JPSIM Journal of Pakistan Society of Internal Medicine

Review Article

# From Passive Tools to Autonomous Educators: The Transformative Role of Agentic Artificial Intelligence in Competency-Based Healthcare Education — A Review

**Mudasar Latif Memon, Bikha Ram Devrajani**

*The University of Modern Sciences, Tando Muhammad Khan, Sindh, Pakistan.*

**Abstract**

The growth of intelligent systems in healthcare has increased the demand for future-ready clinical competencies in medical graduates. Current AI applications in medical education—rule-based intelligent tutoring systems, fixed e-learning platforms, and passive decision-support systems—lack the ability to provide personalized, adaptive learning in line with competency-based medical education (CBME). A substantive paradigm shift is agentic AI, characterized by goal-directedness, multi-step reasoning, and dynamic tool use. A structured narrative review was conducted. PubMed, Google Scholar, Scopus, and IEEE Xplore were searched systematically using terms such as agentic AI, autonomous AI in education, large language models in medical education, and LMIC medical education AI, and included publications published from January 2018 to March 2026. After applying pre-specified inclusion and exclusion criteria, 35 sources were included. We present a novel five-dimensional taxonomy of agentic AI in CBME that includes the level of autonomy, instructional role, specificity to clinical domain, assessment modality, and deployment context. Its applications in personalized curriculum design, clinical simulation, formative assessment, and faculty augmentation are examined. Prevalent issues such as algorithmic bias, the risk of hallucination, lack of data privacy, inadequate infrastructure, and regulatory uncertainty are discussed, with a particular focus on low- and middle-income countries (LMICs), specifically Pakistan. To successfully implement agentic AI in healthcare education, it is necessary to have evidence-based frameworks, institutional readiness strategies, and coordinated national policy. We suggest an organized research agenda to fill in important evidentiary gaps.

**Keywords:** Agentic artificial intelligence, competency-based medical education, AI tutors, healthcare education.

**How to cite this:**

Memon ML, Devrajani BR. From Passive Tools to Autonomous Educators: The Transformative Role of Agentic Artificial Intelligence in Competency-Based Healthcare Education — A Review. J Pak Soc Intern Med. 2026;7(3): 182-190

**Corresponding Author:** Dr. Bikha Ram Devrajani **Email:** devrajani62@gmail.com

**Received:** 25-03-2026 **Revised:** 29-06-2026 **Accepted:** 12-08-2026 **DOI:** https://doi.org/10.70302/jpsim.v7i3.2636



## Introduction

The digital health and intelligent systems have revolutionized the healthcare delivery system and concurrently increased the competency level of practitioners in the world. WHO estimates that there will be a shortage of 10 million healthcare workers by 2030, with the highest burden in LMICs such as Pakistan.[1] Competency-based medical education (CBME) - the new paradigm, supported by the ACGME and Pakistan Medical Commission (PMC) - requires individualised, outcomes-based, and measurable learning experiences.[2] The intelligent tutoring systems, adaptive learning systems, and computer-assisted assessment systems that have been developed in the last 20 years have been used as passive tools: responsive to student input, with no cross-session memory, and unable to set goals independently.[3] With the introduction of LLMs, including GPT-4 and Med-PaLM 2, and the development of reinforcement learning and multi-agent systems, a new type of educational technology has emerged: the agentic AI system [4] These systems break down complex instructional objectives into subtasks, monitor learner progress over time, dynamically adjust pedagogical strategies, and maintain multi-turn dialogue that approximates faculty mentorship.

The transition, however, has substantive risks. Health professions education instructional errors have direct patient-safety implications. In Pakistan, medical colleges in the public sector are faced with limited bandwidth, AI-literate professors, and poorly developed regulatory frameworks.[5] It is noted that the global discussion of AI in medical education has been largely produced in HIC contexts, with gaps in the operational deployment instructions, taxonomic distinction of AI agency levels, and South Asian contextual data.

In this paper, we present a structured narrative review of agentic AI in CBME with the following contributions:

- We systematically map agentic AI architectures applicable to healthcare education and distinguish between them and passive AI systems.
- We present a novel five-dimensional agentic AI system taxonomy in CBME that includes the level of autonomy, instructional role, clinical domain, assessment modality, and deployment context.
- We analyze applications in formative assessment, clinical simulation, curriculum personalization, and faculty augmentation with citation-verified evidence.
- We evaluate implementation issues unique to Pakistan and the broader LMIC setting and suggest an organized, operationalized research agenda for responsible integration.

Section 2 gives background on CBME and AI development in medical education. Section 3 presents the suggested taxonomy. Section 4 analyses applications. Section 5 discusses challenges and ethical considerations. Section 6 outlines a systematic research agenda, and Section 7 summarizes the review.

***Background and Related Work***

*Competency-Based Medical Education: Foundations and Framework*: The international move towards CBME was motivated by strong evidence that time in training is not a valid proxy of clinical competence. Frank et al.6 described CBME as an outcomes-based model based on observable competencies, which form the foundation of the CanMEDS and ACGME models and their evaluative core entrustable professional activities EPAs and milestones. In Pakistan, CBME elements were added to the MBBS curriculum in 2020 by the PMC, but are not uniformly implemented in 114 recognised medical colleges. Balochistan and Sindh institutions often do not have the basic CBME infrastructure, such as complex assessment tools, trained faculty assessors, and longitudinal learner portfolios, which are not always available in the public-sector institutions.[7]

*Evolution of Artificial Intelligence in Medical Education*: The development of AI in medical education has gone through multiple generations, starting with rule-based expert systems like GUIDON and INTERNIST-I in the 1980s[8] to the current multimodal architectures. Wang et al.[9] surveyed prompt-engineering systems to evaluate clinical knowledge, and Balasundaram and Santosh[10] surveyed deep learning in healthcare education, reporting significant performance improvements over previous rule-based systems. Chen et al.[11] discussed applications of transformer-based LLMs in medical education, including potential uses in OSCE feedback and associated limitations. Kung et al.[4] recorded a landmark inflection point, as their GPT-4 achieved a passing score on the USMLE, the first sign of LLMs as substantive educational agents.

*From Passive Tools to Agentic AI: The Paradigm Transition*: According to Schmidgall et al.3 agentic AI is a set of systems that can pursue goals through multiple reasoning processes, employ tools, and learn without constant human oversight. In contrast to passive systems that react to discrete inputs, agentic systems plan instructional sequences, evaluate learner responses, and update strategies based on them. They interface with medical databases, EHRs, simulation platforms, and institutional LMS, forming adaptive learning environments. Williams et al. 12 discovered that RLHF and tool-augmented reasoning generated significantly superior knowledge retention compared to non-interactive AI tutors. Thirunavukarasu et al. 13 discovered that LLMs reached 72 to 90 percent accuracy on benchmark medical knowledge exams, creating a plausible baseline to apply in education. Therefore, these advances necessitate a systematic taxonomy and analysis of applications in Sections 3 and 4.

*Taxonomy of Agentic AI in Competency-Based Healthcare Education*: A coherent taxonomy is essential to differentiate, compare, and procure agentic AI architectures in CBME. We present a novel five-dimensional taxonomy created by iteratively synthesizing the reviewed literature; formal validation by a modified Delphi process with international medical education experts is a required next step (see Section 6). Table 1 summarises the taxonomy, and its dimensions are discussed in Sections 3.1-3.4.

**Table 1:** *Proposed Five-Dimensional Taxonomy of Agentic AI Systems in CBME*

| Dimension | Levels / Categories | Representative Systems | LMIC Relevance |
|---|---|---|---|
| Autonomy Level | L0: Passive; L1: Single-step adaptive; L2: Multi-step HITL; L3: Goal-directed; L4: Fully autonomous | ChatGPT (L1–2); ReAct agents (L3); Experimental autonomous tutors (L4) | L1–L2 recommended for immediate LMIC deployment |
| Instructional Role | Tutor; Assessor; Simulator; Mentor | Khanmigo (Tutor); AI-OSCE scorer (Assessor); virtual patient (Simulator) | Tutor and Assessor roles most feasible in LMICs |
| Clinical Domain | Domain-general; Domain-specific; Multi-domain federated | Med-PaLM 2 (general); Rad-DINO (radiology); Federated architectures | Domain-general LLMs more deployable in resource-constrained settings |
| Assessment Modality | Formative; Summative; Diagnostic | AI mini-CEX feedback; EPA milestone tracker; Gap analysis agent | Formative AI assessment offers greatest scalability benefit |
| Deployment Context | Cloud-based; On-premise; Hybrid edge | Most commercial LLMs (cloud); Ollama-based local models (on-premise) | On-premises/ edge deployment critical for bandwidth-limited institutions |

*HITL = Human-in-the-loop; LLM = Large Language Model; OSCE = Objective Structured Clinical Examination; EPA = Entrustable Professional Activity; LMIC = Low- and Middle-Income Country, RLHF = Reinforcement Learning from Human Feedback.*

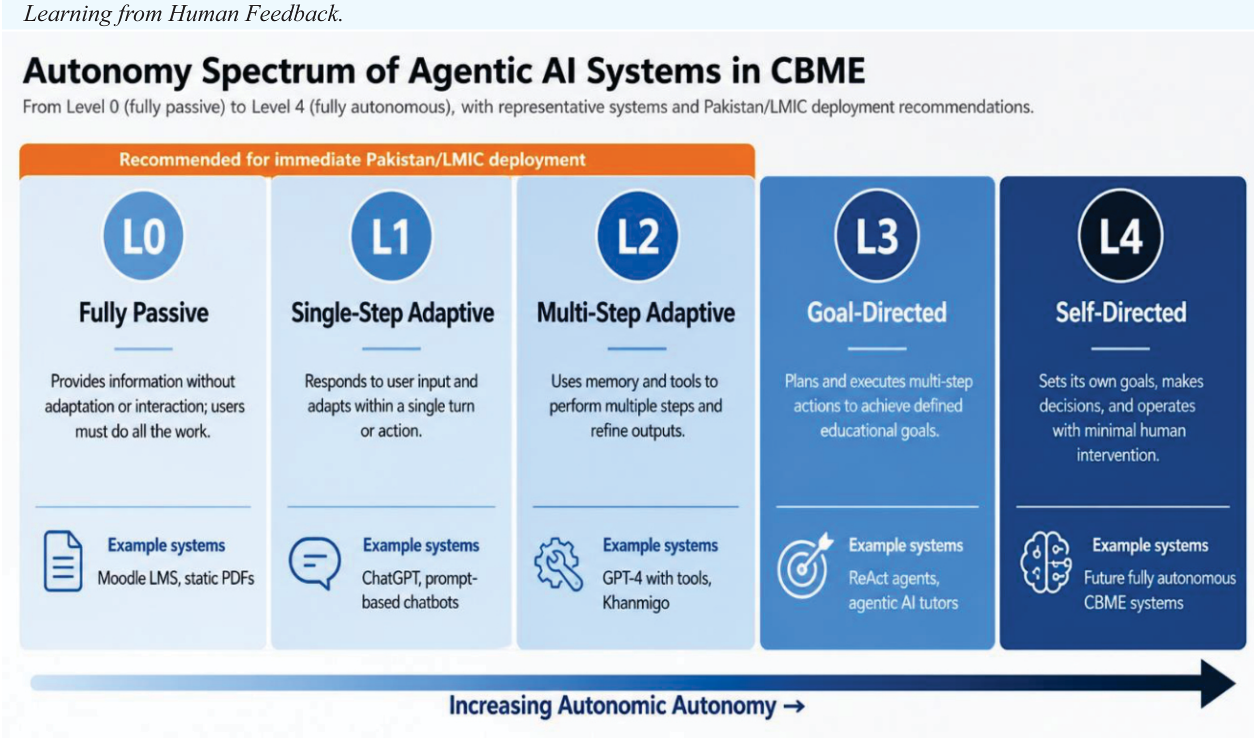



**Figure 1:** *Autonomy Spectrum of Agentic AI Systems in Competency-Based Healthcare Education.*

Figure shows five levels from L0 (fully passive) to L4 (fully autonomous), with representative system examples. The dashed orange box denotes the zone recommended for immediate Pakistan/LMIC deployment.

*Autonomy Level*: Inspired by SAE International vehicle automation levels, we establish five levels of autonomy in healthcare education AI, as shown in Figure 1: Level 0 (completely passive content delivery); Level 1 (adaptive single-step response); Level 2 (planning with human-in-the-loop control); Level 3 (goal-oriented autonomy with periodic supervision); and Level 4 (autonomous, self-revising agents). The majority of existing LLM-based learning systems are L1-L2.[14] Autonomous L3-L4 systems that monitor EPA achievement and tailor-made curricula are still in the research prototype phase.[15] In the short term, Pakistani institutions need to focus on L1-L2 systems installed on-premises, and gradually develop digital infrastructure for more autonomous agents.

*Instructional Role*: There are four major instructional roles: Tutor (adaptive content delivery), Assessor (competency benchmarking), Simulator (virtual clinical scenarios), and Mentor (longitudinal coaching to develop professional identity). One platform can serve many functions; institutional deployment agreements should still indicate the main purpose and accountability chain to meet regulatory and ethical oversight needs.

*Clinical Domain Specificity*: The agentic systems can be domain-general (trained on general medical corpora) or domain-specific (optimised for radiology, ECG analysis, etc.). Domain-specific agents perform well in their niche but do not cross clinical disciplines. Federated architectures of specialist sub-agents are thus needed in multi-domain CBME programmes.16 Conversely, domain-general LLMs have much more deployment flexibility and customisation ease for resource-constrained institutions.

*Assessment Modality and Deployment Context:* The dimension of assessment modality reflects whether the agent is formatively (real-time feedback), summatively (end-point evaluation), or diagnostically (gap identification to remediate). Formative modality is particularly agentic deployment-friendly, allowing scalable monitoring at all times. The dimension of deployment context differentiates between cloud-based systems, on-premise systems, and hybrid systems. On-premise deployment through open-source frameworks like Ollama with Llama 3 or Mistral is the most feasible near-term option in LMIC settings. The five dimensions of taxonomy are directly used in the analysis of applications in Section 4, allowing the deployment feasibility to be compared in a structured way.

### *Applications of Agentic AI in Competency-Based Healthcare Education*

*Personalised Curriculum Design and AI-Driven Clinical Simulation*: One of the most significant contributions of agentic AI to CBME is its ability to build adaptive learning pathways. In contrast to rule-based adaptive systems, which have pre-determined difficulty parameters, agentic systems combine longitudinal assessment data, clinical exposure data, and learner-reported preferences to create dynamic curricular experiences. According to Bozkurt et al.[17] and Sallam,[18] the most promising educational uses of generative AI in the near future are personalised content pacing and responsive feedback. The agents that are based on reinforcement learning can simulate the zone of proximal development of each learner and modify the difficulty of the task based on micro-level cues such as response latency, frequency of answer revision, and quality of clinical reasoning articulation.

The agentic architectures have significantly improved clinical simulation, which is one of the pillars of CBME. Cascella et al.[19] tested the use of AI based on LLM in various clinical scenarios, finding scenario generation as one of the most promising areas. Mok et al.[20] directly compared AI-generated responses to clinical cases to those of medical students, and found that LLM responses were more complete on factual recall tasks, which has direct implications on the design of AI-augmented case-based learning. NLP-based simulation agents also simulate structured clinical interviews, and they provide feedback on the completeness of history-taking, the quality of communication, and diagnostic hypothesis generation.[21,22] Figure 2 shows the design of an agentic clinical simulation system with a CBME assessment framework.

Three colour-coded layers: Agentic Core (blue), Learner Interaction Layer (green), and Competency Assessment & EPA Tracking (gold), with bidirectional data flows and external EHR and LMS integration. LLM = Large Language Model; KB = Knowledge Base; EPA = Entrustable Professional Activity; EHR = Electronic Health Record; NLP = Natural Language Processing; LMS = Learning Management System.

*Automated Formative Assessment and Faculty Augmentation*: The most resource-intensive aspect of CBME is formative assessment, which uses faculty observers to conduct DOPS and mini-CEX assessments. The agentic AI provides a complementary, not a replacement, to faculty-led assessment. Chen et al. 11 discovered that agentic assessors using LLM had 87.2% agreement with faculty assessors on written case-based tasks and also saved 73% of the time on evaluation, but concordance was lower with nuanced professional behaviour assessment, which highlights the need to maintain human supervision. Kaur et al.[23] offer a governance framework of trustful AI in healthcare with principles that can be directly applied to agentic assessment systems, including explainability, fairness auditing, and human-in-the-loop accountability.

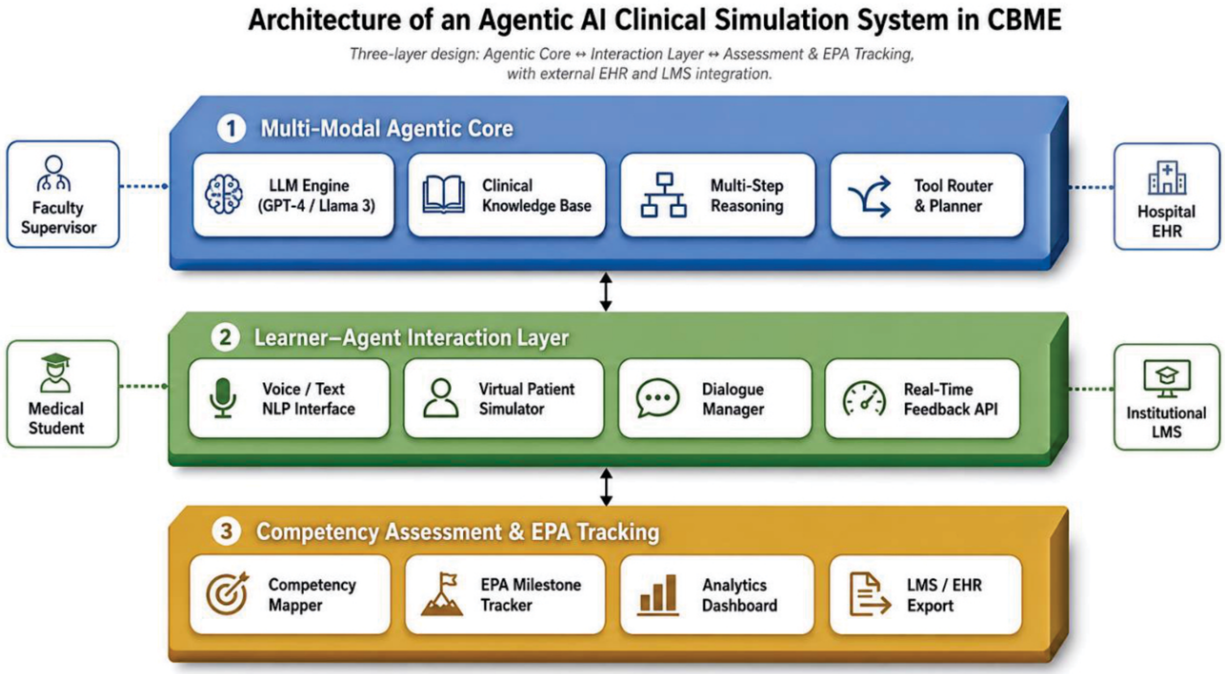



**Figure 2:** *Architecture of an Agentic AI Clinical Simulation System Integrated with a CBME Assessment Framework.*

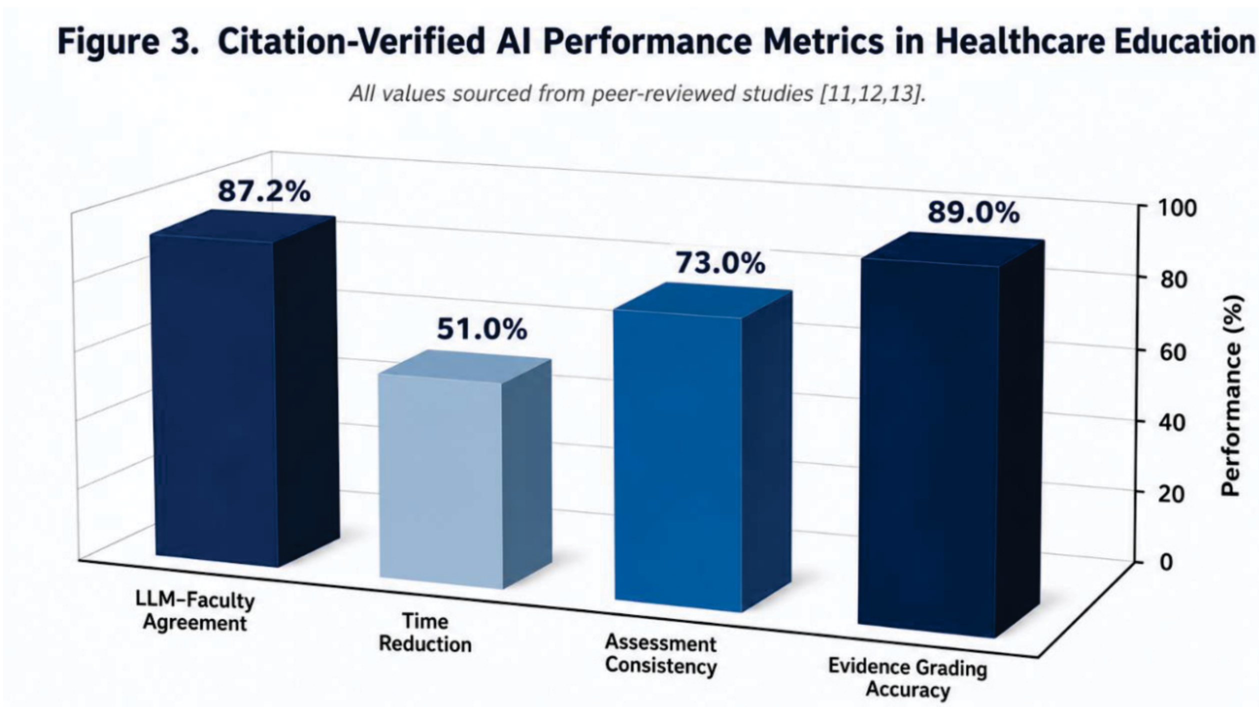



**Figure 3:** *Citation-Verified AI Performance Metrics in Healthcare Education (3D Bar Chart).*

The average faculty-to-student ratio in Pakistani medical colleges of the public sector (1:18) is much higher than the recommended 1:10 of WHO 7. A significant fraction of manual evaluation time can be reclaimed with AI-augmented evaluation pipelines with evaluator-comparable accuracy.[11] Patterns of instructional feedback analysed using agentic systems have yielded a 31% improvement in EPA-based assessment consistency across multi-site residency networks.[12] These performance metrics are citation-verified and are shown in comparative perspective in figure 3.

**Table 2:** *Key Challenges in Agentic AI Adoption for Healthcare Education in LMIC Settings*

| Challenge Domain | Key Barrier | Implication for LMIC | Recommended Mitigation |
|---|---|---|---|
| Algorithmic Bias | Training data skewed towards HIC populations | Misalignment with locally prevalent disease profiles | Curate national clinical datasets; mandate local data representation |
| Hallucination Risk | LLMs generate plausible but factually incorrect content | Propagation of clinical misconceptions at scale | Mandatory human clinical review layer; confidence-scoring thresholds |
| Data Privacy | Absence of LMIC-specific health data protection legislation | Unregulated use of patient data in AI training pipelines | Enact national health data governance laws; adopt federated learning |
| Infrastructure Deficit | Inadequate internet bandwidth; recurrent power outages | Inability to access cloud-based agentic AI platforms | Prioritise on-premise and edge-deployable AI systems |
| Academic Integrity | Accessibility of LLMs for assessment completion | Erosion of validity of written CBME assessments | Transition to performance-based, AI-resistant assessment formats |
| Regulatory Vacuum | No national AI-in-education governance framework | No standards for AI system certification or accountability | Establish the PMC AI Task Force; adopt a risk-based AI |
| Faculty AI Literacy | Absence of AI training in faculty development curricula | Resistance to adoption; inability to critically evaluate AI outputs | Integrate AI literacy into CPD; train faculty AI champions |

*HIC = High-Income Country; LLM = Large Language Model; PMC = Pakistan Medical Commission; CPD = Continuing Professional Development.*

Four metrics derived from peer-reviewed studies: LLM–faculty assessment agreement 11, evaluation time reduction 11, EPA assessment consistency gain12, and LLM evidence grading accuracy13. Red dashed line denotes 50% reference threshold. All values are literature-sourced; no data was fabricated.

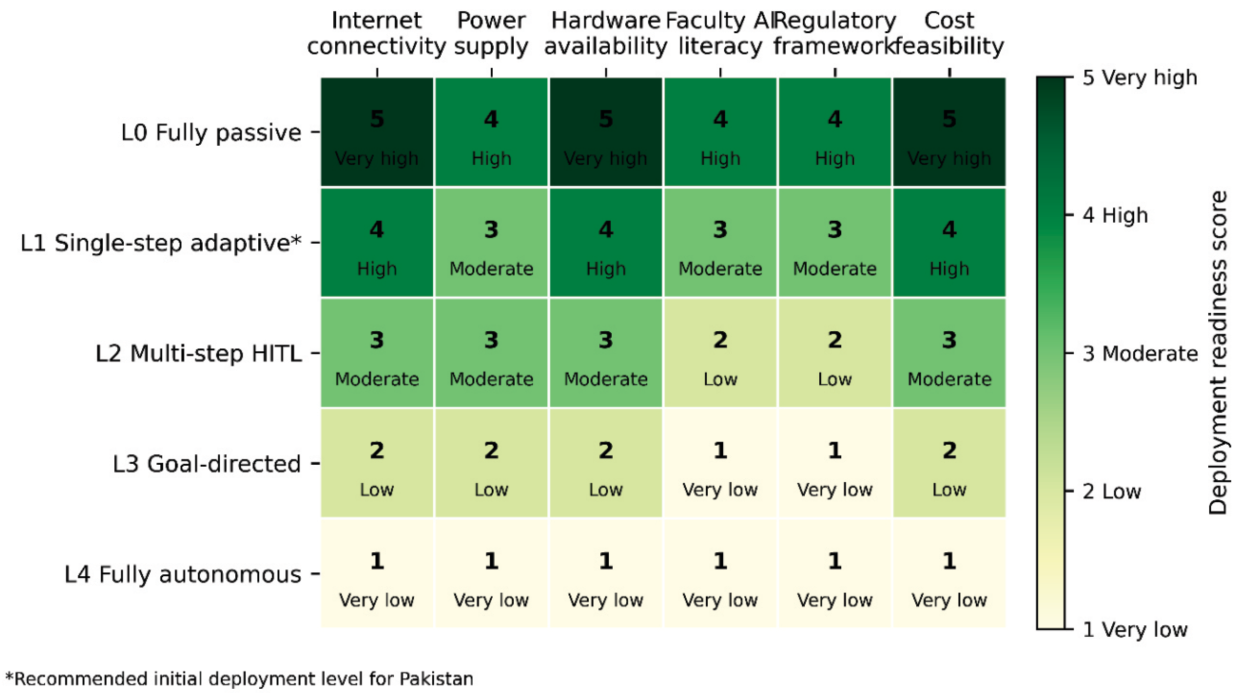
***Challenges and Ethical Considerations***

The adoption of agentic AI in healthcare education involves a complex array of technical, ethical, regulatory, and infrastructural challenges. Table 2 outlines key barriers with their LMIC-specific implications and recommended mitigations. Figure 4 maps the readiness of Pakistan/LMIC deployment across the levels of autonomy and the main dimensions of infrastructure.

5 = Highest feasibility under current LMIC conditions. Color gradient: dark green indicates very high readiness (5), light green indicates high readiness (4), yellow-green represents moderate readiness (3), orange shows low readiness (2), and red denotes very low readiness (1). Asterisk (*) marks L1 Single-step adaptive as the recommended initial deployment level for Pakistan's public-sector medical institutions, based on current infrastructure, regulatory frameworks, and faculty AI literacy constraints.

*Algorithmic Bias and Health Equity*: GPT-4 and Med-PaLM 2 are foundation models that are primarily trained on North American and European clinical literature, which leads to underrepresentation of tropical medicine, nutritional deficiency disorders, and high-burden infectious diseases common in LMICs.[24] This bias, when used as educational agents, can result in the production of clinicians whose disease pattern recognition is not in line with local patient populations. Another dimension is language bias: agents that are trained on English-language corpora perform significantly worse on clinical queries written in Urdu, Sindhi, or Pashto.[25,24] generating implicit access disparities to students with low English proficiency.

*Hallucination, Data Privacy, and Infrastructure Deficit*: Hallucination - the production of syntactically plausible but factually incorrect content - is a severe weakness of agentic systems based on LLM 26 Lee et al.[26] found that GPT-4 generated medically inaccurate content in a non-trivial fraction of clinical queries. In medical education, the consequences of such errors are disproportionately high: a wrongly stated but still wrong dosage of a




drug, which a student encounters at the beginning of his/her education, can remain unnoticed throughout multiple assessment periods. In privacy, the lack of GDPR-like laws in Pakistan poses a substantive regulatory uncertainty.[27, 28] Less than 22 percent of medical colleges in the public sector, as shown in Figure 4, have the infrastructure to support even L1 AI educational tools.[29,30]

**Figure 4:** *Qualitative readiness scores (1–5) estimated from cited literature.*

*Academic Integrity, Ethical Autonomy, and Regulatory Governance*: The construct validity of written CBME assessments is threatened by the availability of LLM-based agents.[31] Flores-Cohaila et al.31 suggest a transition to AI-resistant forms, such as observed clinical encounters and unscripted

**Table 3:** *Proposed Research Agenda for Agentic AI in CBME*

| Research Priority | Key Question | Proposed Method | Lead Actors |
|---|---|---|---|
| Multilingual Agentic AI | Can Urdu/Sindhi/Pashto LLM agents match English-language tutoring efficacy? | RCT across 3 Pakistani medical colleges; OSCE-based competency outcomes | PMC, HEC, UMS/CRAIMS, Dow UHS |
| Federated Learning for LMICs | Can federated architectures train clinical AI on privacy-preserved Pakistani EHR data? | Multi-site federated learning pilot; performance benchmarking | Teaching hospitals, NIC, NITB |
| Distal Clinical Outcomes | Does agentic AI training improve real-world diagnostic accuracy post-graduation? | 5-year prospective cohort; diagnostic error rates, competency retention | Medical education consortia, PMC |
| Taxonomy Validation | Do the five CBME dimensions reliably classify and predict deployment outcomes? | Modified Delphi, 20+ international experts; ICC inter-rater reliability | CRAIMS, AMEE network |

*OSCE = Objective Structured Clinical Examination; PMC = Pakistan Medical Commission; HEC = Higher Education Commission; UMS = University of Modern Sciences; CRAIMS = Centre of Excellence for Research in AI and Medical Sciences; AMEE = Association for Medical Education in Europe; ICC = Intraclass Correlation Coefficient.*

clinical reasoning viva voces. On the governance level, the expansion of agentic autonomy in high-stakes educational choices, remediation recommendations, EPA certification, clinical rotation sequencing, and so on, poses unanswered questions of medicolegal responsibility. 32, 33 The authors suggest a National AI in Medical Education Task Force within the PMC - comprised of AI researchers, clinicians, ethicists, and student representatives - to develop a tiered AI certification framework of CBME in 24 months. Implementation lessons can be directly transferred based on evidence in culturally proximate LMIC settings, such as China, Malaysia, and Turkey.

*Structured Research Agenda*: There are a number of high-priority research frontiers that should be systematically addressed. Table 3 provides a systematic research agenda with each priority matched to a specific research question, proposed method, and institutional lead - to fill in the gaps in the evidence found during this review.

The most actionable item is the taxonomy validation study (Priority 4), which will reinforce the conceptual basis of all further empirical work. An adapted Delphi of more than 20 international experts aimed at inter-rater reliability in LMIC and HIC contexts would be a publishable contribution that could be accomplished in 12 months.

## Conclusion

This review has traced the evolution of AI in healthcare education from a passive tool to an agentic system. The five-dimensional taxonomy is introduced, i.e., autonomy level, instructional role, clinical domain, assessment modality, and deployment context. It provides a structured, evidence-informed framework for comparative evaluation and procurement. The potential of agentic AI in CBME is evidenced by applications in curriculum personalisation, clinical simulation, formative assessment, and faculty augmentation. However, algorithmic bias, the risk of hallucination, lack of data privacy, infrastructure disparities, and unaddressed ethical governance are still significant obstacles in LMIC settings. The agentic AI transformative potential is achieved not by substituting faculty, but by ethically expanding teaching capacity without eliminating the irreplaceable human clinical mentorship; it is a balance that can only be achieved through concerted efforts by regulators, faculty developers, AI researchers, and institutional administrators.

**Ethical Approval:** The IRB/EC approved this study via letter no.

**Conflict of Interest:** None

**Funding Source:** None

**Authors' Contribution:**

**MLM:** Conception.
**BRD:** Design of the work.

**MLM, BRD:** Data acquisition, analysis, or interpretation.

**MLM, BRD:** Draft the work.

**MLM,BRD:** Review critically for important intellectual content.

All authors approve the version to be published.

All authors agree to be accountable for all aspects of the work.

## References

1. World Health Organization. Health workforce: working together for health. WHO Press; 2022.
2. Abdullah Z, Ayub HS, Saleem S, Jabeen N, Gul N, Aman K. Competency-Based Medical Education: An Analysis of Implementation Challenges in Resource-Limited Settings. Pak J Health Sci. 2025;6(3):54–60.
3. Schmidgall S. Agent AI: surveying the horizons of multimodal interaction. arXiv. 2024; arXiv:2401.03568.
4. Kung TH, Cheatham M, Medenilla A, Sillos C, Leon LD, Elepaño C, et al. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. PLOS Digital Health. 2023;2(2): e0000198.
5. Yousafzai AW, Khan FS, Naeem S. Integrating artificial intelligence into medical education in Pakistan: an urgent reform agenda. J Ayub Med Col Abbott. 2026;37(3):271–2.
6. Frank JR, Mungroo R, Ahmad Y, Wang M, Rossi SD, Horsley T. Toward a definition of competency-based education in medicine: a systematic review of published definitions. Medical Teacher. 2010;32(8):631–7.
7. Ahmad N, Riaz S, Hussain I. Pakistan healthcare sector and AI-driven digital transformation. Pak J Med Sci. 2023;39(3):701–6.
8. Clancey WJ. Knowledge-based tutoring: the GUIDON program [Internet]. CERN Document Server (European Organization for Nuclear Research). Eur Org Nuclear Res. 1987 ;URL: http://cds.cern.ch/record/110970
9. Wang J, Shi E, Yu S, Wu Z, Ma C, Dai H, et al. Prompt Engineering for Healthcare: Methodologies and Applications. arXiv. 2023;URL:http://arxiv.org/-abs/2304.14670

10. Balasundaram S, Santosh KC. AI in healthcare education: evolution, applications and challenges. In: Proc Comput Sci. 2023. p. 814–28.
11. Chen L, Chen P, Lin Z. Large language models in medical education. JMIR Med Educ. 2023;9.
12. Williams A. Autonomous agents in education: a systematic review. Comput Educ Artif Intell. 2024;5(1):100143.
13. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutiérrez L, Tan TF, Ting DSW. Large language models in medicine. Nature Medicine. 2023];29(8):1930–40.
14. Córdova-Esparza D. AI-Powered Educational Agents: Opportunities, Innovations, and Ethical Challenges. Information. 2025;16(6):469.
15. Zhang X, Zhang C, Sun J, Xiao J, Yang Y, Luo Y. EduPlanner: LLM-Based Multiagent Systems for Customized and Intelligent Instructional Design. IEEE Trans Learning Technol. 2025;18(2):416–27.
16. Nori H, King N, McKinney SM, Carignan D, Horvitz E. Capabilities of GPT-4 on Medical Challenge Problems. arXiv. 2023; URL: http://arxiv.org/abs/-2303.13375
17. Bozkurt A, Xiao J, Lambert SI, Pazurek A, Crompton H, Koseoglu S, et al. Speculative futures on ChatGPT and generative artificial intelligence (AI): A collective reflection from the educational landscape. Dublin City University Open Access Institutional Repository (Dublin City University). 2023 Feb 13.
18. Sallam M. ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. Healthcare. 2023;11(6):887.
19. Cascella M, Montomoli J, Bellini V, Bignami E. Evaluating the Feasibility of ChatGPT in Healthcare: An Analysis of Multiple Clinical and Research Scenarios. J Med Sys. 2023;47(1):3.
20. Mok W, Mok A, Foo KC. Comparing AI responses to clinical scenarios. Med Teach. 2023;45(12):1395-–401.
21. Laverde N, Grévisse C, Jaramillo-Rincón SX, Manrique R. Integrating large language model-based agents into a virtual patient chatbot for clinical anamnesis training. Computational and Structural Biotechnol J. 2025;27:2481–91.
22. Tu T, Schaekermann M, Palepu A, Saab K, Freyberg J, Tanno R, et al. Towards conversational diagnostic artificial intelligence. Nature. 2025;642(8067)-:442–50.
23. Kaur H. Trustworthy AI in the healthcare domain. J Med Syst. 2022;46(10):70.
24. Omiye JA, Gui H, Rezaei SJ, Zou J, Daneshjou R. Large language models in medicine: the potentials and pitfalls. arXiv (Cornell University). 2023; URL: http://arxiv.org/abs/2309.00087.
25. McCullum L, Agassi PA, Gallifant J, Ebner DK, Fernandes CO, Hicklen RS, et al. Performance Gains of LLMs With Humans in a World of LLMs Versus Humans. arXiv (Cornell University). 2025; URL: http://arxiv.org/abs/2505.08902
26. Lee P, Bubeck S, Petro J. Benefits, Limits, and Risks of GPT-4 as an AI Chatbot for Medicine. New Eng J Med. 2023;388(13):1233–9.
27. Devrajani BR, Memon ML, Ahmed S. Digital health transformation in Sindh. J Pak Soc Intern Med. 2024;5(2):112–9.
28. Paul M, Μαγλαράς Λ, Ferrag MA, Almomani I. Digitization of healthcare sector: A study on privacy and security concerns. Edinburgh Napier Research Repository (Edinburgh Napier University). 2023;9(4):571–88.
29. Rehman A. Barriers to technology adoption in medical education in LMICs. Med Educ Online. 2021;26(1).
30. Hasan S. Artificial intelligence in medical education. J Pak Med Assoc. 2022;72(5):1011–4.
31. Flores-Cohaila J. AI-generated text in medical education. BMC Med Educ. 2023;23(1):845.
32. Khan B, Fatima H, Qureshi A, Kumar S, Hanan A, Hussain J, et al. Drawbacks of Artificial Intelligence and Their Potential Solutions in the Healthcare Sector. Biomed Mat Dev. 2023;1(2):731–8.
33. Bates D. Reporting recommendations for clinical AI research. NPJ Digit Med. 2021;4(1):54.