



Algorithm as Medical Teacher: AI and the Future of Medical Teaching

Pushparaj Shetty ^{*1}, Jeevan Divakaran ², Ila Chauhan ³

- 1. Associate Professor, Department of Anatomy, Trinity School of Medicine, St. Vincent*
- 2. Professor, Department of Pathology, Medical University of the Americas, Nevis*
- 3. Professor and Chair, Department of Clinical Skills, Medical University of the Americas, Nevis.*

***Correspondence to:** Pushparaj Shetty. Associate Professor, Department of Anatomy, Trinity School of Medicine, St. Vincent.

Copyright

© 2026 **Pushparaj Shetty**. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received: 15 April 2026

Published: 12 May 2026

DOI: <https://doi.org/10.5281/zenodo.20131976>

Abstract

Background: For generations, the medical educator has stood at the center of a complex professional role — transmitting knowledge, building clinical skills, shaping professional identity, and mentoring the next generation of physicians. Artificial intelligence (AI) is now advancing at a pace that challenges each of these functions in concrete, evidenced ways, and not merely at the margins.

Aim: This perspective article examines how AI will, within three to five years, take on the core functions of the medical educator — drawing on current trial data, deployed technologies, and emerging educational frameworks rather than speculation.

Key Arguments: We work through six domains of a medical educator functions: personalized knowledge instruction, clinical simulation and skills training, assessment and feedback, curriculum design, mentorship and pastoral support, and professional identity formation. For each, we describe how AI is already performing or closely approximating these roles, and where the human educator remains irreplaceable.

Conclusion: The pressing question is no longer whether AI can do much of what a medical educator does — the evidence suggests it can. The real question is how institutions govern this shift without sacrificing the humanistic values that medicine depends on.

Keywords: artificial intelligence; medical education; large language models; AI tutoring; clinical simulation; assessment; curriculum design; medical educator role.

Introduction

Ask any medical educator what their job involves and you will get an answer that extends well beyond lecturing. They design curricula, supervise clinical work, give feedback on student performance, run assessments, mentor trainees through identity crises and career decisions, and quietly model what it means to be a good doctor. For a long time, this constellation of tasks was assumed to be irreducibly human — grounded in lived clinical experience, relational wisdom, and moral judgment that no machine could plausibly replicate.

That assumption has not aged well. Over the past three years, a convergence of large language models (LLMs), multimodal AI systems, adaptive learning platforms, and AI-powered simulation has produced tools that can tutor a medical student through a neuroanatomy pathway at midnight, simulate a patient in acute distress for history-taking practice, grade an OSCE note with near-human reliability, and generate examination questions indistinguishable from those written by experienced faculty. None of this is speculative — these tools are deployed in real curricula, serving real students.

The scale of the shift is becoming clearer in the data. A bibliometric analysis spanning publications from 2000 to 2024 identified exponential growth in AI-enabled medical education research, with the fastest acceleration occurring after 2022 — when generative AI became publicly accessible (Wang et al., 2025). At the same time, a 2024 international survey of more than 4,500 students across 192 medical, dental, and veterinary faculties found that over 75% had received no formal AI education in their curriculum, even as two in three physicians now use AI tools in practice — a 78% increase from 2023 (Ahmed et al., 2025). There is a striking disconnect between how deeply AI has penetrated clinical work and how slowly it has been formally integrated into the training that prepares people for that work.

The corporate world offers a cautionary parallel. Kodak pioneered digital imaging but declined to act on it, fearing disruption to its core film business; Nokia dominated mobile telecoms but moved too slowly when smartphones changed the game (Lucas & Goh, 2009; Wang, 2016). In both cases, the failure was not a lack of capability or foresight — it was strategic hesitation in the face of transformative change. Medical education faces a version of the same dilemma. The technology is here. The question is whether institutions will shape how it is used or simply find themselves overtaken by it.

To appreciate the scale of what is already underway, it helps to look beyond medical education specifically and consider the broader labor market picture. Figure 1 — taken from Anthropic's 2026 report *Labour Market Impacts of AI: A New Measure and Early Evidence* — plots theoretical AI capability against observed AI usage across 22 occupational categories. The blue area represents the share of job tasks that large language

models could theoretically perform based on the Eloundou et al. task-exposure scoring system; the red area shows actual, work-related automated usage recorded across Anthropic's platforms. The gap between the two is striking. For knowledge-intensive occupations most relevant to medical education — management (91.3%), legal (89.0%), computer and mathematics (94.3%), and education and library (61.7%) — theoretical capability is consistently high yet observed deployment lags far behind. Even in the most AI-active category (computer and mathematics), actual task coverage reaches only around 35.8% of what is theoretically feasible. The report's authors describe this gap as a measure of displacement risk yet to materialize: "As capabilities advance, adoption spreads, and deployment deepens, the red area will grow to cover the blue" (Anthropic, 2026, para. 14). For medical educators, this figure is not merely descriptive — it is a forecast. Education and library occupations sit at 61.7% theoretical exposure, and the trajectory of closing that gap is precisely what this article addresses.

This article does not argue that AI should or will replace medical educators completely. What it does argue is that within three to five years, AI will perform the functional core of the educator's role with sufficient fidelity that institutions which have not reckoned with this will be poorly prepared — educationally and ethically. We examine each major educator function in turn, ground our analysis in current evidence, and consider what genuinely remains beyond algorithmic reach.

Theoretical capability and observed usage by occupational category

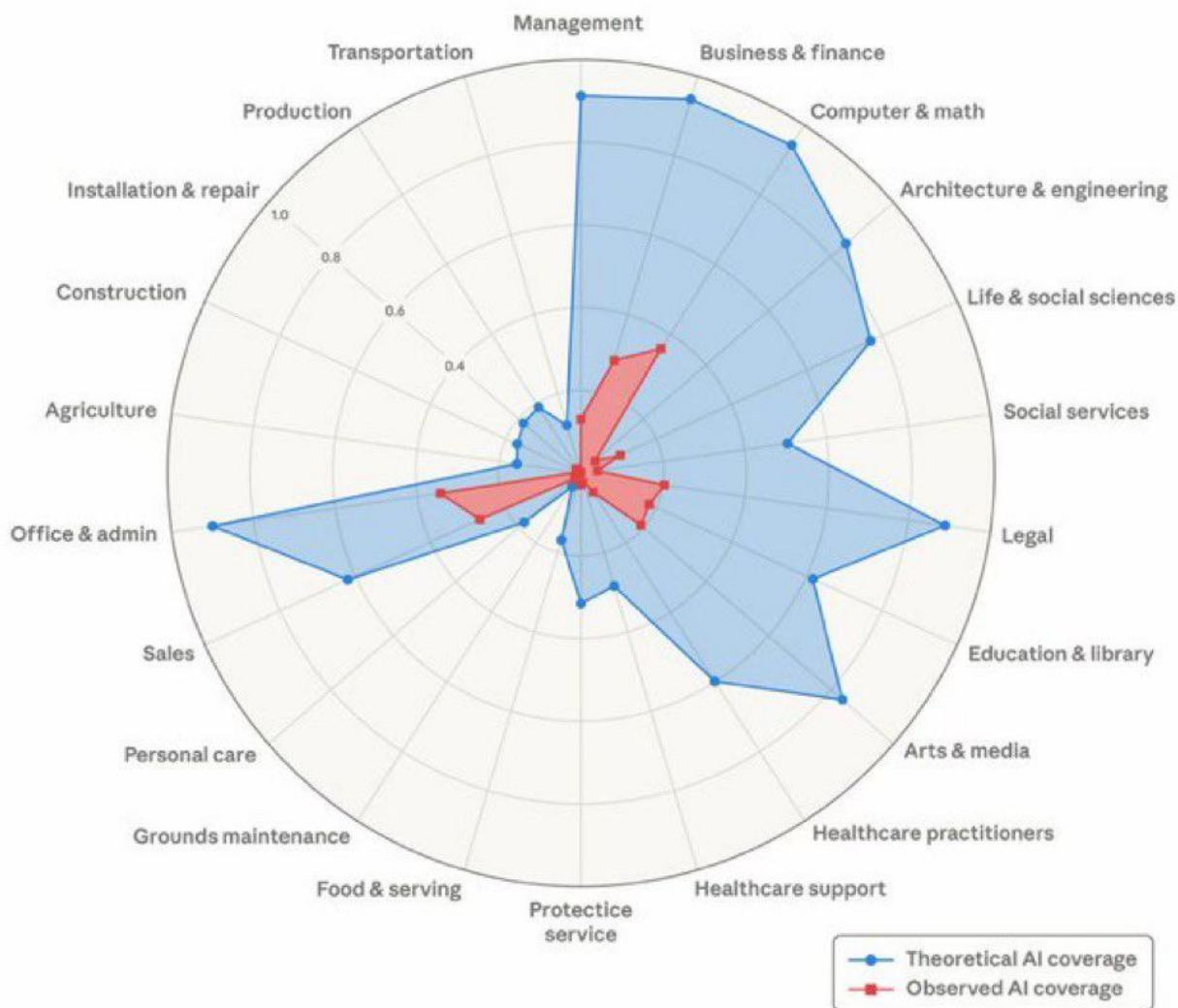


Figure 1. Theoretical AI capability and observed AI usage by occupational category. The blue area shows the share of job tasks theoretically feasible for large language models (based on Eloundou et al., 2023 β -scores); the red area shows actual automated, work-related usage recorded on Anthropic's Claude platform across the Anthropic Economic Index. Note the persistent gap between capability and deployment across all occupational groups. Reprinted from *Labour Market Impacts of AI: A New Measure and Early Evidence* (Figure 2), by Anthropic, 2026 (<https://www.anthropic.com/research/labor-market-impacts>). Open access.

What Medical Educators Actually Do: A Functional Taxonomy

Before we can ask how much of the educator's role AI can take on, we need to be clear about what that role actually consists of. Drawing on frameworks from the Academy of Medical Educators and the Association of American Medical Colleges, we identify six core functional domains:

- I. Personalized knowledge instruction and tutoring
- II. Clinical skills instruction and simulation
- III. Assessment: formative, summative, and programmatic
- IV. Curriculum design and learning environment construction
- V. Mentorship, coaching, and pastoral support
- VI. Professional identity formation and role modeling

These are not watertight categories — in practice, they blur into each other constantly. A bedside teaching session might touch all six simultaneously. But pulling them apart analytically lets us ask, for each one, how much AI can already do, how much it will likely do within five years, and what remains genuinely human.

Domain I — Personalized Knowledge Instruction and tutoring

Start with the function most obviously at risk: knowledge instruction. The traditional medical lecture — delivered to a cohort of 100 to 150 students at a fixed time, in a fixed order, at a pace set by the lecturer rather than the learner — has been under pressure for years from problem-based and team-based learning. AI is now mounting a more fundamental challenge. It offers something no lecturer can: a pedagogically sophisticated, infinitely patient tutor that is available around the clock and that calibrates its explanations to each individual student's current knowledge state.

There is now good evidence for what this looks like in practice. Thomas Thesen and Soo Hwan Park at Dartmouth's Geisel School of Medicine deployed a retrieval-augmented generation (RAG) AI teaching assistant — NeuroBot TA — in a second-year neurology course across two consecutive cohorts totaling 190 students. Rather than drawing on the open internet, the system restricted its responses to instructor-curated course materials, which meant it was accurate, contextually relevant, and curriculum-aligned. What the researchers found was striking: conversation volume with the AI increased by 329.6% in the three days before examinations, suggesting students were using it as an intensive just-in-time tutor precisely when they needed it most — evenings, weekends, the hours when no faculty member is available (Thesen & Park, 2025).

This kind of personalization is something human educators simply cannot deliver at scale. AI-driven learning pathways, which adapt in real time based on a student's performance data, have been associated with mastery rates up to 30% higher than those achieved through conventional instruction (Khakpaki, 2025). GPT-4o can explain the same concept — take the pathophysiology of acute kidney injury — at the right level of complexity for a first-year student, a final-year clerk, and a foundation year doctor, switching between them in seconds (Wu et al., 2024).

Looking five years out, the emerging Socratic tutoring paradigm seems particularly significant. Rather than just providing answers, next-generation AI tutors are being designed to guide students through diagnostic reasoning by asking them targeted questions — pushing them to justify their thinking rather than simply accepting the AI's output. At Dartmouth, this is already partially in place. Given the pace of LLM development and the ongoing pressure of rising student-to-faculty ratios, it is entirely plausible that within five years, AI tutoring will be the primary mode of individualized knowledge instruction across many of the world's medical schools.

The licensing examination benchmark is worth noting too. Current LLMs achieve around 76.8% accuracy on USMLE-style questions across anatomical topics, compared with 44.4% for the earlier GPT-3.5 — well above passing thresholds for students (Mavrych et al., 2025). GPT-4 has passed all three USMLE steps without specialist training, and its explanatory outputs contained novel clinical insights in approximately 90% of cases (Kung et al., 2023). An AI that can pass medical licensing examinations and articulate its reasoning is, functionally, teaching.

Domain II — Clinical Skills Instruction and Simulation

Clinical skills teaching has always required two things: expert demonstration and a safe space to practice. Historically, that meant patient contact — with all its ethical and logistical complexity — expensive standardized patient programs, or physical simulation laboratories. AI is now offering alternatives across all three.

A. Virtual Patients

Standardized patient (SP) programs are logistically demanding. Training actors is costly, scheduling encounters is complicated, and access is deeply unequal — particularly for students in lower-resource settings. LLM-powered virtual patients are not a perfect substitute, but they are a credible one. A 2025 qualitative study at the American University of Antigua assessed ChatGPT as a virtual standardized patient

in a series of live role-play encounters. Students reported that the AI-simulated patient interactions felt clinically realistic, challenged their history-taking and communication skills, and — notably — could convincingly simulate emotionally complex presentations, including patients in distress, confusion, or outright denial (Cross et al., 2025).

The experimental evidence is now hardening too. A quasi-experimental study at the Karolinska Institutet in spring 2025, involving 115 sixth-semester medical students, tested whether AI-generated post-consultation feedback integrated into robotic virtual patient interactions actually improved clinical performance. Students who received AI feedback after their virtual patient encounters scored significantly higher on subsequent OSCEs, with particularly clear improvements in history-taking and clinical communication — which are precisely the skills that normally take a senior clinician time to teach and assess (JMIR Medical Education, 2026).

The MedSimAI framework takes this further. By embedding self-regulated learning principles within an LLM-driven simulation environment, it enables students to engage in iterative deliberate practice with adaptive AI-generated feedback calibrated to their individual learning characteristics. Evidence from this system suggests that AI-based adaptive feedback outperforms static expert solutions in improving diagnostic justification quality, and that personalizing simulation parameters to individual learners represents a genuinely scalable model of clinical training that traditional programs simply cannot match on cost or reach (MedSimAI Research Team, 2025).

B. Procedural Skills and Assessment

Even procedural skills — which many assumed would be the last frontier of AI competence — are now being touched. A 2025 Turkish cross-sectional study assessed ChatGPT-4o and Gemini Flash 1.5 as evaluators of four clinical procedures: intramuscular injection, square knot tying, basic life support, and urinary catheterization. Working from video recordings and standardized checklists, the AI evaluators showed substantial consistency with expert human raters, raising a real prospect of AI-assisted OSCE assessment at scale (Tekin et al., 2025).

Looking forward, the American Medical Association's Precision Education initiative — which awarded grants to eleven institutions in 2025 specifically for AI-driven precision assessment — signals that this direction has institutional as well as technological momentum (American Medical Association, 2025).

For procedural skills, modern manikins can provide realistic tactile (haptic) feedback during simulations. At the same time, AI offers adaptive guidance and real-time feedback, playing an important role not just in assessment but also in procedural and surgical training (Almagharbeh et al., 2026).

Domain III — Assessment: Formative, Summative, and Programmatic

Assessment probably consumes more faculty time than any other educator function. Writing good examination questions, grading written clinical responses, evaluating performance in simulated encounters, giving meaningful individualized feedback — these tasks are relentless, time-intensive, and prone to the kind of inconsistency that comes from human fatigue and variable training. AI is demonstrably capable of taking on significant parts of all of them.

A. Question Generation

The automated generation of multiple-choice questions has attracted more study than almost any other AI application in medical education. A 2025 randomized controlled trial in *BMC Medical Education* compared AI-generated with human-generated MCQs in a high-stakes examination. Blinded faculty reviewers found the AI-generated items to be of comparable quality to the human-authored ones, and student performance across the two sets was statistically indistinguishable (Law et al., 2025). A separate comparative study found that iteratively refined ChatGPT outperformed clinical mentors in generating high-quality interprofessional clinical scenarios for teaching purposes (Tang et al., 2025). These are not modest results. Question-writing currently occupies faculty committees for weeks; AI can produce aligned, validated items in seconds. Even when not used directly, AI is now widely involved in creating learning materials—such as for problem-based learning, OSCEs, and integrated clinical cases—where many new cases are needed (Almagharbeh et al., 2026)

B. Grading Clinical Documentation

Grading complex written responses is harder — it requires understanding context, clinical reasoning, and the subtleties of what a student has actually understood, not just what they have written. Yet here too the evidence is moving. A 2025 case study from UT Southwestern Medical Center's Simulation Center describes the first successful prospective deployment of a generative AI grading system for student post-encounter OSCE notes, involving 245 pre-clerkship students across a ten-station examination. Trained on expert-authored exemplar statements and validated rubrics, the system provided first-pass grading that meaningfully reduced the

workload on trained human evaluators while maintaining assessment integrity (NEJM AI, 2025). At the University of Cincinnati, an AMA-funded project is going further: developing ambient AI that provides real-time feedback on clinical reasoning and communication as students interact with patients, entirely without faculty presence (American Medical Association, 2025).

C. Personalized Formative Feedback

Formative feedback — timely, specific, individualized, actionable — is probably the most powerful single driver of learning in medical education. It is also the thing that faculty consistently under-deliver, for the simple reason that there is never enough time. An AI-assisted feedback tool called 'Sisu Athwala', built using retrieval-augmented generation with custom LLMs at the University of Peradeniya, gave personalized study guidance to undergraduate medical students based on individual performance profiles. An evaluation involving fifteen student counsellors found that between 70% and 90% agreed the system provided clear, actionable plans, identified individual weaknesses with specificity, and offered genuinely novel improvement suggestions — the kind of response that usually requires dedicated one-to-one faculty time (Jayawardena et al., 2025).

The Aquifer clinical education platform has deployed a similar AI feedback system that assesses students' clinical summary statements against expert-derived rubrics and returns specific guidance within seconds. These are not pilot projects sitting in university innovation offices — they are live, in curricula, delivering at scale what faculty cannot (Aquifer, 2025).

Domain IV — Curriculum Design

Curriculum design sits at the cognitively demanding end of the educator's work. It requires integrating evidence-based pedagogy, evolving clinical standards, accreditation requirements, learner needs, and institutional constraints — then weaving all of that into a coherent, sequenced learning experience. It has traditionally been the function most resistant to technological substitution. The assumption was that you needed the kind of judgment that only comes with years of clinical and educational experience.

The evidence is now complicating this assumption. A systematic review examining 24 articles published to early 2024 documented widespread adoption of AI tools — particularly ChatGPT — for curriculum design across educational disciplines (Chokkakula et al., 2025). In medical education specifically, AI is being used to map curricular content against competency frameworks, identify gaps in learning objectives, generate case-based learning scenarios calibrated to clinical prevalence, and model programmatic assessment architectures.

What is emerging beyond this is the concept of 'living curriculum design' — AI systems that learn continuously from student performance data and propose curriculum changes in real time. If cohort-level performance on renal physiology is declining, such a system can identify this, trace it to specific teaching episodes, and propose pedagogical adjustments before the next cohort begins. No curriculum committee operates at this speed or with this data density. Wang et al.'s (2025) bibliometric analysis identifies AI-enabled curriculum reform as among the fastest-growing research domains in the field, with projections of continued acceleration through 2028.

The Lancet Digital Health, in a 2025 Viewpoint, noted another dimension worth taking seriously: AI's contribution to research training. LLM-assisted literature review, data synthesis, and manuscript drafting are reshaping how medical students and trainees are prepared to contribute to the evidence base of their field. This is curriculum design at the level of what it means to be a clinical scientist, not just a clinician (The Lancet Digital Health, 2025).

Domain V — Mentorship, Coaching, and Pastoral Support

Mentorship is where the terrain becomes genuinely complicated. Students consistently identify meaningful mentoring relationships as among the most formative experiences of their training. And yet mentorship is the function most explicitly dependent on authentic human relationship — on being known, being witnessed, and being accompanied by someone who has something real at stake in your development.

AI can perform the functional components of mentorship with a degree of competence that is uncomfortable to acknowledge. A 2025 paper in *Frontiers in Medicine* documented ChatGPT functioning in a mentorship-adjacent role — providing personalized academic guidance that students found contextually sensitive and practically useful (Chokkakula et al., 2025). The system could synthesize a student's academic history, learning preferences, and performance trajectory, maintaining context across interactions in a way that many students — particularly those in large cohorts who rarely have meaningful one-on-one faculty contact — said they had never experienced from a human mentor.

There is a real equity argument here. Students from underrepresented backgrounds, those in geographically isolated institutions, or those dealing with difficulties they feel unable to raise with a supervisor often face significant barriers to mentorship. Evidence suggests that medical students frequently prefer to disclose academic and personal difficulties to an AI first, before approaching a human — they perceive the AI as non-judgmental and reliably available (Ahmed et al., 2025). For students who would otherwise have no

mentorship at all, imperfect AI mentorship may genuinely be better than nothing. That is a real, uncomfortable institutional question.

But something important distinguishes AI coaching from true mentorship. The literature on empathy is consistent on this: AI can simulate the cognitive dimensions of empathy — understanding, predicting, and responding to emotional states — but it cannot generate genuine empathic experience. This is not a technical gap waiting to be closed; it is a philosophical distinction rooted in the fact that empathy requires subjective experience and authentic concern for another person's wellbeing, neither of which AI possesses (Halpern & Brown, 2021). Real mentorship, understood as the human relationship through which professional identity is shaped rather than just a service that provides guidance, remains beyond algorithmic reach.

The question institutions must sit with is this: when the realistic alternative is not AI mentorship versus rich human mentorship, but AI mentorship versus no mentorship at all, what is the ethical choice?

Domain VI — Professional Identity Formation and Role Modeling

This is the domain where AI runs hardest into its own limits. Professional identity formation — the gradual development of the values, commitments, and ways of being that characterize an excellent physician — does not happen through lectures or assessments. It happens through witnessing. Students need to see how a senior clinician handles a mistake, navigates an ethical dilemma, or stays present with a patient who is frightened and confused. They need to encounter people who embody the ideals of the profession under real conditions of pressure, uncertainty, and loss.

AI cannot do this. It cannot model professional values in the embodied sense because it has no professional identity, no genuine stake in patient outcomes, and no lived experience of what it means to carry clinical responsibility. As Mehta and colleagues (2025) argue in *Perspectives on Medical Education*, the educator's irreplaceable function in this era is precisely to model empathy-based care, to help students distinguish algorithmic outputs from clinical judgment, and to guard against the cognitive deskilling that over-reliance on AI could cause. A humanistic approach to medicine resists the reduction of patients to data points — and that lesson cannot be taught by something that processes data for a living.

AI can, however, play a supporting role here too. By generating realistic ethical dilemmas, cases involving moral complexity, or interprofessional conflict scenarios, it can provide the raw material that human educators then use to facilitate structured reflection. The distinction matters: AI generates content for professional identity formation; the human educator remains the catalyst who gives that content meaning through relationship, experience, and example.

Students do not just need information; they need someone to become like. They need direction, identity, and meaning — someone who not only teaches medicine but represents what it means to practice it well. That remains, for now and foreseeably, a human function.

The Hybrid Horizon: What This Means for Institutions and Educators

Taken together, the evidence in this article describes a near-term future in which AI performs the functional core of the medical educator's role across at least five of the six domains examined — knowledge instruction, clinical simulation, assessment, curriculum design, and coaching — with enough fidelity to be a genuine alternative rather than merely a supplement to human faculty in many settings. This is not a prediction about the distant future. Deployed systems operating in live curricula right now already demonstrate these capabilities. The three-to-five-year horizon is about their consolidation, scaling, and normalization.

In November 2025, a panel convened by the Penn Leonard Davis Institute — drawing together faculty from Stanford, Northwestern, and NYU — put the challenge plainly: AI is reshaping how students learn and how faculty teach so rapidly that medical schools can no longer treat it as optional. Panelists called for parallel testing of AI against conventional teaching methods, more granular study of AI's differential effects across stages of training, and a fundamental redesign of competency assessment frameworks (Penn Leonard Davis Institute, 2025).

For individual educators, the transition demands something beyond adapting to a new tool. If AI takes over knowledge transmission, assessment, and simulation, the human educator's distinctiveness must increasingly reside in the things AI cannot do: moral witnessing, professional identity formation, the governance of AI's role in education itself, and the cultivation of clinical wisdom that requires lived experience to transmit. This is what Mehta and colleagues (2025) call the 'straddle generation' challenge — educators who trained in a pre-AI curriculum now having to master, deploy, and govern AI's presence in the institution simultaneously. Medical educators should think less about defending territory and more about redefining value. This means shifting from transmitting facts to teaching clinical reasoning, from lecturing to facilitating dialogue around uncertainty, from content expert to curator and interpreter. The AAMC's principles for responsible AI use in medical education, published in 2025, provide a governance starting point — with commitments to transparency, equity, data protection, and the preservation of human-centered care as non-negotiable constraints (Association of American Medical Colleges, 2025).

The opportunity here is significant. Unlike Kodak or Nokia, medical educators are not facing simple obsolescence — they are facing a redefinition of their role that, handled well, could be genuinely liberating.

The functions that consume most of an educator's time and carry the least pedagogical leverage — delivering content, writing questions, marking assessments — are precisely the ones AI can take. The functions that carry the most leverage — mentoring, modelling, facilitating insight — are the ones that remain human. The question is whether institutions will create the conditions that allow this reorientation to happen.

Risks, Limitations, and Ethical Concerns

A responsible account of AI's potential in medical education cannot omit its very real risks and limits.

First, AI hallucination — the confident generation of factually incorrect clinical information — remains a material danger in high-stakes learning contexts. Even the most capable current LLMs produce errors that require expert human oversight to catch. Deploying AI tutors without robust verification mechanisms risks propagating misinformation to students who, by definition, do not yet have the expertise to identify it (Harvard Medicine Magazine, 2024).

Second, algorithmic bias is a structural equity problem, not just a technical one. AI systems trained on datasets that underrepresent patient populations or clinical contexts will systematically disadvantage the students and patients those populations represent. Research at Massachusetts General Hospital is actively training AI systems to identify and correct bias in faculty evaluations of students — a sobering acknowledgment that bias operates on both sides of the human-AI interface (Harvard Medicine Magazine, 2024).

Third, cognitive deskilling is a real and documented risk of AI dependence. Students trained heavily on AI outputs may struggle with the kind of ambiguous, uncertain, real-world complexity that medicine presents. Medicine relies on intuition and nuanced judgment built through effortful clinical reasoning — and AI threatens to shortcut precisely the cognitive work through which that judgment develops. As Richard Schwartzstein at Harvard Medical School observes, interpreting tests and working inductively are how students build critical thinking; AI risks making that work feel unnecessary (Harvard Medicine Magazine, 2024). Students trained heavily by AI may struggle with ambiguity, uncertainty, and real-world complexity. Medicine relies on intuition and nuanced decision making and AI systems, while powerful, lack the intuition and contextual judgement that human educators develop over years of clinical practice.

Fourth, there is the question of accountability. When AI provides incorrect teaching or clinical guidance, who is responsible? The developer who built the system? The institution that deployed it? The educator who integrated it into a course? The student who relied on it? Unlike a human teacher, AI systems are often opaque — it can be genuinely difficult to trace how a particular conclusion was reached or where an error originated. In a field where accountability is tied to professional standards and legal liability, this ambiguity has real

consequences for patient safety. AI cannot model human emotions in a truly authentic way, potentially creating a generation of clinicians who are technically competent but less patient centered.

Fifth, AI cannot simulate authentic empathy and should not try. The philosophical literature is clear that empathic AI is not merely technically unachieved — it may be genuinely impossible, because empathy requires subjective experience and conscious emotional investment that AI, by its nature, does not have (Halpern et al., 2021). A generation of students taught predominantly by AI systems risks developing technical competence without the patient-centered orientation that medicine requires.

Sixth, data governance is non-trivial. Using students' personal learning data to drive adaptive algorithms requires transparent governance, informed consent, and serious protection against surveillance, profiling, and commercial exploitation (Pohn et al., 2025). If AI provides incorrect teaching or clinical guidance, it raises questions about responsibility. Who is accountable—the developer, the institution, or the student? This ambiguity complicates both education and patient safety. Is the fault with the developer, who designed and trained the system? The institution, which approved and implemented its use? The educator, who integrated it into teaching? Or the student, who relied on it? Unlike a human teacher, AI systems are opaque, making it difficult to trace how a particular conclusion was reached or where an error originated. The implications for patient safety are even more serious. Medicine operates within a framework where accountability is tied to professional standards and legal liability. If a clinical decision influenced by AI leads to harm, determining liability becomes complex. Without clear guidelines, there is a risk of accountability gaps, where responsibility is diffused across multiple actors, potentially leaving patients inadequately protected and errors insufficiently addressed.

Seventh, students don't just need information; they need someone to become like, but AI cannot mentor. It cannot inspire. It cannot shape professional identity. Through ongoing dialogue and real-world exposure, mentors guide learners through uncertainty, failure, and growth. AI, by contrast, lacks continuity of relationship, emotional investment, and genuine accountability. Inspiration in medicine is sparked by encountering individuals who embody the ideals of the profession—clinicians who demonstrate compassion under pressure, integrity in difficult decisions, and resilience in the face of loss. These lived examples shape not only what students know, but who they aspire to become. Students do not just need accurate answers; they need direction, identity, and meaning. They need someone who not only teaches medicine but represents what it means to practice it well.

Finally, there is the global equity problem. Resource-rich institutions with the infrastructure, capacity, and funding to deploy high-quality AI educational systems will gain significant advantages over lower-resource

institutions — potentially widening the very disparities that AI's advocates promise it will close (The Lancet Digital Health, 2025).

Conclusion

AI is now capable of performing the functional core of the human educator role across multiple domains, and the trajectory of capability improvement makes clear that within three to five years, this will not be a marginal phenomenon but a structural one. The evidence in this article — from AI tutors at Dartmouth and Karolinska, from AI assessors at UT Southwestern and in Turkish universities, from AI feedback platforms operating at scale in live curricula globally — shows that this transformation is already well underway.

The question before medical educators, institutions, and accreditation bodies is not whether to engage with this transformation — that choice has already been made by the technology itself. The question is how to govern it: which educator functions should AI be permitted to take on, under what conditions, with what accountability mechanisms, and with what non-negotiable commitments to the human values that medicine fundamentally requires.

Medical educators should focus less on lecturing facts and more on teaching clinical reasoning and helping students navigate uncertainty, moving from “what to think” to “how to think.” We should learn how AI tools work, integrate them into teaching, and teach students how to critically evaluate AI outputs. In an era of misinformation and AI hallucinations, educators must act as curators (what matters, what doesn't) and validators (what's accurate, what's misleading). They should become mentors and role models. Survival will depend on a willingness to evolve—from transmitters of knowledge to facilitators of clinical reasoning, from lecturers to mentors, and from content experts to curators and interpreters of increasingly complex information ecosystems.

Let us be the architects of the future of healthcare; the future is AI-augmented, not AI-replaced. In contrast to Kodak and Nokia, the opportunity for medical educators is not merely to avoid decline, but to redefine their value in ways that technology cannot replicate, ensuring their continued relevance in an AI-augmented future.

References

-
- ❖ Ahmed, S., Saha, A., Majumdar, S., & Chunder, R. (2025). Artificial intelligence in medical education: Promise, pitfalls, and practical pathways. *Advances in Medical Education and Practice*, 16, 689–700. <https://doi.org/10.2147/AMEP.S500000>
 - ❖ American Medical Association. (2025). Precision education: AMA ChangeMedEd initiative grant recipients. American Medical Association. <https://www.ama-assn.org/education/changemeded-initiative/precision-education>
 - ❖ Almagharbeh WT, Ahmed WR, Alfanash HA, Alnawafleh KA, Tajoury OH. Touching the future: How haptic technology and AI are revolutionizing nursing simulations: A systematic scoping review. *J Prof Nurs*. 2026 Mar-Apr;63:18-25. doi: 10.1016/j.profnurs.2025.12.010. Epub 2026 Jan 6. PMID: 41796557
 - ❖ Anthropic. (2026, March 5). Labor market impacts of AI: A new measure and early evidence. [Figure 2: Theoretical capability and observed exposure by occupational category]. Anthropic Economic Research. <https://www.anthropic.com/research/labor-market-impacts>
 - ❖ Aquifer. (2025, October). AI-powered feedback: Enhancing clinical reasoning in medical education. Aquifer Inc. <https://aquifer.org/blog/ai-powered-feedback-enhancing-clinical-reasoning-in-medical-education>
 - ❖ Association of American Medical Colleges. (2025). Principles for the responsible use of artificial intelligence in and for medical education. AAMC. <https://www.aamc.org/about-us/mission-areas/medical-education/principles-ai-use>
 - ❖ Brodeur, P. G., Buckley, T. A., Kanjee, Z., Goh, E., Restrepo, D., Cabral, S., & Schwartzstein, R. M. (2024). Superhuman performance of a large language model on the reasoning tasks of a physician. *arXiv*. <https://doi.org/10.48550/arXiv.2412.10849>
 - ❖ Cabral, S., Restrepo, D., Kanjee, Z., Schwartzstein, R. M., Goh, E., Brodeur, P. G., & Rodman, A. (2024). Clinical reasoning of a generative artificial intelligence model compared with physicians. *JAMA Internal Medicine*, 184(5), 581–583. <https://doi.org/10.1001/jamainternmed.2024.0295>
 - ❖ Chokkakula, S., Chong, J., Yang, M., Jiang, H., Yu, L., Han, S., Attitalla, I., Yin, T., & Zhang, J. (2025). Quantum leap in medical mentorship: Exploring ChatGPT's transition from textbooks to terabytes. *Frontiers in Medicine*, 12, Article 1517981. <https://doi.org/10.3389/fmed.2025.1517981>
 - ❖ Cook, D. A., Overgaard, J., Pankratz, V. S., Del Fiol, G., & Aakre, C. A. (2025). Virtual patients using large language models: Scalable, contextualized simulation of clinician-patient dialogue with feedback. *Journal of Medical Internet Research*, 27, e68486. <https://doi.org/10.2196/68486>
-

-
- ❖ Cross, J., Kayalackakom, T., Robinson, R. E., Vaughans, A., Sebastian, R., Hood, R., Lewis, C., Devaraju, S., Honnavar, P., Naik, S., & Anand, N. (2025). Assessing ChatGPT's capability as a new-age standardized patient: Qualitative study. *JMIR Medical Education*, 11, e63353. <https://doi.org/10.2196/63353>
 - ❖ Geathers, J., Hicke, Y., Chan, C., Rajashekar, N., Sewell, J., Cornes, S., Kizilcec, R. F., & Shung, D. L. (2025). Benchmarking generative AI for scoring medical student interviews in objective structured clinical examinations (OSCEs). *arXiv*. <https://doi.org/10.48550/arXiv.2501.13957>
 - ❖ Halpern, J., & Brown, J. E. H. (2021). AI chatbots cannot replace human interactions in the pursuit of more inclusive mental healthcare. *SSM – Mental Health*, 1, Article 100017. <https://doi.org/10.1016/j.ssmmh.2021.100017>
 - ❖ Halpern, J., Johansson Sturm, T., & Montemayor, C. (2021). In principle obstacles for empathic AI: Why we can't replace human empathy in healthcare. *BMC Medical Ethics*, 22, Article 34. <https://doi.org/10.1186/s12910-021-00602-0>
 - ❖ Harvard Medicine Magazine. (2024, October). How generative AI is transforming medical education. Harvard Medical School. <https://magazine.hms.harvard.edu/articles/how-generative-ai-transforming-medical-education>
 - ❖ Holderried, F., Stegemann-Philipps, C., Herrmann-Werner, A., Festl-Wietek, T., Holderried, M., Eickhoff, C., & Zierott, L. (2024). A language model-powered simulated patient with automated feedback for history taking: Prospective study. *JMIR Medical Education*, 10, e59213. <https://doi.org/10.2196/59213>
 - ❖ Jayawardena, R., Perera, S., & Dissanayake, V. (2025). Artificial intelligence based personalised student feedback system 'Sisu Athwala' to enhance exam performance of medical undergraduates. *PLOS ONE*, 20(6), e0327410. <https://pmc.ncbi.nlm.nih.gov/articles/PMC12677440/>
 - ❖ JMIR Medical Education. (2026). AI-generated feedback following social robotic virtual patient interactions and medical student performance: Nonrandomized quasi-experimental study. *JMIR Medical Education*, 12, e90368. <https://doi.org/10.2196/90368>
 - ❖ Khakpaki, A. (2025). Advancements in artificial intelligence transforming medical education: A comprehensive overview. *Medical Education Online*, 30(1), Article 2542807. <https://doi.org/10.1080/10872981.2025.2542807>
 - ❖ Kung, T. H., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepaño, C., Madriaga, M., Aggabao, R., Diaz-Candido, G., Maningo, J., & Tseng, V. (2023). Performance of ChatGPT on
-

-
- USMLE: Potential for AI-assisted medical education using large language models. *PLOS Digital Health*, 2(2), Article e0000198. <https://doi.org/10.1371/journal.pdig.0000198>
- ❖ Law, A. K., So, J., Lui, C. T., Choi, Y. F., Cheung, K. H., & Hung, K. K. C. (2025). AI versus human-generated multiple-choice questions for medical education: A cohort study in a high-stakes examination. *BMC Medical Education*, 25(1), Article 208. <https://doi.org/10.1186/s12909-025-05208-x>
 - ❖ Lucas, H. C., & Goh, J. M. (2009). Disruptive technology: How Kodak missed the digital photography revolution. *Journal of Strategic Information Systems*, 18(1), 46–55. <https://doi.org/10.1016/j.jsis.2009.01.002>
 - ❖ Mavrych, V., Ganguly, P., & Bolgova, O. (2025). Using large language models (ChatGPT, Copilot, PaLM, Bard, and Gemini) in gross anatomy course: Comparative analysis. *Clinical Anatomy*, 38(2), 200–210. <https://doi.org/10.1002/ca.24244>
 - ❖ MedSimAI Research Team. (2025). MedSimAI: Simulation and formative feedback generation to enhance deliberate practice in medical education. *arXiv*. <https://arxiv.org/abs/2503.05793>
 - ❖ Mehta, N., Mehta, S., Rubenstein, A., & Wood, S. K. (2025). Not replaced, but reinvented: AI education pathways to prepare future physicians to lead healthcare transformation. *Perspectives on Medical Education*, 14(1), 849–859. <https://doi.org/10.5334/pme.2233>
 - ❖ NEJM AI. (2025). Rubrics to prompts: Assessing medical student post-encounter notes with AI [Case study]. *NEJM AI*. <https://doi.org/10.1056/AIcs2400631>
 - ❖ Penn Leonard Davis Institute. (2025, November). AI pushes medical schools into new era, but are they prepared? University of Pennsylvania. <https://ldi.upenn.edu/our-work/research-updates/ai-pushes-medical-schools-into-new-era-but-are-they-prepared/>
 - ❖ Pohn, B., Mehnen, L., Fitzek, S., Choi, K.-E. A., Braun, R. J., & Hatamikia, S. (2025). Integrating artificial intelligence into pre-clinical medical education: Challenges, opportunities, and recommendations. *Frontiers in Education*, 10, Article 1570389. <https://doi.org/10.3389/feduc.2025.1570389>
 - ❖ Sharifi, S. V., & Boushehrinejad, A. G. (2025). Enhancing medical education through artificial intelligence: Opportunities, challenges, and future directions. *European Journal of Clinical Medicine and Pharmacological Research*, 9(3), 112–128. https://www.ejcmpr.com/article_225793.html
 - ❖ Tang, Q., Feng, R., Zheng, B., Zhou, J., Li, G., & Zhou, Y. (2025). Iteratively refined ChatGPT outperforms clinical mentors in generating high-quality interprofessional education clinical scenarios:
-

-
- A comparative study. *BMC Medical Education*, 25(1), Article 845. <https://doi.org/10.1186/s12909-025-07845-y>
- ❖ Tekin, M., Yurdal, M. O., Toraman, C., Korkmaz, G., & Uysal, I. (2025). Is AI the future of evaluation in medical education? AI vs. human evaluation in objective structured clinical examination. *BMC Medical Education*, 25(1), Article 641. <https://doi.org/10.1186/s12909-025-07241-4>
 - ❖ The Lancet Digital Health. (2025). How can artificial intelligence transform the training of medical students and physicians? *The Lancet Digital Health*, 7(10), e745–e752. [https://doi.org/10.1016/S2589-7500\(25\)00082-2](https://doi.org/10.1016/S2589-7500(25)00082-2)
 - ❖ Thesen, T., & Park, S. H. (2025). A generative AI teaching assistant for personalized learning in medical education. *npj Digital Medicine*, 8(1), Article 627. <https://doi.org/10.1038/s41746-025-02022-1>
 - ❖ UNESCO. (2023). ChatGPT and artificial intelligence in higher education: Quick start guide. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000385146>
 - ❖ Wang, J. (2016). A study of Nokia's decline in the mobile phone market. *International Journal of Business Administration*, 7(5), 60–70. <https://doi.org/10.5430/ijba.v7n5p60>
 - ❖ Wang, Y., Chang, C., Shi, W., Liu, H., Huang, X., & Jiao, Y. (2025). How AI is transforming medical education: Bibliometric analysis. *JMIR Medical Education*, 11, Article e75911. <https://doi.org/10.2196/75911>
 - ❖ Wu, Y., Zheng, Y., Feng, B., Yang, Y., Kang, K., & Zhao, A. (2024). Embracing ChatGPT for medical education: Exploring its impact on doctors and medical students. *JMIR Medical Education*, 10, Article e52483. <https://doi.org/10.2196/52483>
 - ❖ Yauy, K., Lavigne, E., Lopez, A., Frandon, J., Blaizot, G., Gabellier, L., & Adham, S. (2025). AI-standardized clinical examination training on OSCE performance. *NEJM AI*, 2(8), Article AIoa2500066. <https://doi.org/10.1056/AIoa2500066>