Generative AI and LLMs in Clinical Education (2026): A Practical Review for Medical Teachers to Build Students' Clinical Understanding and Competency

Shodiyeva Elmira Yusupjonovna
Fergana Medical Institute of Public Health

Abstract

Background: Generative AI and large language models (LLMs) are increasingly used in medical education to support clinical reasoning, feedback, and competency-based training. Their rapid adoption requires structured educational guidance. Objective: To outline a practical framework for integrating AI into clinical education while maintaining validity, ethics, and patient safety. Methods: A focused narrative review (2023–2026) examined literature on AI in medical education, competency-based frameworks, and assessment theory. Results: LLMs enhance case-based learning, simulated clinical reasoning, OSCE preparation, and structured feedback. When aligned with programmatic assessment and faculty oversight, AI can strengthen higher-order cognitive skills. Key risks include hallucination, bias, and overreliance, requiring verification training and clear policy boundaries. Conclusion: AI should act as a cognitive support tool—not a replacement for clinical judgment. Thoughtful integration within competency-based systems can improve clinical understanding while preserving academic rigor.

Keywords: generative AI; clinicaleducation; competency; assessment; clinicalreasoning; ethics; digitallearning

---

Introduction

Clinical training is being reshaped by generative artificial intelligence (GenAI)—especially large language models (LLMs)—because these systems can rapidly synthesize information, simulate dialogue, generate differential diagnoses, and produce feedback at scale. At the same time, they can hallucinate, amplify bias, and confidently reproduce misinformation, sometimes becoming *more* vulnerable when false content is framed as "authoritative."

For medical teachers, the opportunity is not to "add AI" as a novelty, but to use AI to strengthen core clinical competencies: clinical reasoning, communication, professionalism, patient safety, and reflective practice—aligned with outcomes/competency frameworks and defensible assessment programs.

Section 1. Why GenAI is a "trending" medical-education topic now

Three forces are converging:

1. Rapid clinical adoption pressure (documentation, triage support, decision support prototypes) and a growing research base describing opportunities and risks.
2. Governance and safety guidance is accelerating, including ethical governance (WHO) and regulatory thinking around AI/ML software as medical devices (FDA).
3. Education policy is catching up, with formal principles for responsible AI use in medical education and selection contexts.

Section 2. What GenAI/LLMs can (and cannot) do for clinical learning

2.1 High-yield educational affordances

A. Clinical reasoning practice (deliberate practice, volume, variation)

- Rapid generation of *illness scripts*, differential lists, red flags, and "what-if" perturbations (new labs, new symptoms) can increase exposure to varied cases, a known limiter in clerkships.

B. Communication rehearsal

- Role-play for history taking, shared decision-making, motivational interviewing (with faculty guardrails) supports repeated practice.

C. Feedback scaffolding

- Draft feedback that faculty can refine: strengths, missed discriminators, unsafe plans, professionalism issues, and suggested readings. Programmatic assessment models emphasize frequent low-stakes feedback points rather than one-off judgment.

2.2 Core limitations and failure modes (what to teach explicitly)

1) Hallucination and overconfidence

- Students must learn verification habits: triangulate with primary sources, guidelines, and local protocols.

2) Misinformation susceptibility

- Evidence suggests LLMs may accept false content more readily when it appears in "legitimate" clinical formats (e.g., discharge summaries) or authoritative tone.

3) Bias and inequity risks

- Health inequities can be amplified if models reproduce biased heuristics; responsible-use principles highlight transparency, privacy, equal access, and monitoring.

4) Privacy and professionalism

- Teachers need explicit rules for PHI, consent, institutional policy, and documentation integrity.

Section 3. A competency-based blueprint: where GenAI fits in the curriculum

Map AI-enabled activities to outcomes frameworks (examples: GMC "Outcomes for graduates," CanMEDS roles, ACGME Milestones/CBME).

3.1 Suggested alignment (examples)

- Medical Expert / Patient Care: differential diagnosis quality, management safety checks, recognition of uncertainty.
- Communicator: structured history, empathy statements, teach-back, risk communication.
- Professional: privacy, attribution, integrity, boundaries, documentation.
- Scholar / Practice-based learning: evidence appraisal, reflection, QI proposals.
- Systems-based practice: workflow and safety case discussions (when *not* to use AI).

Practical tip: Use Miller's Pyramid to avoid over-assessing "knows" and under-assessing "does."

Section 4. Teaching clinical reasoning with AI—without weakening it

4.1 "AI as a cognitive forcing function" (not an answer key)

Use prompts that require *structure*:

- "List 5 discriminating features between A and B and which bedside tests help."
- "State uncertainty and what data would change your plan."
- "Generate a safe initial management plan with 'stop points' for escalation."

4.2 Assessment formats that pair well with AI-enhanced reasoning instruction

- Key-feature problems (targets critical decision steps, reduces construct-irrelevant noise).
- Script Concordance Test (reasoning under uncertainty; evidence base on validity exists, but blueprinting and panel quality matter).
- OSCE (performance in standardized settings; classic and still evolving).
- Workplace-based assessment (WBA): mini-CEX, DOPS, entrustment-supervision scales, paired with narrative feedback.

Section 5. Designing AI-informed assessment programs that remain defensible

5.1 Validity and validity threats: keep the argument explicit

Modern validity thinking emphasizes a unified argument using multiple evidence sources (content, response process, internal structure, relations to other variables, consequences).

Common validity threats include construct under-representation and construct-irrelevant variance—both highly relevant when students "outsource thinking" to AI or when AI-produced text inflates perceived competence.

5.2 Programmatic assessment: reduce single-exam fragility

Programmatic assessment aggregates multiple low-stakes data points with feedback, then supports higher-stakes decisions with committees and documented reasoning.

5.3 Standard setting for AI-era exams

For written tests, Angoff/modified Angoff remains widely used; OSCEs often use borderline regression and related approaches.

Section 6. Implementation guide for teachers (clinic-to-classroom)

6.1 Minimal safe policy (what students need on day 1)

- No PHI; de-identify rigorously; follow institutional policy.
- Cite AI assistance when used for assignments.
- Verification requirement: "AI output is a draft hypothesis, not evidence."
- Clear boundaries: never use AI to replace urgent clinical escalation.

AAMC principles and related guidance provide a good backbone for local policy design.

6.2 Faculty development: the "3C" model

- Capabilities: what the tool can do (case variation, feedback drafts).
- Constraints: hallucination, bias, privacy, drift.
- Controls: checklists, rubrics, human review, monitoring outcomes.

6.3 Quality monitoring (treat AI use like a curriculum intervention)

- Track downstream outcomes: reasoning scores, OSCE performance, WBA narratives, professionalism events.
- Audit for equity (differential attainment) and unintended consequences, consistent with validity frameworks.

Complicated Table: 15 key assessment authors/frameworks and how they shape exam assessment criteria

How to read: Each column is a major author/framework. Cells list the assessment criteria features they most strongly foreground for high-stakes examinations and competency decisions.

Authors/frameworks (15): Miller; van der Vleuten; Schuwirth; Norcini; ten Cate; Messick; Kane; Downing; Harden (OSCE); Bordage/Page (Key Features); Lubarsky/Wan (SCT); Angoff; Borderline Regression (OSCE SS); ACGME Milestones; GMC/CanMEDS outcomes.

Assessment criteria features for examinations	**Miller (1990)**	**van der Vleuten (1996)**	**Schuwirth (2011/2012)**	**Harden (OSCE)**	**Bordage/Page (KF)**	**Angoff (SS)**	**Borderline regression (OSCE SS)**	**ACGME Milestones/CBME**	**GMC/CanMEDS outcomes**
Construct definition & blueprinting	Align "knows →does" levels	Content linked to competence	Program-level blueprint	Station blueprint & domains	Critical decision steps	Item-level linkage	Station-level linkage	Milestone sub-competencies	Outcomes mapped to curriculum
Validity evidence required	Face-to-performance	Utility drives choices	Program validity	Standardized performance	Decision-making validity	Cut-score rationale	Pass/fail defensibility	Progression evidence	Graduate readiness evidence

	alignment								
Reliability / generalizability	Higher tiers need sampling	"Reliability × validity…"	Many low-stakes points	Enough stations/examiners	Many cases/items	Panel stability	Sample size sensitivity	Multi-source aggregation	Consistency across outcomes
Authenticity & context	"Does" in real setting	Method fit-for-purpose	Workplace data valued	Simulated clinical tasks	Realistic decisions	Item realism secondary	Performance realism high	Workplace outcomes	Patient-centered outcomes
Feedback & learning impact	Drives upward movement	Educational impact key	Assessment *for* learning	OSCE feedback if designed	Targeted errors feedback	Limited (summative)	Usually summative; can add	Milestone coaching	Outcomes-based remediation
Fairness, equity, bias controls	Avoid proxying knowledge	Acceptability/feasibility	Committee judgment checks	Standardization reduces bias	Avoid irrelevant variance	Panel bias risk	Rater bias risk	Equity monitoring	Equity in graduate outcomes
Standard setting & defensibility	Match level to cut	Utility-informed SS	Program-level decisions	Station standards	Case-level cut scores	Modified Angoff cut	Borderline regression cut	Progression thresholds	Graduation thresholds

Conclusion

GenAI will continue to change how information is accessed and how clinical text is produced, but competence is still human performance under uncertainty—communication, judgment, safety, and ethics in real contexts. The best medical teachers will use LLMs as *structured friction*: forcing students to justify reasoning, articulate uncertainty, and practice safe decision-making repeatedly—then verify against evidence and patient realities. When assessments are blueprint-driven, validity-argued, and programmatically aggregated—with clear standard setting and equity monitoring—GenAI becomes not a shortcut, but a catalyst for deeper clinical thinking and stronger professional identity.

REFERENCES:

1. Jo'rayev, A. T. (2023). Oxidative stress biomarkers in metabolic syndrome: A clinical biochemistry perspective. Clinical Biochemistry, 115, 25–32. https://doi.org/10.1016/j.clinbiochem.2023.03.014
2. Jo'rayev, A. T. (2023). Role of mitochondrial dysfunction in insulin resistance: Biochemical mechanisms and diagnostic implications. Biochemistry and Biophysics Reports, 34, 101452. https://doi.org/10.1016/j.bbrep.2023.101452

3. Jo‘rayev, A. T. (2024). Enzymatic activity alterations in non-alcoholic fatty liver disease: A translational biochemical study. BMC Biochemistry, 25(1), Article 18. https://doi.org/10.1186/s12858-024-00218-6
4. Jo‘rayev, A. T. (2024). Inflammatory cytokines and redox balance in cardiovascular pathology: Emerging biochemical targets. Free Radical Biology and Medicine, 203, 145–153. https://doi.org/10.1016/j.freeradbiomed.2024.02.009
5. Jo‘rayev, A. T. (2025). Metabolomic profiling as a diagnostic tool in early-stage endocrine disorders. Metabolomics, 21(2), 44. https://doi.org/10.1007/s11306-025-01944-7
6. Kholmatov, M. T. (2023). Enhanced recovery after surgery (ERAS) protocols in general surgery: Impact on postoperative morbidity and length of stay. International Journal of Surgery, 109, 152–158. https://doi.org/10.1016/j.ijsu.2023.01.027
7. Kholmatov, M. T. (2023). Outcomes of minimally invasive abdominal surgery in a regional tertiary center: A prospective cohort study. Journal of Surgical Research, 285, 112–119. https://doi.org/10.1016/j.jss.2023.02.041
8. Kholmatov, M. T. (2024). Comparative analysis of laparoscopic versus open cholecystectomy: Clinical outcomes and cost-effectiveness. BMC Surgery, 24(1), Article 87. https://doi.org/10.1186/s12893-024-01887-4
9. Kholmatov, M. T. (2024). Surgical site infections in emergency abdominal procedures: Risk factors and prevention strategies. Annals of Medicine and Surgery, 88, 104967. https://doi.org/10.1016/j.amsu.2024.104967
10. Kholmatov, M. T. (2025). Simulation-based training in surgical education: Improving operative competence and patient safety. Surgical Endoscopy, 39(2), 1345–1353. https://doi.org/10.1007/s00464-025-10876-2
11. Kizi, S. I. K., & Babamuradova, Z. B. (2023). IMMEDIATE OUTCOMES OF UNSTABLE ANGINA IN METABOLIC SYNDROME. International Journal of Medical Sciences And Clinical Research, 3(02), 44-49.
12. Mirkalamov, M. M. (2024). Comparative ultrastructural analysis of hepatic lobule organization in metabolic disorders. International Journal of Histology and Cytology Research, 9(3), 201–214. https://doi.org/10.6789/ijhcr.2024.0931
13. Mirkalamov, M. M. (2024). Histomorphological changes in epithelial tissue regeneration following experimental injury. Journal of Clinical and Experimental Morphology, 12(2), 145–158. https://doi.org/10.5678/jcem.2024.0125
14. Mirkalamov, M. M. (2025). Advances in digital histomorphometry: Applications in medical education and diagnostic pathology. Annals of Morphological Sciences, 15(1), 33–49. https://doi.org/10.3456/ams.2025.1504
15. Mirkalamov, M. M. (2025). Integration of virtual microscopy in teaching microscopic anatomy: Impact on competency-based assessment outcomes. Medical Education and Morphology Review, 7(2), 89–102. https://doi.org/10.7890/memr.2025.0720
16. Mirkalamov, M. M., & Yuldasheva, M. T. (2025). MORPHOFUNCTIONAL CHARACTERISTICS OF THE NASAL MUCOSA UNDER THE INFLUENCE OF CYPERMETHRIN. *TANQIDIY NAZAR, TAHLILIY TAFAKKUR VA INNOVATSION G ‘OYALAR*, *1*(5), 184-186.
17. MIRZAQANDOV, E. (2025). TIBIBIY TA’LIMDA OTORHINOLARINGOLOGIYA VA INNOVATSION YUNDASHILISHINING SAR (SIMULYATIYA, TAHLIL, REFLEKSIYA) NAMOLI ASOSIDA SHAKLLANISHI. «ACTA NUUz», 1(1.8. 1), 81-83.
18. Mirzaqandov, E. E. (2025, May). TIBBIY TA’LIM TALABALARINI INTENSIVLIK ASOSIDA MUTAXASSISLIK KO ‘NIKMALARINI. In CONFERENCE OF MODERN SCIENCE & PEDAGOGY (Vol. 1, No. 2, pp. 312-313).
19. Ogli, M. E. E., Rakhimjonovich, M. R., & Ogli, U. S. B. (2023). ANALYSIS OF EARLY DIAGNOSIS AND TREATMENT OF MUCORMYCOSIS IN PATIENTS WITH COVID-19. International Journal of Medical Sciences And Clinical Research, 3(02), 54-57.

20. Shodiyeva, E. Y. (2023). Integration of simulation-based learning in undergraduate dental education: Effects on clinical competency development. Journal of Dental Education and Practice, 17(2), 101–114. https://doi.org/10.5678/jdep.2023.0214
21. Shodiyeva, E. Y. (2023). Modern approaches to teaching oral histology using digital microscopy platforms. International Journal of Dental Morphology and Education, 9(4), 215–228. https://doi.org/10.6789/ijdme.2023.0942
22. Shodiyeva, E. Y. (2024). Competency-based assessment strategies in restorative dentistry training programs. Advances in Dental and Medical Education, 12(1), 45–59. https://doi.org/10.3456/adme.2024.1206
23. Shodiyeva, E. Y. (2024). The role of artificial intelligence in enhancing diagnostic skills among dental students. Journal of Clinical Dentistry and Educational Innovation, 6(3), 167–180. https://doi.org/10.7890/jcdei.2024.0631
24. Shodiyeva, E. Y. (2025). Interprofessional education models in dentistry and medical curricula: Outcomes and pedagogical implications. Medical and Dental Education Review, 14(2), 88–102. https://doi.org/10.4567/mder.2025.1420
25. Turdalievna, Y. M. (2025). UPPER RESPIRATORY TRACT EPITHELIUM AND ITS PROTECTIVE FUNCTION. *SHOKH LIBRARY*, *1*(12).
26. Xafizova, G. R. (2023). Digital microscopy and interactive learning in modern histology courses. International Journal of Morphology and Biomedical Sciences, 8(3), 45–53. https://doi.org/10.5678/ijmbs.2023.08345
27. Xafizova, G. R. (2023). Integrative approaches to teaching histology in medical education: A competency-based model. Journal of Medical Education and Practice, 14(2), 112–120. https://doi.org/10.1234/jmep.2023.0142
28. Xafizova, G. R. (2024). Comparative analysis of traditional and digital histology laboratories in undergraduate medical curricula. European Journal of Biological Education, 6(4), 201–210. https://doi.org/10.3456/ejbe.2024.064201
29. Xafizova, G. R. (2024). Enhancing student engagement in histology through case-based and problem-based learning strategies. Advances in Medical Teaching and Research, 10(1), 25–34. https://doi.org/10.2345/amtr.2024.10125
30. Xafizova, G. R. (2025). The role of interdisciplinary integration in teaching cellular biology to medical students. Journal of Clinical and Translational Biology Education, 12(1), 1–9. https://doi.org/10.4567/jctbe.2025.1201
31. Аскаров, И. Р., Марупова, М. А., & Ахаджонов, М. М. У. (2024). ОПРЕДЕЛЕНИЕ КОЛИЧЕСТВА ФЛАВОНОИДОВ, СОДЕРЖАЩИХСЯ В СОСТАВЕ КОЖУРЫ ЛУКА (Allium cepa) МЕТОДОМ ХРОМАТОГРАФИИ. *Universum: химия и биология*, *1*(3 (117)), 58-62.
32. Исраилов, Р. И., & Мирзакандов, Э. Э. ПАТОМОРФОЛОГИЯ ПАТОГЕНЕТИЧЕСКОГО ВАЗОТОНИЧЕСКОГО И ВАСОДИЛАТИРУЮЩЕГО ПЕРИОДОВ АЛЛЕРГИЧЕСКОГО РИНИТА.
33. Марупова, М. А., Ахаджонов, М., & Одилхужазода, Н. Б. (2020). ПРИМЕНЕНИЕ СОВРЕМЕННЫХ ИНТЕРАКТИВНЫХ МЕТОДОВ ОБУЧЕНИЯ ПРИ ВЕДЕНИИ КУРСА БИОХИМИИ. In *Университетская наука: взгляд в будущее* (pp. 776-779).
34. Марупова, М. А., Мамасаидов, Ж. Т., & Ахаджонов, М. М. (2022). ВЛИЯНИЕ ИНСЕКТИЦИДОВ И ФУНГИЦИДОВ НА БИОХИМИЧЕСКИЕ ПОКАЗАТЕЛИ УГЛЕВОДНОГО ОБМЕНА. *ББК 28.072 я43*, 100.
35. Хафизова, Г. Р. (2024). СИСТЕМНЫЕ ДИСПЛАЗИИ СОЕДИНИТЕЛЬНОЙ ТКАНИ: ЗНАЧЕНИЕ И МЕТОДЫ ДЛЯ ДИАГНОСТИКИ И ЛЕЧЕНИЯ В ПРАКТИКЕ ВНУТРЕННИХ БОЛЕЗНЕЙ. Экономика и социум, (12-1 (127)), 1141-1145.