doctor, or healthcare professional education
Intervention	Studies examining the use or exposure to artificial intelligence (AI) or machine learning tools in medical learning, clinical reasoning, or skills training. Includes AI-assisted simulation, tutoring systems, or diagnostic support used as part of education or training.	Studies focused primarily on AI for assessment, grading, or administrative purposes at the institutional level. Studies centred on AI model development or validation without a student-learning component. Studies that focus primarily on non-AI digital tools (eg, traditional e-learning or LMS platforms).
Result	Studies directly and significantly dealing with the theme of deskilling or upskilling inhibition. Studies reporting or discussing reduced skill acquisition, diminished clinical reasoning, reduced critical thinking among learners, or overreliance on AI tools among learners.	Papers merely mentioning the concept of deskilling or upskilling inhibition without elaboration.
Type of Study	Original Research Articles, Systematic Reviews and Meta-Analyses, Scoping Reviews, Mixed-Methods Studies (for examples including both qualitative and quantitative data), Clinical Education Intervention Studies.	Narrative reviews, Conference proceedings, Grey literature, Editorials or Opinion Pieces, Technical AI Development Papers, Studies Without Clear Educational Outcomes, Duplicate Publications, Non-English Articles, single case reports, Studies Focused Solely on Administrative or Logistical AI Use, articles studying only a small numbers of students (10 or less) akin to case studies or case series.

**Abbreviation:** LMS, Learning Management Software.

how AI-based educational interventions affect clinical skill acquisition and retention in undergraduate medical students. Using this framework supported clarity and consistency during screening by mapping each component of the research question directly onto the eligibility criteria.

## Results

Database searches across PubMed ( $n = 146$ ), Scopus ( $n = 145$ ), and Embase ( $n = 129$ ) identified a total of 420 records. After removing 165 duplicates, 255 unique records remained for title and abstract screening as shown in the PRISMA Flow diagram (Figure 1). Following screening, 246 records were excluded, and nine full-text articles were sought for retrieval, all of which were successfully obtained. Of these, five were excluded at full-text review because they did not adequately address deskilling, upskilling inhibition, or related concepts. The remaining total of four studies met the inclusion criteria without meeting any exclusion criteria and were included in the final synthesis. These four studies encompassed a total of 408 medical students (Table 3).

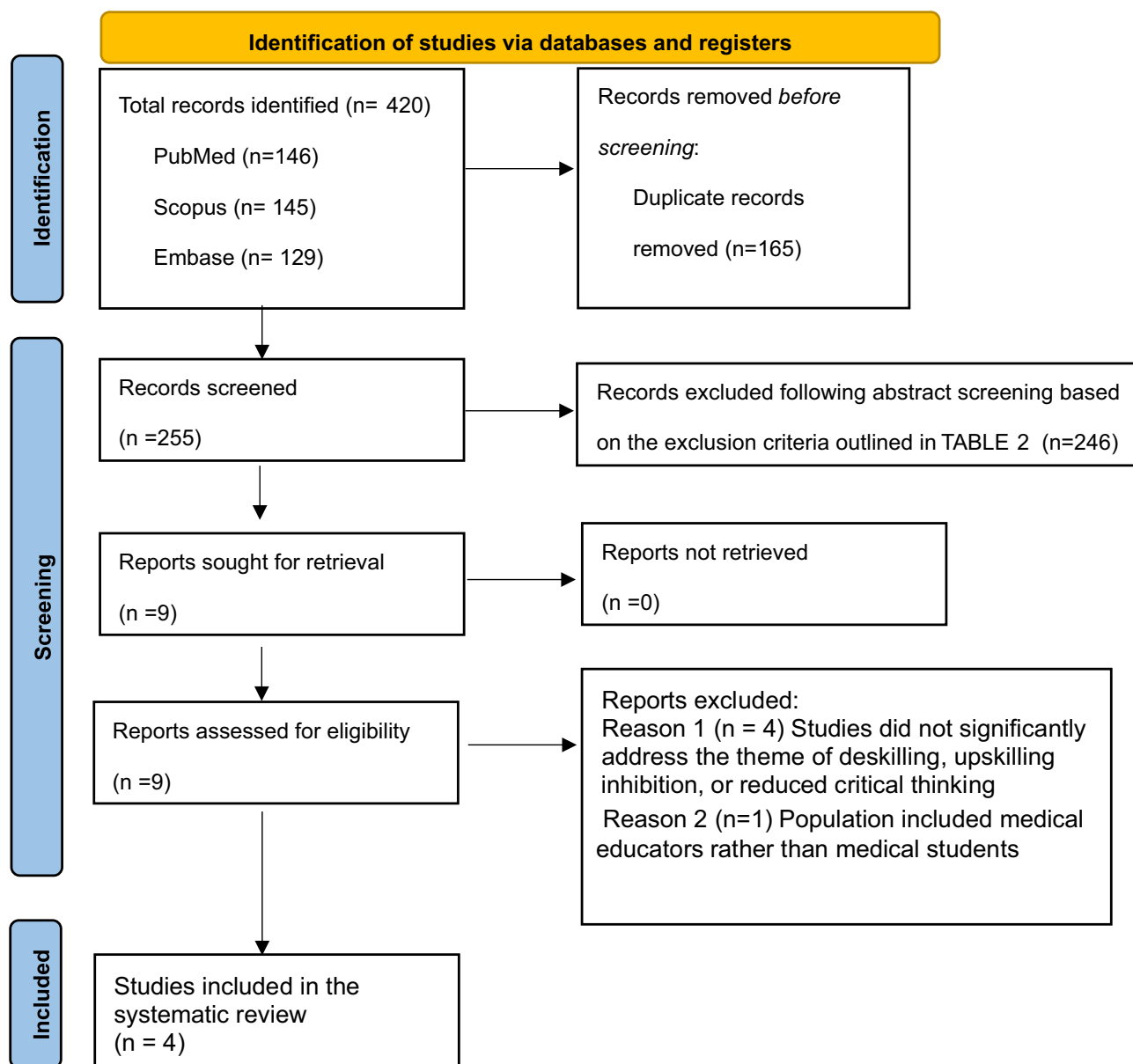


Figure 1 PRISMA-type flowchart displaying the study search and selection process.<sup>17</sup>

**Table 3** Summary of Key Outcomes, Study Design and Aims, Subject Type and Focus of AI Use and the Sample Population

Study (Year)	Study Design & Aims	Focus of AI Use	Sample (n Medical Students)	Sample Size Relative to Other Studies	Key Outcomes (Extracted Directly from Each Study)
Li et al, 2025 <sup>18</sup>	Controlled comparative study. Aim: Tested generative AI (Kimi 2.0) as a teaching assistant in case-based learning (biochemistry) vs traditional learning.	AI used for content generation, assignment support, and automated marking.	N = 79 (40 control, 39 AI)	■■■■■	AI group completed tasks faster (2.6 h vs 5.5 h) and achieved higher exam scores (77.3 vs 66.5). AI grading aligned closely with teacher scoring. Students rated AI highly for basic knowledge acquisition but low for complex reasoning and innovative thinking.
Çiçek et al, 2025 <sup>19</sup>	Randomised Controlled Trial. Aim: Compared ChatGPT-generated feedback vs expert feedback during spaced CEQ clinical reasoning skill acquisition. Test impact on immediate/delayed performance.	AI-generated formative feedback.	N = 129 (65 control, 64 AI)	■■■■■■■ ■	No significant difference in overall Key-Features scores (immediate or 10-day delayed). Expert feedback outperformed AI in complex UTI cases. AI exposure increased students' scepticism concerning perceived AI accuracy.
Montagna et al, 2025 <sup>20</sup>	Controlled comparison during a hands-on internal medicine session. Aim: Compared three clinical decision-support systems (CPGs, online repositories, ChatGPT) on case-solving effectiveness and speed.	Clinical decision-making, diagnostic workup, differential diagnosis.	N = 16 (CPG n=6, OR n=5, ChatGPT n=5)	■	ChatGPT group answered the fastest but produced less accurate responses than guidelines. CPG group achieved highest accuracy and completeness. ChatGPT responses showed susceptibility to prompt quality and risk of error propagation.
Nwe et al, 2025 <sup>21</sup>	Cross-sectional, questionnaire-based survey of 1 <sup>st</sup> and 2 <sup>nd</sup> year medical students	Generative AI use for content generation, study support, problem-solving, and perceived support for clinical reasoning.	n= 184 respondents	■■■■■■■ ■■■■■	Students raised concerns for over-reliance (83.2%) and loss of critical thinking skills (77.7%); mixed perceptions regarding clinical decision-making (38% supportive, 43% neutral, 19% opposed). AI was perceived as effective for problem solving (76.1%), decision making (63.6%) and critical thinking (66.8%).

**Note:** Full bar (■) equals an n number of 20 participants.

**Abbreviations:** CEQ, ContExtended Questions; CPG, Clinical Practice Guidelines; UTI, Urinary Tract Infections.

Despite considerable heterogeneity in AI application and outcome domains, three of the four included studies directly examined the impact of generative AI on objective clinical learning performance,<sup>18–20</sup> while one study explored students' self-reported perceptions of AI's educational impact.<sup>21</sup> One controlled study demonstrated marked improvements in task efficiency and examination performance when AI tools supported case-based learning, including a halving of task-completion time and higher final-exam scores compared with traditional learning methods.<sup>18</sup> In the only randomised controlled trial, Çiçek et al compared ChatGPT-generated feedback with expert feedback. Whilst there was no difference in overall clinical reasoning performance immediately or after a 10-day delay, expert feedback remained superior for complex diagnostic cases.<sup>19</sup> AI exposure still significantly increased students' critical appraisal of the AI outputs themselves.<sup>19</sup> In the third identified study, in which a hands-on comparison of three clinical decision-support modalities was studied, ChatGPT enabled the fastest responses yet produced less accurate and less complete clinical decisions compared with guideline-based reasoning.<sup>20</sup> In contrast to these performance-based evaluations, Nwe et al (2025) conducted a cross-sectional survey of preclinical medical students examining perceived educational impact.<sup>21</sup> While a majority of students rated AI as effective for problem-solving, decision-making and critical thinking, substantial proportions also expressed concern regarding over-reliance (83.2%) and potential loss of critical thinking skills (77.7%), with mixed views on its role in clinical decision-making.<sup>21</sup>

Demographic reporting was limited across all studies, with only two providing detailed age characteristics (Table 4). Collectively, the identified studies suggest that generative AI may improve efficiency and support basic knowledge

**Table 4** Characteristics of Study Populations Including Country, Mean Age in years, Gender, Ethnicity, and year of Study

Study	Country	Mean Age in Years ( $\pm$ SD)	Gender	Ethnicity	Year of Study/ Training Stage
Li et al, 2025 <sup>18</sup>	China	Control: 19 $\pm$ 0.9; AI: 18 $\pm$ 0.7	Control 20F/20M; AI 19F/20M	Mixed; balanced distribution stated	Pre-clinical (Year 1–2 equivalent; medical biochemistry course)
Çiçek et al, 2025 <sup>19</sup>	Turkey	Not Reported	66F/49M overall (group breakdown not reported)	Not reported	First-year medical students
Montagna et al, 2025 <sup>20</sup>	Italy	Not Reported	CPG 5F/1M; OR 2F/3M; ChatGPT 3F/2M	Not reported	Senior medical students (final-year internal medicine course)
Nwe et al, 2025 <sup>21</sup>	Malaysia	20.4 $\pm$ 1.27	126F/58M	Not reported	Pre-clinical (Year 1–2 MBBS)

**Abbreviations:** CPG, Clinical Practice Guidelines; OR, Online repositories; F, Female; M, Male.

**Table 5** Summary of Reported Effects of AI on Medical Students' Clinical Skill Development Across Key Competency Domains

Skill Area	Evidence of Improvement with AI	Evidence of Neutral Effect	Evidence of Impairment / Concern
Basic Knowledge acquisition	Strong improvement (Li et al, 2025) <sup>18</sup> High perceived support for simplifying concepts and summarizing material (Nwe et al, 2025) <sup>21</sup>		
Clinical reasoning accuracy		ChatGPT comparable to experts (Çiçek et al, 2025) <sup>19</sup>	Underperforms experts in complex cases (Çiçek et al, 2025; <sup>19</sup> Montagna et al, 2025) <sup>20</sup>
Speed/efficiency	Faster task completion (Li et al, 2025; <sup>18</sup> Montagna et al, 2025) <sup>20</sup>		
Diagnostic accuracy/decision-making	63.6% perceived improvement in clinical decision making (self-reported) (Nwe et al, 2025) <sup>21</sup>		Lower than guideline-based reasoning; error propagation from poor prompts (Montagna et al, 2025) <sup>20</sup>
Retention	No significant difference (Çiçek et al, 2025) <sup>19</sup>		
Critical appraisal of AI	AI exposure increases critical thinking (Çiçek et al, 2025) <sup>19</sup>		78.8% reported concern regarding a potential loss of critical thinking skills (Nwe et al, 2025) <sup>21</sup>
Higher-order reasoning/innovation	76.1% perceived AI effective in assisting in solving clinical problems (Nwe et al, 2025) <sup>21</sup> 66.8% rated AI effective in enhancing critical thinking (Nwe et al, 2025) <sup>21</sup>		Students perceived AI as limited in handling complex reasoning and innovative thinking (Li et al, 2025) <sup>18</sup>
Over-reliance			83.7% of students raised concern of over-reliance on technology (Nwe et al, 2025) <sup>21</sup>

acquisition. However, while students frequently perceive benefits in decision-making and higher order reasoning, objective evidence indicates that performance in complex clinical reasoning and diagnostic accuracy remains inconsistent and, in some cases, inferior to expert-derived guidance (Table 5). Importantly, the perceived benefits reported in higher-order reasoning domains arise from a single cross-sectional self-reported survey, whereas the other included studies employed objective assessments of clinical performance.

## Discussion

This systematic review synthesised evidence from studies involving over 400 medical students across two continents at various stages of their undergraduate training. The identified studies evaluated the impact of AI-supported learning on clinical skill acquisition and retention. Across the studies meeting the inclusion and exclusion criteria, AI tools showed some evidence

of enhanced efficiency and facilitated basic knowledge acquisition. For example, medical students utilising generative AI completed tasks in less than half the time required by peers employing traditional methods, while also achieving higher examination scores within the generative AI cohort.<sup>18</sup> Similarly, the use of ChatGPT for decision support enabled faster completion of case-solving exercises compared to conventional approaches.<sup>19</sup> These findings align with recent literature highlighting the potential of AI to improve both efficiency and effectiveness in routine learning processes within medical education.<sup>14</sup> Complementing these objective findings, a cross sectional survey reported high perceived support for simplifying complex concepts and summarising learning material among preclinical students, although these findings were derived from subjective student perceptions rather than objective performance measures.<sup>21</sup>

Our results provided some evidence of deskilling in medical students when using AI as opposed to more traditional methods. The study by Li et al showed that the AI group of medical students asked fewer questions (10 versus 35) compared to the control group.<sup>18</sup> Although not definitive, this suggests reduced active engagement, potentially with its associated inquiry-based learning. This study also reported lower ratings for complex reasoning and innovation, despite students rating AI highly for basic knowledge, as shown in Table 3. Thus, while medical students completed tasks faster when using AI (2.6 hours versus 5.5 hours) and scored more highly in assessments, their deeper cognitive engagement appeared diminished, consistent with deskilling. The study by Li et al focused on preclinical biochemistry tasks for medical students, which might inherently require less complex reasoning than clinical cases, potentially limiting interpretation of these results. In keeping with this, the study states that knowledge acquisition and testing of this were the key requirements of the medical students at their stage of learning as opposed to development of higher order thinking processes.<sup>18</sup> However, the chosen methodology of case-based learning is well recognised as developing higher order thinking in medical education,<sup>22</sup> and as such the study by Li et al still has relevance when considering evidence both for deskilling and upskilling impairment.

The randomised controlled trial by Çiçek et al included 129 first year medical students and reported significantly lower medical student performance when using AI feedback compared to traditional expert human feedback for more challenging clinical scenarios involving urinary tract infections cases with complications as shown in Table 3.<sup>19</sup> Clinical reasoning skills were a primary focus of this study. This provides some evidence that reliance on AI feedback may have resulted in deskilling because the AI and control groups were compared in real time - essentially assessing pre-existing skills. In particular, this suggests erosion of advanced clinical skills through over-reliance on AI. It is conceivable that the use of language may have led to less good responses in the intervention (ChatGPT) arm of the study. The expert human feedback was provided in Turkish, the language used by the medical students, but the ChatGPT answers were produced in English and then translated into Turkish, an additional step that may have impacted the quality of the output. This study by Çiçek et al only involved first-year medical students with limited clinical experience.<sup>19</sup> Their baseline clinical reasoning ability was therefore likely limited, so differences in complex case performance might reflect prior knowledge gaps rather than AI-induced deskilling. Such an argument can however be countered as the Key-Features approach used to assess clinical decision-making skills by Çiçek et al has been validated even at such early stages of medical training.<sup>23</sup>

In the study by Montagna et al, the medical student group guided by ChatGPT produced less accurate responses than the control group using guideline-based reasoning, although that might reflect the small total subject number of just 16 students.<sup>20</sup> Concerningly, there was evidence of error propagation resulting from poor prompts (Table 3). This suggests that speed of using AI may have been prioritised over depth of learning, risking skill degradation in clinical decision-making. Nevertheless, this study did not report significant differences in the clinical decision making between the ChatGPT intervention cohort compared to those students using traditional methods based on guidelines. However, it is important to note that this sample size was small, with only 16 students, and as such the presence of upskilling inhibition cannot be excluded in this study.

In the final study by Nwe et al, the effectiveness of AI in medical education was based on respondents' feedback collected through an online questionnaire conducted using Google Forms.<sup>21</sup> A large proportion of the surveyed students in that study reported apprehension about over-reliance on technology (83.7%) and the potential loss of critical thinking skills (78.8%). These concerns were widespread despite concurrent reports of perceived benefit, with 76.1% rating AI as effective for problem solving and 66.8% for enhancing critical thinking. This coexistence of high perceived usefulness and high levels of concern suggests that students recognise AI as a valuable educational tool, yet remain aware of its potential risks. Although AI may support learning and analytical development, there appeared to be apprehension that excessive reliance could compromise the development of independent reasoning and autonomous clinical judgement during early medical training. Concerningly, only

3.3% of respondents selected “other” concerns, which included issues such as false or inaccurate information. Given the documented phenomenon of AI-generated misinformation and hallucinated content, the relatively low proportion of students identifying this issue may suggest that risks related to output accuracy may be under-recognised by medical students. This is notable, as uncritical acceptance of inaccurate AI-generated information may undermine opportunities to develop critical appraisal and clinical reasoning skills, with potential downstream implications for patient safety.

It is important to interpret the findings from that study by Nwe et al with appropriate caution. As a cross-sectional questionnaire study, the data reflect subjective student perceptions rather than objectively measured performance outcomes. Furthermore, the survey did not report the use of a validated measurement tool or provide clear definitions of constructs such as critical thinking or decision-making. Consequently, participants’ responses may have reflected individual interpretations of these terms rather than standardised or objectively defined measures of cognitive skills.

Our identified studies also found some evidence to suggest upskilling inhibition with AI in medical students, although this is more challenging to demonstrate, as this requires demonstration of reduced higher-order or advanced skill acquisition when exposed to new tools. In the study by Li et al, there was evidence of impairment in higher-order reasoning, with students perceiving AI as more limited in handling complex case reasoning and innovative thinking as shown in [Table 3](#) and [Table 5](#).<sup>18</sup> The AI intervention improved exam performance, but ratings of advanced cognitive skills remained low, indicating some inhibition of upskilling beyond basic knowledge acquisition required for exams. Limitations in display of more advanced cognitive skills by medical students studied by Li et al may not only be related to generic effects of using AI to learn but may also reflect the limitations of AI itself to produce optimal answers when used for teaching. For example, ChatGPT, a comparable generative AI tool to the Kimi Chat 2.0 used by Li et al, has previously been shown to have limitations specifically in higher order thinking outputs in clinical biochemistry, the subject area assessed in their study.<sup>24</sup> Importantly, the study by Li et al measured student perceptions rather than objectively assessed higher-order reasoning tasks. Potential student biases and pre-existing assumptions about AI’s effect on higher-order reasoning could have influenced these results although the authors did use a pre-study questionnaire to screen out students with extreme positive or negative existing views on the use of AI tools on the medical biochemistry course.<sup>18</sup>

Çiçek et al only reported significant differences in key feature scores between AI and expert feedback groups when considering the more complex case scenarios when comparing immediate retention with delayed performance 10 days later.<sup>19</sup> However, even the delayed test was only 10 days later, which may have been too short to identify significant retention differences between groups for less complex cases whilst the CEQ format used in this study may not have fully assessed higher-order reasoning.

Finally, in the article by Montagna et al, guideline-based systems outperformed AI in all domains studied, achieving higher accuracy and completeness compared to ChatGPT.<sup>20</sup> AI use did not show enhancement of advanced diagnostic reasoning or evidence of improved decision-making, which could be consistent with upskilling inhibition. This study could be critiqued, however, for comparing AI with formal guidelines, as guidelines are designed for accuracy and may serve as an effective gold standard for diagnostic approaches. As such, the measured differences may not be attributable solely to inhibition of students’ ability to learn advanced reasoning.

Our identified studies have shown that AI may be considered a double-edged sword in medical education: effective for basic knowledge acquisition yet potentially less reliable for advanced clinical reasoning. While AI clearly offers a personalised learning environment - allowing learners to input prompts and receive rapid feedback<sup>14</sup> - concerns regarding cognitive erosion, particularly through over-reliance on AI, have been increasingly emphasised.<sup>1,8</sup> Such concerns are especially pertinent in undergraduate medical education, where foundational reasoning skills are still developing. These risks can be conceptualised within a framework comprising two mechanisms: deskilling and upskilling inhibition.<sup>11</sup> Deskilling refers to the erosion of previously acquired competencies due to reduced practice, whereas upskilling inhibition denotes missed opportunities to develop novel skills when tasks are delegated to AI rather than undertaken by the learner.

The studies we identified provide some evidence for both deskilling and upskilling inhibition when AI tools are used in medical student education. These may occur when medical students bypass cognitively demanding processes - such as generating differential diagnoses or interpreting ambiguous findings - in favour of rapid AI-generated solutions. These concerns can be further understood through cognitive theory, particularly dual-process theory. Experienced clinicians employ clinical reasoning through an interplay of intuitive, rapid System 1 thinking and slower, analytical System 2

processes.<sup>9</sup> With expertise, practitioners fluidly transition between these systems, enabling pattern recognition and deliberate analysis.<sup>10</sup> However, the ready availability of AI tools offering apparent System 1-type answers may impede the development of System 2 reasoning, limiting opportunities to consolidate higher-order skills. Over time, this could impair diagnostic flexibility and the ability to individualise care - hallmarks of expert clinical practice. One positive counterpoint emerged from the study by Çiçek et al, which reported that AI exposure increased students' critical appraisal of AI outputs, suggesting potential metacognitive benefits.<sup>19</sup> However, this effect appeared contingent upon deliberate instructional design. In contrast, unstructured or excessive reliance on AI may risk fostering passive learners who accept algorithmic outputs without verification - a phenomenon documented in other cognitive domains.<sup>11</sup>

The integration of AI into medical curricula and its adoption by medical students is approaching near-universal levels.<sup>14</sup> Therefore, attempts to restrict its use may be impractical. Despite this, from an educational perspective, AI should be regarded as a complement rather than a replacement for traditional learning strategies. Appropriate safeguards are essential to prevent reliance on shortcut learning and to promote independent reasoning. Students should be explicitly informed about the limitations and potential risks associated with AI outputs at the outset of their medical training. Although the most conspicuous examples of AI "hallucinations", such as fabricated references, appear to be diminishing with technological advancements, they remain a genuine phenomenon that learners must understand.<sup>25</sup> Many medical students remain unaware of the severe academic consequences of plagiarism.<sup>26</sup> Additional education in this important area represents an opportunity to reinforce the dangers of excessive dependence on AI.

Medical curricula should also emphasise professional responsibility, particularly the obligation to verify algorithmic recommendations generated by AI systems. This principle is consistent with guidance issued by the World Health Organization on the ethical use of AI in health education.<sup>7</sup> Of paramount importance is the potential risk to future patients: AI-generated responses cannot be fully relied upon, and excessive dependence without critical appraisal may ultimately compromise patient safety.

Theoretical cognitive-science research may provide a grounding on how AI use by medical students could be limited in specific ways to enhance rather than inhibit the development of critical thinking. Cognitive Load Theory, for example, describes how students build expertise when instructional methods reduce unnecessary cognitive load while preserving sufficient challenge (the germane load) to promote meaningful learning.<sup>27</sup> Applied to AI-supported education, AI systems could be constrained so that they do not provide complete or polished answers too early so that students still engage fully with the underlying reasoning processes. One practical approach could be to require that each student generative AI prompt included a standard instruction prohibiting the generation of fully formed solutions, obliging the generative AI tool to omit key steps that the learner must supply. As students progress with their learning, even that moderated form of AI assistance could then be gradually withdrawn, requiring learners to assume increasing responsibility for clinical reasoning. Such an approach would align with the concept of scaffolding based on Vygotsky's Zone of Proximal Development.<sup>28,29</sup> With scaffolding, the student is guided by a "more knowledgeable other" who provides temporary support that is progressively removed as the learner masters more advanced levels of understanding. Traditionally that more knowledgeable other has been an educator, but large language models could potentially assume such a role as well in the future.

## Limitations

This systematic review is subject to several limitations. Despite comprehensive screening, only four studies met the inclusion criteria without triggering any exclusion criteria. For a domain of such importance - given its potential to profoundly influence future patient care - the paucity of research is striking. It is particularly surprising that so few studies have examined the potential for AI to cause harm, such as through deskilling or upskilling inhibition, among medical students. The small number of eligible studies, combined with heterogeneity in AI applications precludes direct comparisons or meta-analysis.

The demographic distribution shown in Table 5 of the included studies - three of which were conducted in Asia and none from predominantly English-speaking countries - is in contrast to a long recognised over representation of medical education literature from North America and Europe.<sup>30</sup> However, this simultaneously underscores the lack of representation from many other regions worldwide in this important area of research. Consequently, caution is warranted when extrapolating these findings globally. Furthermore, the evidence presented focuses exclusively on short-term outcomes.

Given that medical education spans approximately five years at medical school and continues throughout professional practice, long-term studies examining skill retention and independent clinical judgement are urgently needed.

## Future Research Directions

Future research should prioritise longitudinal studies to evaluate the sustained impact of AI exposure on clinical competence. Greater involvement of medical students themselves in study design and the inclusion of their perspectives on autonomy, confidence, and trust in AI would enhance the relevance and applicability of findings. It is often assumed that contemporary learners enter medical school with extensive prior exposure to AI tools, a factor that may contrast sharply with the experience of many senior educators currently conducting research. Failure to account for this potential generational disparity risks producing misleading conclusions. A major potential limitation for any future study in this field concerns the adoption of adequate control groups. The use of AI by medical students and university students in general is already pervasive.<sup>14</sup> A recent major study including over 23,000 higher education students from six different continents demonstrated that over 70% of them had used ChatGPT.<sup>31</sup> Therefore careful attention to access to AI for groups in future studies will be important to allow adequate comparison between control and intervention groups. Further investigation into deskilling and upskilling inhibition remains critical, given the profound implications of these phenomena and the current paucity of empirical data. Even where evidence currently exists (such as in the three studies identified in our systematic review), updated study results may be needed to assess subsequent updates in generative AI tools used. For example, two of the studies we identified used ChatGPT-3.5, a version which has already been superseded by newer models. Finally, research encompassing geographically diverse cohorts and incorporating cross-cultural analyses would strengthen the interpretability and generalisability of future findings in this important field.

## Conclusion

AI-supported learning offers demonstrable benefits in terms of efficiency and foundational knowledge acquisition for medical students. However, evidence for its role in fostering higher-order reasoning and complex diagnostic accuracy appears less robust, with some limited evidence that traditional expert-led teaching results in superior outcomes in these domains.

The deliberate integration of safeguards within medical curricula may help mitigate unintended consequences of AI used by medical students, such as deskilling and upskilling inhibition. As AI-supported systems increasingly influence how such students engage with clinical problems, clear pedagogical boundaries are required to ensure that gains in efficiency do not occur at the expense of independent reasoning and reflective judgement. Specific technical approaches have already been proposed for such safeguards. Thurzo et al, for example, describes a Trustworthy Ethical Firewall Architecture for medical and educational AI systems, incorporating formal ethical constraints, transparent audit mechanisms, and structured human-oversight escalation protocols. Such frameworks offer a practical model for aligning AI-enabled learning with educational objectives while preserving human responsibility.<sup>32</sup>

Substantially more research is required to inform evidence-based strategies for the safe and effective incorporation of AI into medical education. This should consider not only immediate learning outcomes but also long-term implications for future patient care. This need is increasingly urgent, as the pace of AI development and adoption within educational settings continues to outstrip its associated curricular adaptation and formal governance. The design and value structures of advanced AI systems are largely shaped by a small number of external actors, leaving educators and training institutions with limited control over technologies that are rapidly becoming embedded in learning environments.<sup>2</sup> Without timely development of the corresponding oversight, medical education risks responding reactively to technological change rather than actively shaping AI use in a manner that safeguards professional formation and public trust.

## Abbreviations

AI, Artificial Intelligence; LMS, Learning Management Software; CEQ ContExtended Questions; CPG, Clinical Practice Guidelines; UTI, Urinary Tract Infections; CPG, Clinical Practice Guidelines; OR, Online repositories.

## Disclosure

The authors report no conflicts of interest in this work.

## References

- Natali C, Marconi L, Dias Duran LD, Cabitza F. AI-induced deskilling in medicine: a mixed-method review and research agenda for healthcare and beyond. *Artificial Intelligence Rev.* 2025;58(11):356. doi:10.1007/s10462-025-11352-1
- Thurzo A. How is AI transforming medical research, education and practice? *Bratislava Med J.* 2025;126(3):243–248. doi:10.1007/s44411-025-00063-2
- Araújo B, Gomes SF, Ribeiro L. Critical thinking pedagogical practices in medical education: a systematic review. *Front Med.* 2024;11:1358444. doi:10.3389/fmed.2024.1358444
- Durning SJ, Jung E, Kim DH, Lee YM. Teaching clinical reasoning: principles from the literature to help improve instruction from the classroom to the bedside. *Kor J Med Educ.* 2024;36(2):145–155. doi:10.3946/kjme.2024.292
- Ahmed Abdalhalim AZ, Nureldaim Ahmed SN, Dawoud Ezzelarab AM, et al. Clinical impact of artificial intelligence-based triage systems in emergency departments: a systematic review. *Cureus.* 2025;17(6):e85667. doi:10.7759/cureus.85667
- Liu X, Faes L, Kale AU, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *Lancet Dig Health.* 2019;1(6):e271–e297. doi:10.1016/S2589-7500(19)30123-2
- World Health Organisation. Ethical, social, and professional implications of large language models in medical and nursing education. 2024. Available from: <https://iris.who.int/server/api/core/bitstreams/e9e62c65-6045-481e-bd04-20e206bc5039/content>. Accessed March 27, 2026
- Zhai C, Wibowo S, Li LD. The effects of over-reliance on AI dialogue systems on students' cognitive abilities: a systematic review. *Smart Learn Environ.* 2024;11(1):28. doi:10.1186/s40561-024-00316-7
- Croskerry P. A universal model of diagnostic reasoning. *Acad Med.* 2009;84(8):1022–1028. doi:10.1097/ACM.0b013e3181ace703
- Lambe KA, O'Reilly G, Kelly BD, Curristan S. Dual-process cognitive interventions to enhance diagnostic reasoning: a systematic review. *BMJ Qual Saf.* 2016;25(10):808–820. doi:10.1136/bmjqs-2015-004417
- Rinta-Kahila T, Penttinen E, Salovaara A, Soliman W, Ruissalo J. The vicious circles of skill erosion: a case study of cognitive automation. *J Assoc Inform Syst.* 2023;24:1378–1412. doi:10.17705/1jais.00829
- Eva KW. What every teacher needs to know about clinical reasoning. *Med Educ.* 2005;39(1):98–106. doi:10.1111/j.1365-2929.2004.01972.x
- Schmidt HG, Rikers RM. How expertise develops in medicine: knowledge encapsulation and illness script formation. *Med Educ.* 2007;41(12):1133–1139. doi:10.1111/j.1365-2923.2007.02915.x
- Simoni J, Urtubia-Fernandez J, Mengual E, et al. Artificial intelligence in undergraduate medical education: an updated scoping review. *BMC Med Educ.* 2025;25(1):1609. doi:10.1186/s12909-025-08188-2
- Bramer WM, Rethlefsen ML, Kleijnen J, Franco OH. Optimal database combinations for literature searches in systematic reviews: a prospective exploratory study. *Syst Rev.* 2017;6(1):245. doi:10.1186/s13643-017-0644-y
- Roberts NW, Christenson RH, Price CP. Searching for evidence: a guide to finding the evidence in laboratory medicine. *Annals Clin Biochem.* 2014;51(3):326–334. doi:10.1177/0004563214521161
- Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ.* 2021;372:n71. doi:10.1136/bmj.n71.
- Li L, Zhang W, Zhang K, et al. The role of generative AI tools in case-based learning and teaching evaluation of medical biochemistry. *BMC Med Edu.* 2025;25(1):1185. doi:10.1186/s12909-025-07567-z
- Çiçek FE, Ülker M, Özer M, Kiyak YS. ChatGPT versus expert feedback on clinical reasoning questions and their effect on learning: a randomized controlled trial. *Postgrad Med J.* 2025;101(1195):458–463. doi:10.1093/postmj/qgae170
- Montagna M, Chiabrando F, De Lorenzo R, Rovere Querini P. Impact of clinical decision support systems on medical students' case-solving performance: comparison study with a focus group. *JMIR Med Educ.* 2025;11:e55709. doi:10.2196/55709
- Nwe TM, Shamshol NAH, Jaafar NN, et al. The role of artificial intelligence in medical training, with its applicability, efficiency, potential, and challenges among preclinical students. article. *Res J Pharm Technol.* 2025;18(6):2508–2516. doi:10.52711/0974-360X.2025.00358
- McLean SF. Case-based learning and its application in medical and health-care fields: a review of worldwide literature. *J Med Educ Curricular Dev.* 2016;3:JMECD.S20377. doi:10.4137/jmeed.S20377
- Bordage G, Page G. The key-features approach to assess clinical decisions: validity evidence to date. *Adv Health Sci Educ.* 2018;23(5):1005–1036. doi:10.1007/s10459-018-9830-5
- Ghosh A, Bir A. Evaluating ChatGPT's ability to solve higher-order questions on the competency-based medical education curriculum in medical biochemistry. *Cureus.* 2023;15(4):e37023. doi:10.7759/cureus.37023
- Eysenbach G. The role of chatgpt, generative language models, and artificial intelligence in medical education: a conversation with ChatGPT and a call for papers. *JMIR Med Educ.* 2023;9:e46885. doi:10.2196/46885
- Pavlovic A, Rajovic N, Masic S, et al. Assessing attitudes toward research and plagiarism among medical students: a multi-site study. *Phil Ethics Human Med.* 2024;19(1):11. doi:10.1186/s13010-024-00161-z
- Young JQ, Van Merriënboer J, Durning S, Ten Cate O. Cognitive load theory: implications for medical education: AMEE Guide No. 86. *Med Teach.* 2014;36(5):371–384. doi:10.3109/0142159x.2014.889290
- Wood D, Bruner JS, Ross G. The role of tutoring in problem solving. *J Child Psychol Psych.* 1976;17(2):89–100. doi:10.1111/j.1469-7610.1976.tb00381.x
- Vygotsky LS, Cole M, John-Steiner V, Scribner S, Souberman E. *Mind in Society: The Development of Higher Psychological Processes*. Harvard University Press; 1978.
- Wondimagegn D, Whitehead CR, Cartmill C, et al. Faster, higher, stronger - together? A bibliometric analysis of author distribution in top medical education journals. *BMJ Global Health.* 2023;8(6):e011656. doi:10.1136/bmjgh-2022-011656
- Ravšelj D, Keržič D, Tomažević N, et al. Higher education students' perceptions of ChatGPT: a global study of early reactions. *PLoS One.* 2025;20(2):e0315011. doi:10.1371/journal.pone.0315011
- Thurzo A. Provable AI ethics and explainability in medical and educational ai agents: trustworthy ethical firewall. *Electronics.* 2025;14(7):1294.

**Advances in Medical Education and Practice**

### **Publish your work in this journal**

Advances in Medical Education and Practice is an international, peer-reviewed, open access journal that aims to present and publish research on Medical Education covering medical, dental, nursing and allied health care professional education. The journal covers undergraduate education, postgraduate training and continuing medical education including emerging trends and innovative models linking education, research, and health care services. The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <http://www.dovepress.com/advances-in-medical-education-and-practice-journal>

**Dovepress**  
Taylor & Francis Group