Original Paper

How Medical Students Use AI for Learning: Cross-Sectional, Survey-Based Quality Improvement Evaluation

Rebekah Cole[1], MEd, PhD; Paul A Hemmer[1], MACP, MPH, MD; Lisa Moores[1], MACP, MD; David Mears[1], PhD; Anita Samuel[1], PhD; Christina La Croix[1], DO; Elizabeth Pearce[2], MA; Eric Elster[1], MD

[1]School of Medicine, Uniformed Services University of the Health Sciences, Bethesda, MD, United States
[2]Henry Jackson Foundation for the Advancement of Military Medicine, Bethesda, MD, United States

Corresponding Author:

Rebekah Cole, MEd, PhD
School of Medicine, Uniformed Services University of the Health Sciences
4301 Jones Bridge Road
Bethesda, MD 20814
United States
Phone: 1 301-295-3017
Email: rebekah.cole@usuhs.edu

Abstract

Background: AI is increasingly shaping medical student learning, yet institutions have limited data to guide the integration of these tools into medical education. As part of a quality improvement initiative, we evaluated students' AI use to inform institutional transformation toward AI-supported competency-based medical education.

Objective: We aimed to characterize how first-year medical students use AI; identify opportunities for improving responsible AI-supported learning; and guide curricular, faculty development, and policy initiatives.

Methods: We administered a program evaluation survey to first-year medical students following a neuroscience module. The survey assessed students' AI use, perceived effects on learning, and verification practices. To contextualize the findings, we conducted an exploratory cohort-level comparison of exam performance with that of prior cohorts.

Results: Of 186 students, 171 (91.9%) reported using AI. Students most commonly used AI for exam preparation, content review, concept clarification, and practice question generation. Most students perceived AI as improving study efficiency (152/171, 88.9%) and helping in promoting understanding of difficult concepts (148/171, 86.5%), and 82.5% (141/171) agreed that they used AI tools to actively test their understanding. However, 21.6% (37/171) rarely or never verified AI-generated information, pointing to an important target for educational improvement. Cohort-level exam performance was similar to or modestly higher than that of prior cohorts.

Conclusions: This quality improvement initiative found that AI was already deeply embedded in students' learning and identified verification and responsible AI use as priorities for curricular planning, faculty development, and institutional guidance within the broader transition toward AI-supported competency-based medical education.

JMIR Med Educ 2026;12:e109763; doi: 10.2196/109763

Keywords: artificial intelligence; AI; undergraduate medical education; medical students; generative AI; self-regulated learning; strategic planning; quality improvement

Introduction

Background

AI technologies are increasingly being integrated across the medical education continuum, from admissions processes to teaching, learning, and assessment [1-3]. AI has been used to support personalized learning, provide real-time feedback, create curricular content, facilitate assessment, and power innovations such as virtual and robotic standardized patients [1-3]. The emergence of large language models has further expanded the reach of these technologies [1,4]. Tools such as ChatGPT, Google Gemini, Claude, and Microsoft Copilot have made AI more accessible than ever, and medical students are increasingly incorporating these tools into their learning practices [1,4-6].

As access to AI expands, these technologies are reshaping how students study, learn, and engage with educational

content [4-6]. This shift raises questions about the nature and quality of students' learning, including concerns about meaningful learning, cognitive offloading, information verification, and the validity of existing assessment approaches in AI-supported learning environments [7-10].

The Liaison Committee on Medical Education requires medical schools to engage in continuous quality improvement and ongoing curricular evaluation [11]. Given the growing influence of AI on student learning, there is an increasing need for medical schools to evaluate how AI affects learning processes and educational outcomes [2,11]. Equally important is the need to understand students' authentic learning behaviors rather than relying on assumptions or speculation about how AI may be used [2,5,6]. Such evidence is essential for informing educational policy, curricular design, assessment practices, and faculty development efforts in an increasingly AI-supported educational environment.

Competency-Based Medical Education as a Framework for AI-Supported Learning

As students' AI use continues to expand, competency-based medical education (CBME) provides a framework for integrating AI into learning, assessment, and competency development [12-14]. The relationship is mutually reinforcing: CBME guides the development of outcomes, curriculum, and assessment, while AI helps operationalize key CBME principles such as individualized learning, continuous feedback, programmatic assessment, and data-informed progression decisions through adaptive learning, learning analytics, and personalized support [13-17].

Several educational theories suggest that AI can function as more than an information resource. Self-regulated learning theory highlights AI's potential to help learners identify knowledge gaps, monitor progress, receive feedback, and set learning goals [18,19]. Adaptive and learner-centered theories similarly suggest that AI can personalize content, feedback, and learning pathways while supporting autonomy and individualized development [16,17,20].

Institutional Transition Toward AI-Supported CBME

Given these symbiotic intersections between AI and CBME, the Uniformed Services University (USU) School of Medicine launched a strategic planning initiative in November 2025 to guide its transition toward AI-supported CBME. This quality improvement effort sought to better understand how students were using AI to support their learning and generate data that could inform future curriculum redesign and educational planning. For other schools, accreditation and reaccreditation cycles may provide a natural opportunity to undertake similar curriculum transformation efforts, as these processes often prompt comprehensive review of educational outcomes, competencies, and assessment systems.

As we began our strategic planning efforts, we found that prior studies had documented medical students' use of and attitudes toward generative AI, including perceived benefits and concerns [1,5,6,8]. However, less is known about how students incorporate AI into their day-to-day learning, including their learning workflows, prompting practices, and verification of AI-generated information.

To address this knowledge gap and inform the broader medical education community, we conducted a quality improvement study examining patterns of AI use among medical students. The objective of our study was to characterize how students integrate AI into their learning workflows, contextualize cohort-level academic performance during a period of widespread AI use, and generate insights to inform the design of an AI-supported CBME curriculum.

Methods

Overview

We conducted a cross-sectional, survey-based quality improvement evaluation within a single undergraduate medical education program. Reporting was guided by the SQUIRE (Standards for Quality Improvement Reporting Excellence) 2.0 guidelines [21], with applicable items adapted to this baseline, survey-based evaluation within a broader educational quality improvement initiative. The curriculum consists of a 16-month preclerkship phase organized into integrated, systems-based modules, followed by a 12-month clerkship phase and a 16-month postclerkship phase [22]. The evaluation was implemented during the 7-week neuroscience module, the fourth of 7 first-year preclerkship modules. The preclerkship phase was selected to characterize AI use early in medical school and inform subsequent curricular planning. We selected the neuroscience module in particular because its timing in the academic calendar aligned with the designated period for our program evaluation.

USU has partnered with Google since June 2024, providing all students and faculty with institutionally supported access to a suite of generative AI tools, including NotebookLM, Gemini, and Google Workspace. The class of 2027 completed the neuroscience module in March 2024, before these tools became available. The classes of 2028 and 2029 had access to the institutionally supported tools during their respective neuroscience modules. Students are also permitted to use commercially available AI tools of their choosing. Preclerkship course content is delivered through the Canvas learning management system, with recorded lectures housed in Panopto.

Relevant assessment methods within the preclerkship curriculum include National Board of Medical Examiners (NBME) Customized Assessment Services exams, administered as 80- to 100-item midterm and final exams, as well as faculty-developed exams consisting of 30- to 50-item single-best-answer multiple-choice exams.

This evaluation was designed to characterize patterns of AI use among first-year medical students and inform institutional strategic planning related to CBME implementation. Data were collected using a Google Forms survey (Multimedia Appendix 1) administered at the conclusion of

the 7-week neuroscience module during the spring of 2026. Survey responses were deidentified before analysis to protect students' confidentiality.

Ethical Considerations

USU's Institutional Review Board determined that this activity did not meet the criteria defining research and classified it as program evaluation (DBS.2026.1093). Because the project was conducted as a program evaluation for quality improvement purposes rather than as human subject research, informed consent was not required. Survey data used for this evaluation were deidentified to protect participant privacy and confidentiality. The students received no compensation for their completion of the survey. The evaluation was conducted in accordance with the ethical principles outlined in the Declaration of Helsinki.

Institutional Improvement Context

This survey represented the first phase of a broader quality improvement initiative supporting USU's transition toward AI-supported CBME. To guide this effort, the dean convened an AI Curriculum Reform Task Force composed of PhD- and MD-level faculty charged with understanding evolving student AI use and informing the institution's transition to a competency-based educational model.

Our survey sought to establish a baseline understanding of student AI use and inform subsequent initiatives related to curriculum redesign, faculty development, assessment modernization, AI literacy, and the development of AI-supported educational systems. As such, this evaluation served as a foundational step in the School of Medicine's broader transformation toward a more adaptive, data-informed, competency-based learning environment.

Participants

Participants were first-year medical students (N=186) who completed the 7-week neuroscience module during the spring 2026 semester. The neuroscience module is the fourth of 7 modules in the preclerkship curriculum at USU. In 2026, the module was shortened by 3 instructional days compared with previous years. At the time of survey administration, students were approximately 10 months into medical school and had previously completed the foundation, musculoskeletal, and cardiopulmonary-renal modules.

Instrument

Our survey, the first phase of the broader institutional quality improvement initiative, was developed using a modified Delphi process involving basic science and clinical faculty. After iterative refinement, the survey was piloted with first- and fourth-year medical students to evaluate clarity, relevance, and content validity. The final instrument (Multimedia Appendix 1) included 22 closed-ended items and 6 free-text items addressing AI use; learning workflows; perceived educational impact; verification behaviors; and AI literacy, including prompting skills. Of the 6 free-text items, 4 were substantive open-ended questions, and 2 collected contact information for optional follow-up activities. Responses provided by students to 2 open-ended questions were included in the thematic analysis: one asking AI users about the impact of AI tools on their learning (n=81) and one asking about helpful prompting strategies (n=52). Responses to the question about future AI training for students who reported no AI use were not included in the thematic analysis. Responses to the question about additional training to improve AI literacy were also excluded because only 14 students responded.

The survey was administered through Google Forms as part of the required end-of-module evaluation. The screening question "Did you use AI tools (LLMs such as Gemini, NotebookLM, ChatGPT, Claude) to assist your studying during the module?" was completed by all 186 enrolled students (100%). Of these 186 students, 15 (8.1%) reported not using AI to support their studying during the module and were therefore excluded from the full survey analysis. An administrative staff member compared automatically collected email addresses to identify duplicate submissions. Five students completed the survey twice; for each, the initial submission was excluded, and only the final submission was retained for analysis. The email addresses were used solely to identify duplicates and were removed after data cleaning. Administrative staff removed all identifying information before providing the survey data to the research team; therefore, the research team analyzed only deidentified responses and could not link individual students to their survey responses. Students could voluntarily provide their name and contact information if they wished to participate in follow-up activities.

Comparative Cohort Data (Classes of 2026, 2027, 2028, and 2029)

Because survey responses were deidentified, AI use was not linked to individual student performance. Instead, we conducted an exploratory cohort-level comparison between the study cohort and 3 earlier cohorts. The classes of 2026, 2027, 2028, and 2029 completed the neuroscience module in spring 2023, 2024, 2025, and 2026, respectively. Although the class of 2028 had initial access to institutionally supported AI tools, AI use was not measured in earlier cohorts. Analyses included performance on faculty-developed and NBME Customized Assessment Services midterm and final exams. These comparisons were intended to contextualize academic performance during a period of widespread AI use rather than estimate the effect of AI use on academic performance.

Data Analysis

We analyzed the quantitative data using descriptive statistics, including frequencies, percentages, and response distributions for all survey items. Comparative analyses using ANOVA were conducted across class cohorts to examine differences in academic performance. We conducted a sensitivity analysis excluding all examination scores from students who repeated the module to assess whether their inclusion influenced the cohort comparisons. All analyses were performed using Jamovi (version 2.3.28.0).

We assessed the open-ended survey responses using thematic analysis [23]. Two members of the research team (RC and EP) independently reviewed and coded the responses. Related codes were grouped into broader themes reflecting recurring patterns in students' experiences with and perceptions of AI-supported learning. Any coding discrepancies were discussed by the 2 reviewers and resolved through consensus.

Results

Overview

All 186 students enrolled in the neuroscience module completed the initial survey items (response rate=100%). Of these 186 students, 171 (91.9%) reported using AI during the module, completed the full survey, and were included in the analysis. Table 1 summarizes students' AI use patterns, prompting behaviors, verification practices, and self-reported AI literacy.

Table 1. AI use patterns, AI literacy, and prompting behaviors among medical students. Data are from an end-of-module survey of AI use among Uniformed Services University class of 2029 students (n=171).

Domain	Participants, n (%)
Prior AI experience	
Had used AI before the module	164 (95.9)
Moderate to extreme baseline familiarity	156 (91.2)
Current AI literacy	
Intermediate	94 (55)
Beginner	44 (25.7)
Advanced or expert	30 (17.5)
Novice	3 (1.8)
Frequency of use	
Daily	85 (49.7)
Several times per week	65 (38)
Weekly or less	21 (12.3)
Verification of AI output	
Always	14 (8.2)
Often	69 (40.4)
Sometimes	51 (29.8)
Rarely or never	37 (21.6)
AI provided needed information	
Almost always	67 (39.2)
Often	90 (52.6)
Approximately half the time	14 (8.2)
Confidence in prompting	
Somewhat or very confident	115 (67.3)
Neutral	40 (23.4)
Not very confident or not at all confident	16 (9.4)
Typical prompting style	
Some context or background	86 (50.3)
Detailed instructions	51 (29.8)
Simple prompts	34 (19.9)
Need to revise prompts	
Sometimes	97 (56.7)
Often	35 (20.5)
Rarely	32 (18.7)
Always	4 (2.3)
Never	3 (1.8)
Perceived improvement in prompting skills	
Agreed or strongly agreed	82 (48)
Neutral	73 (42.7)

Domain	Participants, n (%)
Disagreed or strongly disagreed	16 (9.4)
Use of AI to write AI prompts	
Strongly agree	17 (9.9)
Agree	38 (22.2)
Neutral	61 (35.7)
Disagree	41 (24)
Strongly disagree	14 (8.2)
Source of AI content used	
Only faculty-provided content	75 (43.9)
Mostly faculty-provided content with some third-party content	54 (31.6)
An equal mix of faculty-provided and third-party content	32 (18.7)
Mostly third-party content	8 (4.7)
Only third-party content	2 (1.2)
Helpfulness of faculty-curated AI tools	
Very helpful	62 (36.3)
Somewhat helpful	47 (27.5)
Neutral	50 (29.2)
Not very helpful or not at all helpful	12 (7)

Reasons for Not Using AI

Among the 15 students who reported no AI use, the most common reason was a preference for their usual study methods (n=12, 80%). Other reasons included concerns about overreliance on AI (n=6, 40%), perceptions that the tools were unnecessary for learning (n=4, 26.7%), lack of familiarity with the tools (n=1, 6.7%), uncertainty about whether AI use was appropriate (n=1, 6.7%), and technical difficulties (n=1, 6.7%). In total, 6.7% (1/15) of the students selected "Other" and clarified that they used AI to build knowledge but not to review material.

Types and Purposes of AI Use

Students reported using AI across a wide range of learning activities, with most relying primarily on faculty-provided materials and using AI for content review, concept clarification, practice question generation, and exam preparation. Additional characteristics of AI use are shown in Table 2.

Table 2. Characteristics of AI use among medical students from responses to an end-of–preclerkship module survey regarding the use of AI related to the curriculum, Uniformed Services University class of 2029 students (n=171)[a].

Domain and most common responses	Participants, n (%)
Primary reasons for AI use	
Creating practice questions	123 (71.9)
Reviewing or summarizing complex material	122 (71.3)
Clarifying concepts	122 (71.3)
Identifying high-yield topics	99 (57.9)
Saving time	98 (57.3)
AI tools used	
Gemini	132 (77.2)
Faculty-developed NotebookLM notebooks	100 (58.5)
Student-created NotebookLM notebooks	103 (60.2)
ChatGPT	46 (26.9)
Canvas Gemini Virtual Patient	27 (15.8)
Study activities supported by AI	
Exam preparation	147 (86)
Review of course material	139 (81.3)
Lecture review	126 (73.7)
Lecture preparation	70 (40.9)
Study groups	32 (18.7)

Domain and most common responses	Participants, n (%)
Ways in which students interacted with AI	
Asking basic questions	110 (64.3)
Self-assessment	97 (56.7)
Primarily reading outputs	83 (48.5)
Iteratively refining prompts	80 (46.8)
Critically evaluating outputs	53 (31)
Asking AI to identify knowledge gaps	45 (26.3)

[a]Percentages may exceed 100% because participants could select multiple responses.

Perceived Impact of AI on Learning

Students generally perceived AI tools as beneficial to their learning. Most reported that AI helped them understand difficult concepts, made studying more efficient, and helped them identify knowledge gaps. Additionally, 82.5% (141/171) agreed that they used AI tools to actively test their understanding (Table 3).

Table 3. Student perceptions of the educational value of AI tools from responses to an end-of–preclerkship module survey regarding the use of AI related to the curriculum, Uniformed Services University class of 2029 students (n=171).

Statements	Strongly disagree, n (%)	Disagree, n (%)	Neutral, n (%)	Agree, n (%)	Strongly agree, n (%)	Agree or strongly agree, n (%)
AI tools helped me understand difficult concepts.	4 (2.3)	4 (2.3)	15 (8.8)	82 (48)	66 (38.6)	148 (86.5)
AI tools made my studying more efficient.	4 (2.3)	3 (1.8)	12 (7)	66 (38.6)	86 (50.3)	152 (88.9)
AI tools helped me identify gaps in my knowledge.	5 (2.9)	5 (2.9)	22 (12.9)	70 (40.9)	69 (40.4)	139 (81.3)
I used AI tools to actively test my understanding.	4 (2.3)	5 (2.9)	21 (12.3)	56 (32.7)	85 (49.7)	141 (82.5)

Common challenges included difficulty writing effective prompts (72/171, 42.1%), inaccurate AI-generated information (40/171, 23.4%), tools being time-consuming to use (28/171, 16.4%), concerns about trustworthiness (25/171, 14.6%), and difficulty using AI effectively (18/171, 10.5%). A smaller number of students reported preferring other study methods (10/171, 5.8%), inconsistent outputs (2/171, 1.2%), or a steep learning curve associated with effective AI use (3/171, 1.8%). Additionally, 15.8% (27/171) of the students stated that they experienced no challenges using AI.

Qualitative Findings

Thematic analysis of open-ended responses identified 4 major themes (Table 4): AI as a cognitive partner, perceived efficiency and study value, prompt engineering as an emerging literacy, and verification and responsible use. Only 14 students responded to the free-text item on additional AI literacy training needs; these responses were not included in the thematic analysis. Collectively, these findings suggest that students view AI as a valuable learning resource while recognizing the importance of effective prompting, critical evaluation, and appropriate use.

Table 4. Themes identified from open-ended survey responses from qualitative responses to an end-of–preclerkship module survey regarding the use of AI related to the curriculum, Uniformed Services University class of 2029 students.

Theme	Description	Representative responses
AI as a cognitive partner	AI used as a tutor, thought partner, and source of personalized explanations and clarifications	"A specialized tutor in your pocket;" "Safety backboard;" "Endless question bank"
Perceived efficiency and study value	Students' perceptions of time savings and more effective studying	"Life changing;" "Game changer;" "Essential;" "Saves me so much time;" "Makes my study more effective"
Prompt engineering as an emerging literacy	Students emphasized specificity, context, and iterative prompt refinement	"Being as specific as possible;" "Trial and error;" "Using AI to help hone my prompts"
Verification and responsible use	Students recognized limitations and the need to cross-check outputs	"AI can hallucinate;" "I cross check;" "AI often requires correction"

Cohort-Level Comparison of Exam Performance Across the Classes of 2026 to 2029

To contextualize academic performance during a period of widespread AI use, we conducted an exploratory comparison of exam scores for the class of 2029 with those of the classes of 2026 to 2028 using ANOVA. These analyses compared successive cohorts rather than AI-exposed and AI-unexposed groups as the class of 2028 also had access to institutionally supported AI tools. Significant cohort differences were examined using Tukey post hoc comparisons.

Twenty-three students repeated their first year after failing: 8 originally in the class of 2026 joined the class of 2027, 6 originally in the class of 2027 joined the class of 2028, and 9 originally in the class of 2028 joined the class of 2029. For students who repeated the neuroscience module after extending their training by 1 year, we excluded their first-attempt scores and analyzed only their repeat-attempt scores to preserve the independence of observations. At the time of analysis, no students taking the neuroscience module for the first time with the class of 2029 had declared an extension of training that would require them to repeat the module. Two students in the class of 2026 began the module but did not take the final exams, affecting the df for the 2 final exam comparisons.

As shown in Table 5, the class of 2029 demonstrated higher performance on the faculty-developed midterm and the NBME final, although the observed effect sizes were small to moderate (η^2_p=0.024-0.060), and differences were not observed for all exams. Excluding both first- and repeat-attempt scores for students who repeated the module produced minimal changes in η^2_p values (absolute changes of 0.001-0.006) and did not change the statistical significance of the results.

Table 5. Comparison of midterm and final exam performance across student cohorts (classes of 2026-2029).

Exams	Class of 2026 (n=163), mean (SD)	Class of 2027 (n=174), mean (SD)	Class of 2028 (n=172), mean (SD)	Class of 2029 (n=186), mean (SD)	Between-class differences (ANOVA)
NBME[a] midterm (0-100)	84.67 (6.97)	84.66 (6.84)	84.29 (6.37)	83.83 (6.64)	No significant differences ($F_{3,691}$=0.62; P=.61; η^2_p=0.003)
Faculty-developed midterm (0-100)	79.10 (7.35)	79.19 (8.33)	78.39 (7.98)	83.54 (9.04)	Class of 2029>all prior cohorts ($F_{3,691}$=14.78; P<.001; η^2_p=0.060)
NBME final (0-100)	84.89 (7.60)	84.84 (8.24)	85.86 (6.14)	87.58 (6.84)	Class of 2029>classes of 2026 and 2027 ($F_{3,689}$=5.6; P<.001; η^2_p=0.024)
Faculty-developed final (0-100)	77.51 (11.11)	80.80 (10.02)	80.22 (9.14)	80.16 (8.03)	Class of 2026<all subsequent cohorts ($F_{3,689}$=3.89; P=.009; η^2_p=0.017)

[a]NBME: National Board of Medical Examiners.

Discussion

Principal Findings

Our findings indicate that first-year medical students have integrated AI into multiple aspects of their learning. Consistent with emerging reports of widespread AI adoption, students used AI for content review, concept clarification, self-assessment, examination preparation, and creation of personalized study resources [24-27]. These findings align with recent scoping reviews documenting a rapid increase in undergraduate medical students' use of large language models for personalized learning, study support, and examination preparation [1].

Rather than focusing on attitudes or perceived usefulness, this evaluation examined self-reported learning behaviors and suggests that AI is becoming a routine component of medical education. Because performance outcomes were limited to 4 exams, conclusions about the impact of AI on learning should be interpreted cautiously. Cohort-level examination performance showed either no differences or small improvements relative to prior cohorts, although effect sizes were small and may not represent meaningful educational benefits [28]. Exploratory cohort-level comparisons showed no consistent evidence of lower examination performance in the study cohort; however, these comparisons cannot establish the effect of AI use on learning.

The increasing use of AI by medical students presents both opportunities and challenges for medical education. Students used AI to review content, clarify concepts, prepare for examinations, and generate practice questions. These capabilities may improve learning efficiency and allow student-faculty interactions to focus on higher-order skills such as application, synthesis, and clinical reasoning [16,17,24,29]. This widespread adoption aligns with a recent global meta-analysis showing high levels of student enthusiasm and AI use across medical education settings [30]. However, the same study identified a "knowledge-attitude gap," with students exhibiting only moderate understanding of how AI systems function [30].

At the same time, many students reported not routinely verifying AI-generated information. Although NotebookLM may reduce confabulation by grounding responses in faculty-provided materials, AI outputs remain susceptible to errors and omissions [16,31]. These findings highlight the importance of teaching learners to critically evaluate AI-generated content and ensuring that educational resources are accurate and aligned with learning objectives [16,17,31]. The lack of consistent verification mirrors broader findings in medical education, where automation bias can lead learners to accept flawed outputs uncritically. Emerging scholarship has also warned of "never skilling," the risk that overreliance on answer-generating AI may reduce cognitive effort and impede development of the reasoning skills required for independent clinical practice [32].

Finally, our findings align with ongoing discussions about AI's role in medical education and the evolving role of faculty [3,16,17]. Although AI may enhance learning

efficiency and knowledge acquisition, it cannot replace the human interactions that support clinical reasoning, communication, professional identity formation, and patient-centered care [16,33,34]. As AI becomes more integrated into education and practice, educators must balance its benefits with preserving the human elements essential to medicine [3,16,17].

Our evaluation results suggest that students are extensively using AI to support learning and assessment preparation while continuing to rely heavily on faculty-provided materials. These results reinforce the central role of faculty in identifying, organizing, and communicating foundational knowledge within AI-enabled learning environments [33,34]. However, integrating these tools requires addressing differences in familiarity with generative AI as students may be more familiar with these tools than the faculty teaching them [35]. To bridge this gap, modern faculty development might consider embracing a "shared learning model" and interdisciplinary "communities of practice," where educators and learners jointly navigate the limitations of changing technologies [36].

As AI becomes more integrated into medical education, faculty might likewise consider ensuring that educational materials support reliable AI outputs through clear learning objectives, aligned resources, and accurate transcripts. Faculty also need a foundational understanding of AI capabilities and limitations to effectively incorporate these tools into teaching, assessment, and curriculum design [34-36]. Rather than treating AI solely as a reference tool, medical educators must recognize its influence on clinical reasoning and learning processes. Institutions may therefore benefit from faculty development programs and structured supervision frameworks such as DEFT-AI (diagnosis, discussion, evidence, feedback, and teaching), which promote effective prompting, output verification, and reflective AI use [36].

Finally, our findings highlight the need for structured instruction in responsible AI use. As incoming students arrive with increasing AI experience, medical schools should assess baseline AI literacy and provide guidance on effective use, critical evaluation, verification of AI-generated information, and effective prompting strategies [36-38]. However, instruction alone may be insufficient. Institutions might consider developing integrated AI-enabled learning ecosystems that provide learners with access to vetted tools, curated educational resources, and structured supports that promote responsible, efficient, and evidence-informed AI use. Ongoing evaluation of learner needs and educational outcomes will be essential to ensure these educational approaches remain relevant and effective [39].

Developing educational policy for AI is challenging because both the technology and patterns of learner use continue to evolve rapidly [32]. Our findings suggest that policies might focus less on restricting AI use and more on promoting verification, accountability, and responsible use. Students' strong reliance on faculty-provided materials also highlights the value of trusted AI-enabled learning environments. Given the pace of change, AI governance might be viewed as a continuous quality improvement process, with policies regularly reassessed and refined as technologies, learner behaviors, and educational needs evolve.

Our findings may have relevance beyond medical education as educational institutions across disciplines and countries respond to the rapid integration of AI into students' learning practices. AI is increasingly shaping how learners engage with content and direct their own learning [40]. Educational systems therefore need to evolve in ways that help learners use AI effectively and responsibly while maintaining responsibility for their own learning.

Balancing AI-supported learning with the development of independent competence is particularly important in CBME, where AI may enable more individualized learning and progression. As AI becomes increasingly integrated into these processes, medical educators will need to distinguish between performance supported by AI and competence that learners can demonstrate independently. Programmatic assessment must therefore continue to provide credible evidence that learners are developing the capabilities and increasing independence expected for progression [15,41]. Although specific approaches will vary across disciplines, institutions, and educational contexts, understanding how learners actually use AI can provide an important foundation for designing effective and responsible AI-enabled learning environments.

Limitations and Future Directions

Our evaluation has several limitations. First, it was conducted among first-year students at a single medical school during one 7-week neuroscience module. Patterns of student AI use may vary by course content, duration, instructional approach, and assessment structure, limiting the transferability of these findings to other courses and educational settings. Future studies should therefore examine AI-supported learning across courses of varying duration and structure. In addition, although 18.7% (32/171) of AI users reported incorporating AI into study group sessions, our survey did not distinguish individual AI use within a group from collaborative interaction with AI. Future research across institutions, disciplines, stages of training, and learning contexts could help determine which patterns are broadly shared and how individual vs collaborative AI use influences learning.

Our evaluation also relied primarily on self-reported data and, therefore, reflects students' perceptions of their AI use and its educational value rather than objectively measured learning behaviors. Responses may also be subject to recall and social desirability biases [42]. Administration as part of the required end-of-module evaluation may have increased social desirability bias as students may have felt pressure to report practices they perceived as institutionally preferred. As a result, future studies incorporating objective measures of AI use and learning processes are needed to provide a more complete understanding of how students engage with AI and how these behaviors relate to educational outcomes.

In addition, the cohort-level examination comparisons were exploratory and should not be interpreted as evidence of an effect of AI on learning. Curricular and assessment

changes, differences between cohorts, and AI use among earlier cohorts may have influenced performance. The class of 2028 had access to institutionally supported AI tools, and AI use among earlier cohorts was not measured; therefore, these cohorts cannot be considered AI-unexposed comparison groups. The cohort comparisons included repeat-attempt scores from 23 students repeating their first year of medical school. Although excluding their first-attempt scores avoided counting students twice, prior exposure to the module may have influenced their repeat-attempt performance and biased comparisons between cohorts. Moreover, because students frequently used AI with faculty-provided materials, AI-supported study may have been particularly well aligned with faculty-developed exams, providing an alternative explanation for some observed performance differences. Because survey responses were deidentified, individual AI use patterns could not be linked with academic performance. Consequently, observed differences in examination performance cannot be attributed to AI use. Future studies might link individual AI use behaviors with educational outcomes while accounting for differences in curricular and assessment conditions.

Finally, as AI technologies and patterns of learner use continue to evolve, longitudinal and multi-institutional studies are needed to examine how AI-supported learning changes over time and its relationship with competency development, clinical performance, and other longer-term educational outcomes. Our evaluation approach examining AI use patterns, learning activities, prompting and verification behaviors, and perceived effects on learning may provide a feasible framework for studying AI-supported learning across educational settings.

Conclusions

AI has become deeply integrated into medical students' learning. This evaluation demonstrated widespread adoption and provided insights into how students use AI while highlighting challenges related to prompting, verification, and responsible use. Our findings suggest that medical education must move beyond whether students should use AI toward determining how it can be integrated in ways that support learning while preserving critical thinking. Understanding how learners engage with AI can inform the design of curricula, assessment, learner support, faculty development, and institutional guidance that reflect contemporary learning practices. As AI continues to evolve, ongoing evaluation will be essential to ensure that its integration supports competency development and prepares learners for clinical practice.

Acknowledgments

The opinions and assertions expressed herein are those of the authors and do not reflect the official policy or position of the Uniformed Services University of the Health Sciences, the Henry M. Jackson Foundation for the Advancement of Military Medicine, or the Department of War.

Generative AI Use

Generative AI (Google Gemini and ChatGPT [GPT-6; OpenAI]) was used during manuscript preparation to assist with copyediting, language refinement, rephrasing, formatting, and condensation of text. AI tools were not used to generate original scientific content, analyze data, interpret results, or formulate conclusions. All AI-assisted outputs were reviewed, revised, and approved by the authors, who take full responsibility for the accuracy, integrity, and final content of the manuscript.

Funding

This work received no specific grant or external funding from any funding agency in the public, commercial, or not-for-profit sectors. No funding was required to conduct this quality improvement evaluation.

Data Availability

Deidentified data may be available from the corresponding author on reasonable request and with approval from the Uniformed Services University of the Health Sciences.

Authors' Contributions

All authors contributed to conceptualization and methodology. RC facilitated project management. RC and EP conducted formal analysis, qualitative analysis, and data curation. RC, PAH, LM, DM, and AS drafted the original manuscript. PAH, LM, DM, AS, CL, and EE contributed to supervision, interpretation of findings, and critical review and editing of the manuscript. All authors reviewed, revised, and approved the final manuscript and agree to be accountable for all aspects of the work.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Survey assessing medical students' use of and experiences with AI resources during the neuroscience module.
[DOCX File (Microsoft Word File), 21 KB-Multimedia Appendix 1]

References

1. Gordon M, Daniel M, Ajiboye A, et al. A scoping review of artificial intelligence in medical education: BEME guide no. 84. Med Teach. Apr 2024;46(4):446-470. [doi: 10.1080/0142159X.2024.2314198] [Medline: 38423127]

2. Cole R. Teaching medical students to learn effectively in the age of AI. Med Teach. May 11, 2026:1-4. [doi: 10.1080/0142159X.2026.2670569] [Medline: 42112919]
3. Masters K. Artificial intelligence in medical education. Med Teach. Sep 2019;41(9):976-980. [doi: 10.1080/0142159X.2019.1595557] [Medline: 31007106]
4. Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. PLOS Digit Health. Feb 2023;2(2):e0000198. [doi: 10.1371/journal.pdig.0000198] [Medline: 36812645]
5. Alkhaaldi SM, Kassab CH, Dimassi Z, et al. Medical student experiences and perceptions of ChatGPT and artificial intelligence: cross-sectional study. JMIR Med Educ. Dec 22, 2023;9:e51302. [doi: 10.2196/51302] [Medline: 38133911]
6. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. Healthcare (Basel). Mar 19, 2023;11(6):887. [doi: 10.3390/healthcare11060887] [Medline: 36981544]
7. Zawacki-Richter O, Marín VI, Bond M, Gouverneur F. Systematic review of research on artificial intelligence applications in higher education – where are the educators? Int J Educ Technol High Educ. 2019;16:39. [doi: 10.1186/s41239-019-0171-0]
8. Kasneci E, Sessler K, Küchemann S, et al. ChatGPT for good? On opportunities and challenges of large language models for education. Learn Individ Differ. Apr 2023;103:102274. [doi: 10.1016/j.lindif.2023.102274]
9. Crompton H, Burke D. Artificial intelligence in higher education: the state of the field. Int J Educ Technol High Educ. 2023;20:22. [doi: 10.1186/s41239-023-00392-8]
10. Ouyang F, Zheng L, Jiao P. Artificial intelligence in online higher education: a systematic review of empirical research from 2011 to 2020. Educ Inf Technol. Jul 2022;27(6):7893-7925. [doi: 10.1007/s10639-022-10925-9]
11. Functions and structure of a medical school: standards for accreditation of medical education programs leading to the MD degree. Liaison Committee on Medical Education; 2023. URL: https://education.med.wustl.edu/app/uploads/2023/07/2024-25-Functions-and-Structure-of-a-Medical-School.pdf [Accessed 2026-11-12]
12. Frank JR, Snell LS, Cate OT, et al. Competency-based medical education: theory to practice. Med Teach. 2010;32(8):638-645. [doi: 10.3109/0142159X.2010.501190] [Medline: 20662574]
13. Ten Cate O. Competency-based postgraduate medical education: past, present and future. GMS J Med Educ. 2017;34(5):Doc69. [doi: 10.3205/zma001146] [Medline: 29226237]
14. Carraccio C, Englander R, Holmboe ES, Kogan JR. Driving care quality: aligning trainee assessment and supervision through practical application of entrustable professional activities, competencies, and milestones. Acad Med. Feb 2016;91(2):199-203. [doi: 10.1097/ACM.0000000000000985] [Medline: 26509601]
15. van der Vleuten CP, Schuwirth LW, Driessen EW, et al. A model for programmatic assessment fit for purpose. Med Teach. 2012;34(3):205-214. [doi: 10.3109/0142159X.2012.652239] [Medline: 22364452]
16. Cole R, Peacock JG, Samuel A, Cole J, Duncan J. Embedding artificial intelligence competencies in military medical education: a longitudinal framework. Mil Med. Feb 27, 2026:usag068. [doi: 10.1093/milmed/usag068] [Medline: 41758017]
17. Bojic I, Mammadova M, Ang CS, et al. Empowering health care education through learning analytics: in-depth scoping review. J Med Internet Res. May 17, 2023;25:e41671. [doi: 10.2196/41671] [Medline: 37195746]
18. Zimmerman BJ. Becoming a self-regulated learner: an overview. Theory Pract. 2002;41(2):64-70. [doi: 10.1207/s15430421tip4102_2]
19. Panadero E. A review of self-regulated learning: six models and four directions for research. Front Psychol. 2017;8:422. [doi: 10.3389/fpsyg.2017.00422] [Medline: 28503157]
20. Mayer RE. Applying the science of learning to medical education. Med Educ. Jun 2010;44(6):543-549. [doi: 10.1111/j.1365-2923.2010.03624.x] [Medline: 20604850]
21. Ogrinc G, Davies L, Goodman D, Batalden P, Davidoff F, Stevens D. SQUIRE 2.0 (Standards for QUality Improvement Reporting Excellence): revised publication guidelines from a detailed consensus process. BMJ Qual Saf. Dec 2016;25(12):986-992. [doi: 10.1136/bmjqs-2015-004411] [Medline: 26369893]
22. Torre DM, Dong T, Schreiber-Gregory D, et al. Exploring the predictors of post-clerkship USMLE Step 1 scores. Teach Learn Med. 2020;32(3):330-336. [doi: 10.1080/10401334.2020.1721293] [Medline: 32075437]
23. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol. 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
24. Khakpaki A. Advancements in artificial intelligence transforming medical education: a comprehensive overview. Med Educ Online. Dec 2025;30(1):2542807. [doi: 10.1080/10872981.2025.2542807] [Medline: 40798935]

25. Tangadulrat P, Sono S, Tangtrakulwanich B. Using ChatGPT for clinical practice and medical education: cross-sectional survey of medical students' and physicians' perceptions. JMIR Med Educ. Dec 22, 2023;9:e50658. [doi: 10.2196/50658] [Medline: 38133908]
26. Tran C, Hryciw BN, Moore SW, Chaput A, Seely AJ. Perceptions and use of generative artificial intelligence in medical students: a multicenter survey. J Med Educ Curric Dev. 2025;12:23821205251391969. [doi: 10.1177/23821205251391969] [Medline: 41181167]
27. Turney J, Young TM, Chauhan DR, Beeharry R, Mahmud M. AI use for medical students: impact on clinical skill acquisition and retention. A systematic review. Adv Med Educ Pract. 2026;17:583763. [doi: 10.2147/AMEP.S583763] [Medline: 42006163]
28. Tobler S. Context matters: interpreting effect sizes in education meaningfully. MethodsX. 2024;13:103023. [doi: 10.1016/j.mex.2024.103023] [Medline: 39553737]
29. Fatima SS, Sheikh NA, Osama A. Authentic assessment in medical education: exploring AI integration and student-as-partners collaboration. Postgrad Med J. Nov 22, 2024;100(1190):959-967. [doi: 10.1093/postmj/qgae088] [Medline: 39041454]
30. Visci P, Lupelli F, Calò F, et al. Undergraduate medical students and artificial intelligence: a global systematic review and meta-analysis using a Knowledge-Attitude-Practice framework. Med Teach. Oct 2026;48(10):1791-1803. [doi: 10.1080/0142159X.2026.2663855] [Medline: 42219530]
31. Alansari A, Luqman H. Large language models hallucination: a comprehensive survey. Comput Sci Rev. 2026;61(C). [doi: 10.1016/j.cosrev.2026.100970]
32. Ke Y, Jin L, Ong JC, et al. AI-induced never-skilling in medical education. Nat Med. Jun 2026;32(6):1997-2006. [doi: 10.1038/s41591-026-04438-y] [Medline: 42174254]
33. Sriram A, Ramachandran K, Krishnamoorthy S. Artificial intelligence in medical education: transforming learning and practice. Cureus. Mar 2025;17(3):e80852. [doi: 10.7759/cureus.80852] [Medline: 40255837]
34. Magalhães Araujo S, Cruz-Correia R. Incorporating ChatGPT in medical informatics education: mixed methods study on student perceptions and experiential integration proposals. JMIR Med Educ. Mar 20, 2024;10:e51151. [doi: 10.2196/51151] [Medline: 38506920]
35. McCoy L, Ganesan N, Rajagopalan V, McKell D, Niño DF, Swaim MC. A training needs analysis for AI and generative AI in medical education: perspectives of faculty and students. J Med Educ Curric Dev. 2025;12:23821205251339226. [doi: 10.1177/23821205251339226] [Medline: 40376309]
36. Abdulnour RE, Gin B, Boscardin CK. Educational strategies for clinical supervision of artificial intelligence use. N Engl J Med. Aug 21, 2025;393(8):786-797. [doi: 10.1056/NEJMra2503232] [Medline: 40834302]
37. Boman M, Jhun P, Schaekermann M. Scaffolding for success: blending learning with and about Generative AI in medical education. Med Teach. Dec 2025;47(12):1911-1917. [doi: 10.1080/0142159X.2025.2571041] [Medline: 41379104]
38. Wang L, Wan Z, Ni C, et al. Applications and concerns of ChatGPT and other conversational large language models in health care: systematic review. J Med Internet Res. Nov 7, 2024;26:e22769. [doi: 10.2196/22769] [Medline: 39509695]
39. Ahsan Z. Integrating artificial intelligence into medical education: a narrative systematic review of current applications, challenges, and future directions. BMC Med Educ. Aug 23, 2025;25(1):1187. [doi: 10.1186/s12909-025-07744-0] [Medline: 40849650]
40. Zhao J, Yang K, Gao H, Jiang G. Generative artificial intelligence and the transformation of medical education: a scoping review. Front Educ. 2026;11. [doi: 10.3389/feduc.2026.1907589]
41. Desai SV, Khan S, Lomis K. AI-enabled precision-education systems - transforming lifelong learning in medicine. N Engl J Med. Feb 26, 2026;394(9):838-841. [doi: 10.1056/NEJMp2512935] [Medline: 41728891]
42. Krumpal I. Determinants of social desirability bias in sensitive surveys: a literature review. Qual Quant. 2013;47:2025-2047. [doi: 10.1007/s11135-011-9640-9]

Abbreviations

CBME: competency-based medical education
NBME: National Board of Medical Examiners
SQUIRE: Standards for Quality Improvement Reporting Excellence
USU: Uniformed Services University

Edited by Taiane de Azevedo Cardoso; peer-reviewed by Eleni Samara, Melissa Swee; submitted 16.Aug.2026; initial comments to author 22.Sep.2026; final revised version received 26.Oct.2026; accepted 28.Oct.2026; published 26.Nov.2026

Please cite as:
Cole R, Hemmer PA, Moores L, Mears D, Samuel A, La Croix C, Pearce E, Elster E
How Medical Students Use AI for Learning: Cross-Sectional, Survey-Based Quality Improvement Evaluation
JMIR Med Educ 2026;12:e109763
URL: https://mededu.jmir.org/2026/1/e109763
doi: 10.2196/109763

© Rebekah Cole, Paul A Hemmer, Lisa Moores, David Mears, Anita Samuel, Christina La Croix, Elizabeth Pearce, Eric Elster. Originally published in JMIR Medical Education (https://mededu.jmir.org), 26.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.