Communication

Not peer-reviewed version

The Architectural Shift: Integrating Artificial Intelligence from the Ground Up in Undergraduate Medical Education

Sandeep Reddy *

Posted Date: 19 March 2026

doi: 10.20944/preprints202603.1560.v1

Keywords: artificial intelligence; medical education; curriculum integration; clinical supervision; AI literacy; deskilling

Communication

The Architectural Shift: Integrating Artificial Intelligence from the Ground Up in Undergraduate Medical Education

Sandeep Reddy

School of Public Health and Social Work, Queensland University of Technology, Brisbane, Australia; sandeep.reddy@qut.edu.au

Abstract

As healthcare enters an era defined by algorithmic decision-making, the traditional medical curriculum faces a fundamental challenge. The exponential growth of medical knowledge and the ubiquity of digital health tools render pedagogical models centred on rote memorisation increasingly inadequate. This review argues that Artificial Intelligence (AI) cannot be treated as a supplemental elective or a peripheral module appended to existing curricula. Instead, AI must be integrated as a longitudinal thread woven through every phase of undergraduate medical education (UGME), from foundational sciences to clinical rotations. Drawing on recently published frameworks (2020–2026), this paper proposes a scaffolded pedagogical structure for producing "AI-adaptive" physicians across four curricular phases: pre-clinical foundations, case-based learning, supervised clinical rotations, and reformed assessment. The review examines key frameworks, including the DEFT-AI model for clinical supervision, Bloom's Taxonomy-aligned competency mapping, and trust calibration exercises. Challenges related to faculty development, equity, and the risk of "deskilling" are discussed alongside implementation strategies. The paper concludes that failure to embed AI structurally within medical curricula risks producing a generation of graduates who are either fearful of these technologies or dangerously dependent upon them.

Keywords: artificial intelligence; medical education; curriculum integration; clinical supervision; AI literacy; deskilling

---

1. Introduction: The Imperative for Structural Change

The medical profession is navigating a technological transformation characterised by the convergence of biological and digital domains [1]. The doubling time of medical knowledge, once measured in decades, is now estimated at approximately 73 days [2]. In this environment, the traditional pedagogical objective of knowledge acquisition alone is insufficient; the emerging priority is knowledge management, a function for which AI serves as a primary engine [2]. Despite a broad consensus that AI will reshape diagnostics, clinical workflow, and patient safety, medical education has largely responded with inertia. Where AI is taught, it is frequently confined to elective courses or isolated workshops, positioning the technology as a novelty rather than a core professional competency [3]. This piecemeal approach creates a dangerous disconnect between training and practice. Learners are already deploying Generative AI (GenAI) tools in clinical settings, often without formal supervision or systematic verification frameworks [4].

To prepare graduates for the realities of contemporary and future practice, medical schools must move beyond ad hoc additions. AI integration must be introduced early and embedded structurally into every phase of training [5,6]. The challenge is not simply technical; it is fundamentally pedagogical. Medical educators must reconcile the tension between preserving the humanistic foundations of clinical reasoning and equipping students with the digital competencies demanded by modern practice [11,13]. This review outlines a scaffolded, longitudinal curriculum designed to

 doi:10.20944/preprints202603.1560.v1

mitigate the risks of "deskilling" while cultivating physicians who can seamlessly integrate human clinical judgment with algorithmic intelligence: a concept described as the "centaur" model of practice [4]. The structure draws on recently published curricular frameworks, pedagogical strategies, and risk analyses from 2020 to 2026, and proposes four interconnected phases: pre-clinical foundations, case-based learning integration, supervised clinical application, and reformed assessment.

2. Phase 1: The foundational Years (Pre-Clinical)

2.1. Demystifying the Black Box Through Basic Sciences

The ground-up approach requires that AI literacy commence from the earliest days of medical school, running concurrently with anatomy and physiology. Just as pharmacology cannot be understood without biochemistry, AI tools cannot be safely deployed without a grasp of the fundamentals of data science and algorithmic reasoning [2,3]. In the pre-clinical phase, the curricular focus should align with the lower levels of Bloom's Taxonomy, Remembering and Understanding, while establishing foundational AI literacy [5]. Critically, the goal is not to train every medical student in programming, but to produce "data-literate consumers" who can critically appraise algorithmic outputs [3].

Existing biostatistics and epidemiology modules provide a natural integration point. When students learn about sensitivity, specificity, and predictive values, they should simultaneously be introduced to concepts of algorithmic bias, overfitting, and the distinction between correlation and causation in large datasets [6]. Similarly, as students learn organ system physiology, they should encounter AI tools currently operating in those domains. For example, while studying cardiovascular physiology, students can be introduced to machine learning algorithms that detect atrial fibrillation from ECG data, with a structured discussion of both utility and limitations [6,7]. This approach eliminates the need for separate AI teaching hours, addressing curricular overcrowding, and ensures students learn to assess algorithmic tools in clinical settings from the start of their training. Problem-based learning tutorials in the pre-clinical years also offer a valuable integration point: AI-generated case extensions can be introduced as stimuli for group discussion, enabling students to practise evaluating AI output collaboratively before they encounter such tools in clinical settings [5,6].

2.2. Pedagogical Tool: The RAG-Based Teaching Assistant

To acculturate students to responsible AI use, institutions should deploy sanctioned AI tools for self-directed learning. Thesen and Park [8] demonstrated the utility of a Retrieval-Augmented Generation (RAG) tool grounded in verified curricular materials. Unlike open access chatbots prone to hallucination, RAG systems constrain responses to validated course content. Students use these tools for just-in-time learning, particularly before examinations. This method streamlines querying AI systems and encourages students to verify source citations, reflecting clinical practice in the use of decision-support tools [8]. Additionally, integrating RAG-based platforms fosters early habits of critical appraisal and ethical engagement with digital resources. As students become familiar with these AI tools, they develop the confidence and discernment necessary for safe, effective application in clinical environments.

 doi:10.20944/preprints202603.1560.v1

Table 1. Evolution of AI competencies across the undergraduate medical curriculum (adapted from Cheng et al. [5] and Abdulnour et al. [4]).

Curricular phase	Bloom's level	AI competency focus	Example activity
Pre-clinical (Year 1–2)	Remember, Understand	Data literacy, algorithmic bias, ML fundamentals	RAG-based self-directed learning; bias identification exercises
Case-based learning (Year 2–3)	Apply, Analyse	Prompt engineering, output verification, trust calibration	AI-generated differential critique; virtual patient simulation
Clinical rotations (Year 3–4)	Analyse, Evaluate	Supervised AI use, DEFT-AI framework, human-in-the-loop practice	Structured preceptor debriefs; Centaur/Cyborg mode switching
Assessment & capstone	Evaluate, Create	Critical appraisal, AI-free competency, process evaluation	Show-your-work assessments; algorithmic bias case studies

3. Phase 2: Integration Into Case-Based Learning

3.1. Moving from Theory to Application

As the curriculum transitions toward pathophysiology and diagnostics, AI education must shift from understanding how algorithms work to how they should be used in clinical reasoning. This aligns with the Apply and Analyse levels of Bloom's Taxonomy [5]. In traditional case-based learning (CBL), students receive a static patient vignette. In an AI-integrated curriculum, students should be encouraged to use GenAI tools to generate differential diagnoses, which they must then systematically critique against primary evidence [9].

3.2. The Trust Calibration Exercise

Hough et al. [9] propose the design of "trust calibration" exercises in which students are presented with clinical cases and AI-generated management plans. Crucially, some AI outputs are intentionally flawed or contain embedded biases. Students are required to consult textbooks, clinical guidelines, and published literature as primary sources of evidence to determine whether to accept, modify, or disregard the AI's recommendations. This directly counters automation bias, defined as the tendency to uncritically accept algorithmic output, and prevents the outsourcing of clinical

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 March 2026 doi:10.20944/preprints202603.1560.v1

reasoning [9]. Such exercises develop a disposition of healthy scepticism that is transferable to the broader clinical environment, where information of variable quality is encountered routinely from consultants, textbooks, and electronic decision-support tools alike.

 doi:10.20944/preprints202603.1560.v1

3.3. *High-Fidelity Simulation with Multimodal AI*

Ning et al. [10] describe the potential of multimodal GenAI to create high-fidelity video simulations in which AI-driven virtual patients adjust vital signs and symptom presentations in real time in response to student interventions. In a typical scenario, a student manages a virtual patient whose condition dynamically deteriorates or improves depending on the clinical decisions made. After the simulation, the AI compares the student's performance to thousands of past results and provides personalised, objective feedback on decision speed and diagnostic accuracy; something human tutors can't match at scale [11]. These simulations also offer the advantage of exposing students to rare or high-acuity presentations that may be infrequently encountered during standard clinical placements, thereby broadening the range of clinical experience accessible during training [10,11].

4. Phase 3: Clinical Rotations and the Human-in-the-Loop

4.1. *The Danger of Deskilling*

The transition to clinical rotations represents the highest risk period for what Abdulnour et al. [4] term “never-skilling”: if students rely on AI to generate clinical notes or differential diagnoses without adequate oversight, they may fail to develop the cognitive pathways required for independent clinical reasoning. Simultaneously, prohibiting the use of AI is neither realistic nor desirable, as students are already using these tools [4,9]. The solution lies in structured, supervised AI usage embedded within clinical training. By integrating supervised AI use into clinical rotations, educators can ensure that students develop essential clinical judgment while learning to critically assess algorithmic recommendations. Structured supervision allows preceptors to guide students in verifying AI-generated outputs against established clinical guidelines and literature. This approach not only mitigates the risk of deskilling but also fosters a balanced relationship between technological assistance and human expertise. Ultimately, the goal is to cultivate clinicians capable of leveraging AI responsibly without compromising their independent problem-solving abilities.

4.2. *The DEFT-AI Framework for Clinical Supervision*

Abdulnour et al. [4] propose the DEFT-AI framework for clinical supervision, which transforms a student’s use of AI into a structured teachable moment. The framework ensures that AI remains supplementary to human intelligence rather than a replacement. Table 2 outlines the five components of the framework.

Table 2. The DEFT-AI framework for clinical supervision of AI use during clinical rotations (adapted from Abdulnour et al. [4]).

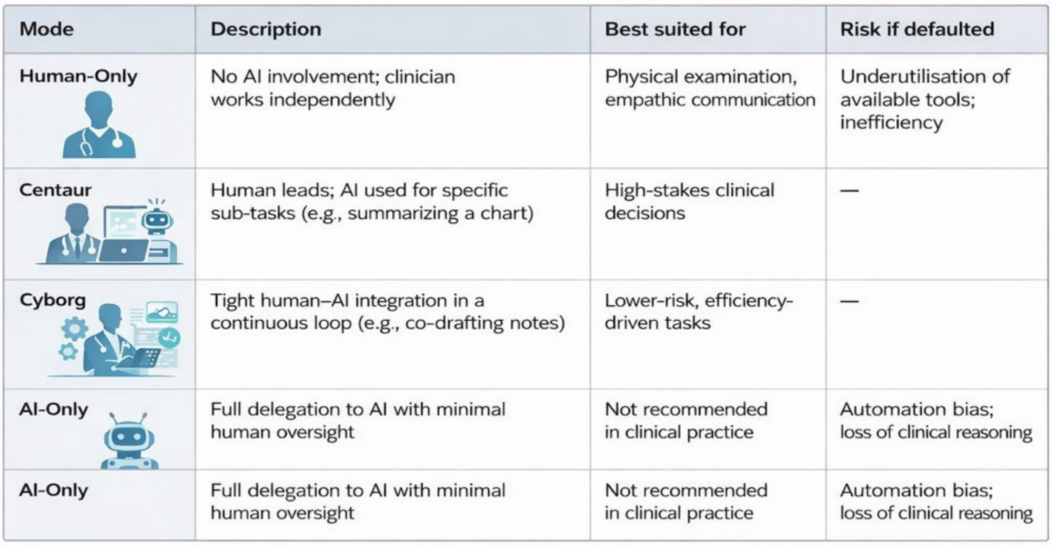
Component	Preceptor prompt	Educational purpose
D – Diagnosis / Discussion	“What AI tool did you use, and what prompts did you enter?”	Forces articulation of the interaction strategy; makes the AI use explicit and auditable
E – Evidence	“How did you verify the AI-generated output?”	Ensures the student cross-referenced AI suggestions with trusted clinical literature or guidelines

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 March 2026 doi:10.20944/preprints202603.1560.v1

F – Feedback	Guided reflection on AI performance	Develops critical appraisal by identifying where AI succeeded or failed (e.g., missed a subtle clinical sign)
T – Teaching	Instruction on prompt engineering techniques	Builds technical skill in eliciting better AI outputs (e.g., chain-of-thought prompting strategies)
AI – Recommendation	Advise on appropriate reliance level	Teaches contextual judgement about when and how much to rely on AI for different clinical tasks

4.3. The Spectrum of Human–AI Collaboration

Abdulnour et al. [4] describe a spectrum of human–AI collaboration ranging from purely human decision-making to fully AI-driven practice. The educationally desirable zone lies in the middle: the Centaur mode, in which the clinician leads and delegates specific sub-tasks to AI while retaining full decision-making authority, is suited to high-stakes decisions; the Cyborg mode, in which human and AI work in tight iterative loops, is suited to lower-risk, efficiency-oriented tasks such as documentation. The goal of the clinical curriculum is to teach students when and how to transition between these modes. Graduates lacking proficiency in this area may resort to unquestioned dependence on artificial intelligence or opt not to utilise available technological resources, both of which are regarded as insufficient professional practices within today's evolving technological context [4].



Mode	Description	Best suited for	Risk if defaulted
Human-Only	No AI involvement; clinician works independently	Physical examination, empathic communication	Underutilisation of available tools; inefficiency
Centaur	Human leads; AI used for specific sub-tasks (e.g., summarizing a chart)	High-stakes clinical decisions	—
Cyborg	Tight human–AI integration in a continuous loop (e.g., co-drafting notes)	Lower-risk, efficiency-driven tasks	—
AI-Only	Full delegation to AI with minimal human oversight	Not recommended in clinical practice	Automation bias; loss of clinical reasoning
AI-Only	Full delegation to AI with minimal human oversight	Not recommended in clinical practice	Automation bias; loss of clinical reasoning

Figure 1. Modes of human–AI collaboration in clinical practice (adapted from Abdulnour et al. [4]).

5. Phase 4: Assessment and Evaluation

5.1. Grading the Process, Not Just the Product

If AI is integrated structurally into the curriculum, assessment methods must evolve in parallel. Traditional multiple-choice examinations are increasingly inadequate, as AI systems can already pass medical licensing exams with high proficiency [12]. Three assessment paradigms merit consideration. First, “show your work” assessments should assume AI access and grade students on prompt quality,

 doi:10.20944/preprints202603.1560.v1

reasoning transparency, and the written rationale for accepting or rejecting AI output [9]. This approach evaluates the cognitive process rather than merely the final answer and reflects the reality that in clinical practice, the quality of one's reasoning is at least as important as the conclusion reached. Second, designated "AI-free zones" should be preserved for high-stakes competencies—physical examination technique, bedside communication, and acute resuscitation scenarios—to ensure baseline human capabilities are maintained and that graduates can function in settings where technology may be unavailable [9]. Third, critical appraisal exercises should test a student's ability to identify algorithmic bias; for example, recognising that a dermatology AI performs poorly on darker skin tones due to imbalances in the training data [9,10]. This tripartite assessment model ensures that students are evaluated not only on their clinical knowledge but also on their capacity to interact safely, critically, and ethically with AI systems.

5.2. The Consequences of Curricular Inaction

The risks of treating AI as a peripheral addition rather than a structural element are considerable. Webster [13] argues that while AI functions as a powerful rule engine, it is fundamentally context-insensitive. A physician who treats AI as an opaque oracle rather than a tool requiring human interpretation will miss the contextual factors—patient values, social determinants, cultural considerations—that ultimately dictate clinical care. Without structured training, students risk falling into what Webster terms "solutionism": the assumption that technology can resolve complex clinical and social problems without an understanding of the underlying mechanisms or biases [13]. Furthermore, inadequately trained physicians may inadvertently propagate algorithmic bias, widening existing healthcare disparities for vulnerable and underrepresented populations [10]. An additional concern is what might be termed the "atrophy of uncertainty tolerance." Clinical practice inherently involves ambiguity, incomplete information, and probabilistic reasoning. If students become habituated to AI systems that present confident outputs regardless of evidential strength, they may lose the capacity to sit with diagnostic uncertainty, a skill that experienced clinicians regard as foundational to safe practice [9,13]. The consequences extend beyond individual competence to system-level effects: if a cohort of graduates enters the workforce unable to function without AI assistance, health system resilience in the event of technical failures, cyberattacks, or resource-limited settings will be materially compromised [2,14].

5.3. Implementation Challenges and Strategies

Embedding AI longitudinally is resource-intensive. Medical schools face a well-documented shortage of faculty who are themselves AI-literate [2]. Three strategies are particularly relevant. First, cross-disciplinary collaboration between medical schools and faculties of engineering, computer science, and data science can accelerate capacity building. "Datathons," which are gatherings where clinicians and data scientists work together to solve health issues, serve as a successful example of extracurricular activities [3]. Second, faculty development must proceed in parallel with curricular reform. Abdulnour et al. [4] argue that current clinical preceptors should engage in co-exploration of AI tools alongside their students, adopting a "teaching while learning" posture rather than waiting for a new generation of AI-native educators. Third, global collaboration is essential. To prevent a digital divide in medical education, particularly affecting low- and middle-income countries, institutions should share open-source curricula, simulation platforms, and training datasets [10]. The UNESCO AI Competency Framework for Teachers, though not designed specifically for medical education, offers a transferable progression model across three developmental levels, acquire, deepen, and create, that could inform analogous frameworks for medical faculty development [14]. A further challenge lies in the ethical governance of AI tools within educational settings. Issues of student data privacy, ownership of AI-generated clinical content, and the potential to reinforce existing biases in training datasets all require institutional-level policy responses that extend beyond the individual classroom [9,10]. These governance structures should be established proactively rather

 doi:10.20944/preprints202603.1560.v1

than reactively, ensuring that ethical considerations are embedded in the design of AI-integrated curricula from the outset.

Table 4. Key implementation challenges and proposed strategies for longitudinal AI curriculum integration.

Challenge	Risk if unaddressed	Proposed strategy
Faculty AI illiteracy	Inability to supervise or model responsible AI use [2]	Co-learning with students; cross-disciplinary partnerships; structured faculty development [3,4]
Curricular overcrowding	AI relegated to elective or single-session workshops	Integration into existing modules (e.g., biostatistics, CBL) rather than additive content [5,6]
Student deskilling / never-skilling	Graduates unable to reason independently without AI [4]	DEFT-AI supervision framework; AI-free assessment zones; trust calibration exercises [4,9]
Equity and resource disparities	Digital divide between well-resourced and under-resourced institutions [10]	Open-source curricula; shared simulation platforms; global academic collaboration [3,10]
Ethical and bias concerns	Propagation of algorithmic bias into clinical decisions [9,10]	Bias identification training embedded in CBL; critical appraisal assessments [6,9]

6. Conclusions

The integration of AI into undergraduate medical education is not a technological update; it is a pedagogical reconstruction. The prevailing model of adding elective AI modules to traditional curricula is insufficient and poses substantial risks. It may produce a generation of physicians who are either fearful of available technologies or dangerously dependent upon them, while lacking the critical faculties to distinguish between these postures. To equip graduates for modern clinical practice, AI should be integrated across all learning environments, from labs to wards. Using structured frameworks like DEFT-AI, aligned with Bloom's Taxonomy and updated assessments, medical schools can develop clinicians who blend empathy and judgment with algorithmic skills. The frameworks reviewed in this paper provide actionable starting points, but they must be adapted to local institutional contexts, resource environments, and regulatory settings. Achieving this will require sustained institutional commitment, cross-disciplinary collaboration, a willingness to reform entrenched assessment practices, and an acceptance that the curriculum itself must be as adaptive as the graduates it aims to produce. Future research should prioritise longitudinal evaluation of AI-integrated curricula to determine whether these pedagogical strategies translate into measurable improvements in clinical competence and patient outcomes.

Author Contributions: The author conceptualised the work, reviewed and synthesised the literature, and wrote this paper.

Research funding: Not applicable.

Informed consent: Not applicable.

Research ethics: Not applicable.

Data availability: Not applicable.

Conflicts of Interest: The author states no conflict of interest.

References

1. Ejaz H, McGrath H, Wong BLH, Guise A, Vercauteren T, Shapey J. Artificial intelligence and medical education: a global mixed-methods study of medical students' perspectives. Digit Health 2022;8:1–11.
2. Grunhut J, Wyatt ATM, Marques O. Educating future physicians in artificial intelligence (AI): an integrative review and proposed changes. J Med Educ Curric Dev 2021;8:1–8.
3. McCoy LG, Nagaraj S, Morgado F, Harish V, Das S, Celi LA. What do medical students actually need to know about artificial intelligence? npj Digit Med 2020;3:86.
4. Abdulnour RE, Gin B, Boscardin CK. Educational strategies for clinical supervision of artificial intelligence use. N Engl J Med 2025;393:786–97.
5. Cheng D, Chan E, Song Y, Cui A, Chang A, Wang W. Framework for AI integration into the undergraduate medical curricula: a scoping review. BMC Med Educ 2026. Advance online publication. https://doi.org/10.1186/s12909-026-08620-1.
6. Grunhut J, Marques O, Wyatt ATM. Needs, challenges, and applications of artificial intelligence in medical education curriculum. JMIR Med Educ 2022;8:e35587.
7. Hedderich DM, Keicher M, Wiestler B, Gruber MJ, Burwinkel H, Hinterwimmer F, et al. AI for doctors—a course to educate medical professionals in artificial intelligence for medical imaging. Healthcare 2021;9:1278.
8. Thesen T, Park SH. A generative AI teaching assistant for personalized learning in medical education. npj Digit Med 2025;8:1–10.
9. Hough J, Culley N, Erganian C, Alahdab F. Potential risks of GenAI on medical education. BMJ Evid Based Med 2025;30:406–8.
10. Ning Y, Ong JCL, Cheng H, Wang H, Ting DSW, Tham YC, et al. How can artificial intelligence transform the training of medical students and physicians? Lancet Digit Health 2025. Advance online publication. https://doi.org/10.1016/j.landig.2025.100900.
11. Davids J, Lam K, Nimer A, Gianarrou S, Ashrafian H. AIM in medical education. In: Lidströmer N, Ashrafian H, editors. Artificial intelligence in medicine. Cham: Springer Nature Switzerland AG; 2021.
12. Pucchio A, Eisenhauer EA, Moraes FY. Medical students need artificial intelligence and machine learning training. Nat Biotechnol 2021;39:388–9.
13. Webster CS. Artificial intelligence and the adoption of new technology in medical education. Med Educ 2020;54:1074–6.
14. Looi MK. Can AI teach medicine? BMJ 2025;389:r822.