An Adaptive Immersive AI Framework for Clinical Training and Performance Evaluation Using Graph-Augmented Large Language Models

[1]Sriharsha Makineni
[1]Independent Researcher Georgetown, TX, USA

Received: 20th Nov 2025 | Received Revised Version: 22nd Dec 2025 | Accepted: 23rd Jan 2026 | Published: 11th Feb 2026

Volume 08 Issue 02 2026 | Crossref DOI: 10.37547/tajet/v8i2-321

Abstract

Healthcare education faces critical challenges including expanding medical knowledge, evolving clinical guidelines, and limited supervised clinical experience. While artificial intelligence (AI) and immersive technologies offer promising solutions, current systems operate in isolation rather than synergistically. This paper presents a systematic review of 165 AI-enabled clinical training systems (2022-2025), examining five technological dimensions: LLM-based virtual patients, VR/AR platforms, graph-augmented AI, adaptive learning, and performance evaluation. Analysis reveals significant fragmentation—67% of systems employ single approaches, only 7% integrate three or more technologies. Critical gaps include: factual hallucinations in LLMs (12-15% error rates), limited adaptability in VR/AR scenarios, single-dimension personalization, and domain-specific assessments. We propose an integrated architectural framework combining graph-augmented LLMs with immersive interfaces, multi-dimensional adaptive learning, and comprehensive performance evaluation. Graph augmentation demonstrates 73% reduction in factual errors; VR/AR systems show 25-40% skill retention improvements. The modular framework addresses identified gaps through bidirectional data flows and evidence-based component integration, providing a research-informed blueprint for next-generation clinical training systems.

Keywords: clinical training, large language models, knowledge graphs, virtual reality, augmented reality, adaptive learning, medical education, systematic review.

Cite This Article: Makineni, S. (2026). An Adaptive Immersive AI Framework for Clinical Training and Performance Evaluation Using Graph-Augmented Large Language Models. The American Journal of Engineering and Technology, 8(2), 87–95. https://doi.org/10.37547/tajet/v8i2-321.

1. Introduction

1.1 Motivation

Medical education confronts escalating complexity: exponential knowledge growth, dynamic treatment paradigms, and restricted hands-on clinical exposure. Traditional apprenticeship models, while effective, lack scalability and demonstrate high variability across institutions [22]. Technology-enabled solutions—conversational AI for patient simulation, immersive VR/AR for procedural training, and adaptive algorithms for personalized learning—have emerged independently but remain largely disconnected.

Current systems exhibit critical limitations: LLM-based virtual patients generate medically inaccurate information without knowledge grounding [15]; VR/AR platforms deliver realistic environments but lack intelligent scenario adaptation [1]; adaptive learning systems adjust single parameters rather than comprehensive educational experiences [13]. This fragmentation prevents realization of synergistic benefits that integrated architectures could provide.

1.2 Contributions

This work makes three principal contributions: (1) systematic review and classification of 165 recent AI-enabled clinical training systems across five technological dimensions; (2) comprehensive gap analysis identifying integration, validation, and scalability deficiencies; (3) evidence-informed architectural framework integrating graph-augmented LLMs, immersive simulation, multi-dimensional adaptation, and holistic assessment. The proposed framework synthesizes empirical findings rather than presenting speculative design, establishing a research foundation for next-generation clinical training system development.

2. Methods

2.1 Literature Search Strategy

A structured search captured AI-enabled clinical training systems across three complementary databases: SciSpace (peer-reviewed journals), Google Scholar (conference proceedings, early-access articles), and arXiv (AI/CS preprints). Search terms combined five thematic areas: *AI Technologies* (Artificial Intelligence, Machine Learning, LLM, GPT, Knowledge Graphs); *Immersive Technologies* (VR, AR, Mixed Reality, Haptics); *Clinical Training* (Medical Education, Simulation, Virtual Patient); *Adaptive Systems* (Personalized Learning, Intelligent Tutoring); *Evaluation* (Performance Assessment, Competency). Boolean operators constructed compound queries targeting technological intersections. Timeframe: 2022-2026 publications, searched late 2024-early 2025.

2.2 Study Selection

Initial retrieval yielded 646 publications. Following duplicate removal (n=127), title/abstract screening eliminated 354 irrelevant studies. Full-text review of remaining articles yielded 165 relevant studies based on inclusion criteria: (1) focus on AI/ML systems for clinical training; (2) detailed system architectures or methodologies; (3) coverage of LLMs, knowledge graphs, VR/AR, adaptive learning, or assessment; (4) implementation details or empirical findings; (5) published post–January 2022; and (6) sufficient technical detail. Exclusion criteria included clinical decision support without an educational component, purely theoretical work, pre–2022 publications, insufficient technical detail, and duplicates. Interrater agreement was high (Cohen's $\kappa = 0.87$). The 165 studies were ranked by relevance across five technological dimensions; the top 30 underwent in-depth architectural analysis.

2.3 Data Extraction and Quality Assessment

Structured extraction covered: (1) *Technical Architecture*: AI/ML models, system components, knowledge integration methods, immersive technologies, adaptation mechanisms, evaluation methods; (2) *Clinical Application*: target users, trained skills, domains, outcomes, study design; (3) *Contributions/Limitations*: technical novelty, identified gaps, future directions. Quality assessment adapted MERSQI criteria [13], evaluating study design, sample characteristics, data type, validity evidence, and analytical rigor. Studies classified as High (n=9), Medium (n=16), or Low (n=5) quality; lower quality studies weighted less in synthesis.

2.4 Synthesis Approach

Narrative synthesis addressed heterogeneous study designs and measurement tools. Analysis categorized findings into five technology types, examining: architectural patterns, quantitative effectiveness data, capability gaps, and integration opportunities. This approach directly informed the proposed framework, grounding design in empirical evidence rather than assumptions.

3. Results and Gap Analysis

3.1 Overview of Current Landscape

The systematic review revealed a fragmented ecosystem with promising individual technologies but minimal crosspollination. Of 30 in-depth analyzed studies: 0 integrated all five technological approaches, 2 (7%) combined three approaches, 8 (27%) used two approaches, and 20 (67%) focused on single approaches. Table I synthesizes key findings across five categories.

3.2 LLM-Based Virtual Patient Systems

Six studies (n=247) demonstrated LLMs' capacity for realistic patient dialogue [15]–[17]. Advanced implementations achieved 85% contextually appropriate responses and 4.2/5.0 usability ratings [20]. The MEDCO framework introduced multi-agent architecture separating patient, knowledge, evaluation, and feedback agents [17]. However, all systems exhibited factual hallucinations (12-

15% error rates) and lacked integration with procedural training or medical knowledge bases.

3.3 Immersive VR/AR Training Platforms

Eleven studies (n=438) established VR/AR effectiveness for procedural skills [1], [7], [8]. Quantitative outcomes included 34% better retention at 3-month follow-up, 28% faster procedure completion, and 42% fewer critical errors compared to mannequin training. Cost-effective implementations using consumer-grade hardware ($1000 total) matched $50,000 systems in skill acquisition [7]. CLiVR demonstrated hybrid approaches integrating AI-powered virtual patients into VR environments [6]. AR applications showed 31% anatomical knowledge improvement [11]. Primary limitation: 9/11 systems used pre-scripted, non-adaptive scenarios lacking intelligent patient responses.

3.4 Graph-Augmented AI Systems

Only two studies addressed LLM hallucination through structured medical knowledge integration [2], [21]. MedCT's clinical terminology graph (400,000+ terms, 2M+ relationships) constrained LLM generation, reducing factual errors by 73% [2]. AIPatient utilized de-identified EHR data to construct patient knowledge graphs driving coherent virtual patient simulations; residents rated graph-augmented patients as significantly more realistic than script-based alternatives (n=34) [21]. Despite dramatic accuracy improvements, neither study explored integration with immersive technologies or comprehensive adaptive learning.

3.5 Adaptive Learning Mechanisms

Five studies (n=285) demonstrated personalization benefits [12]–[14]. Socratic AI's reinforcement learning-based tutoring achieved 18% greater clinical reasoning improvement versus static materials, with largest gains among low-performers (n=67) [13]. Predictive models identified at-risk students with 82% accuracy, enabling interventions that reduced failures by 73% (n=94) [14]. Temporal skill tracking provided real-time feedback, accelerating acquisition by 29% [12]. Limitations: most systems adapted single dimensions (difficulty *or* feedback timing); only one integrated VR/AR [10].

3.6 Performance Evaluation Methods

Four studies (n=178) explored AI-based competency assessment [3], [12], [14], [15]. Temporal AI models enabled continuous micro-assessments tracking skill acquisition, plateaus, and decay [12]. NLP-based communication evaluation achieved 78% agreement with expert raters [15]. Root cause analysis training assessed complex cognitive skills beyond correctness [3]. Machine learning identified specific weaknesses (e.g., rushed information gathering) invisible to traditional assessments [14]. Gaps: domain-specificity, moderate expert agreement (65-80%), lack of standardization, minimal real-time adaptation integration.

3.7 Critical Gaps Identified

Integration Gap: Technological fragmentation prevents synergistic benefits. No studies combined all five approaches; 67% employed single technologies despite clear complementarity.

Validation Gap: Limited empirical rigor: 37% included comparison groups, 17% reported follow-up assessments, 10% correlated simulation with clinical outcomes. Median sample size: 42 participants.

Scalability Gap: 63% were single-institution pilots; 13% multi-institutional. Insufficient attention to deployment costs,

Table I Summary Of Ai-Enabled Clinical Training Technologies

Category	Studies	n	Key Strengths	Critical Limitations
LLM Virtual Patients [15]–[18], [20]	6	247	Realistic dialogue (85% contextual accuracy), infinite patience, scalability, communication feedback	Factual hallucinations (12–15%), no clinical grounding, limited skills integration, text/voice only
VR/AR Platforms [1], [6]–[8], [10], [11]	11	438	Skill retention +25–40%, learning speed +20– 35%, error reduction 15–45%, hands-on practice	Pre-scripted scenarios, hardware costs, motion sickness (8–12%), limited clinical complexity
Graph-Augmented AI [2], [21]	2	89	Factual error reduction 73%, explainable reasoning, clinical consistency, updatable knowledge	Emerging area, development complexity, limited integration, validation challenges
Adaptive Learning [9], [10], [12]–[14]	5	285	Time-to-competency –20–30%, benefits struggling learners, detailed performance insights	Single-dimension adaptation, rare VR/AR integration, black-box algorithms, large data requirements
Performance Evaluation [3], [12], [14], [15]	4	178	Multi-dimensional assessment, granular feedback, continuous tracking, reduced evaluator burden	Domain-specific focus, 65–80% expert agreement, no standardization, limited real-time adaptation

diverse populations, or institutional constraints.

Clinical Grounding Gap: LLM systems lack medical knowledge integration, generating factually incorrect information that graph augmentation dramatically reduces (73%).

Adaptability Gap: Most systems offer narrow personalization (single parameter adjustment) rather than comprehensive multi-dimensional adaptation across scenario selection, difficulty, feedback, and environment.

4. Proposed Architectural Framework

4.1 Design Philosophy and Overview

Based on systematic evidence synthesis, we propose an Integrated Architectural Framework for Next-Generation AIDriven Clinical Training (Fig. 1). Design principles include: (1) *Evidence-based integration*: components reflect demonstrated capabilities; (2) *Modularity*: incremental implementation; (3) *Clinical safety*: knowledge-grounded to prevent misinformation; (4) *Learner-centered*: educationally- driven adaptation; (5) *Scalability*: practical deployment considerations.

The framework integrates four components: *Graph Augmented LLM Engine* (clinically-grounded patient simulation), *Immersive Interface Layer* (experiential multimodal environments), *Adaptive Learning Module* (multi-dimensional personalization), and *Performance Evaluation System* (comprehensive competency assessment). This addresses the integration gap where 67% of reviewed systems employed single approaches.

4.2 Component 1: Graph-Augmented LLM Engine

Purpose: Generate clinically accurate, context-consistent patient simulations grounded in validated medical knowledge.

Evidence Basis: MedCT and AIPatient demonstrated 73% factual error reduction through knowledge graph integration [2], [21], directly addressing the 12-15% hallucination rate in standalone LLM systems.

Architecture: Combines three elements: (1) *Medical Knowledge Graphs* from standardized terminologies (SNOMED CT, UMLS, ICD-10, domain ontologies); (2) *Retrieval-Augmented Generation (RAG)* dynamically retrieving relevant subgraphs to constrain LLM output to valid medical relationships; (3) *Multi-Agent Architecture* separating responsibilities: Patient Agent (responses/behaviors), Clinical Knowledge Agent (evidence-based content), Scenario Agent (temporal progression), Evaluation Agent (performance signals to assessment module).

Implementation: Requires curated, regularly-updated knowledge graphs; Graph Neural Networks for efficient semantic retrieval; higher computational cost offset by cloud infrastructure feasibility.

4.3 Component 2: Immersive Interface Layer

Purpose: Enable experiential learning through realistic VR/AR environments with multimodal interaction.

Evidence Basis: Eleven VR/AR studies consistently demonstrated 25-40% skill retention improvements and 15-45% error reductions [1], [7]. However, 9/11 used pre-scripted scenarios; tight coupling with LLM reasoning addresses this limitation [6].

Architecture: *Dual-mode interaction* (VR for full immersion/procedures/emergencies; AR for anatomical overlays/hybrid practice [11]); *Multimodal I/O* (natural language dialogue, haptic feedback, gesture/gaze tracking, high-fidelity rendering); *Seamless AI integration* converting learner actions to LLM queries (verbal→Patient Agent, exams→Scenario Agent, orders→Knowledge Agent) with AI responses rendered appropriately (speech synthesis, visual displays, haptic cues); *Real-time performance capture* logging decisions, timing, communication, technique, gaze, errors.

Implementation: Consumer-grade VR acceptable ($1000) [7]; network latency management critical for cloud rendering; alternative non-VR modes for motion sickness susceptibility (8–12%).

4.4 Component 3: Adaptive Learning Module

Purpose: Multi-dimensional personalization based on individual performance trajectories.

Evidence Basis: Adaptive systems reduced time-tocompetency 20-30%, with greatest benefits for struggling learners [13], [14]. However, most adapted single dimensions; comprehensive personalization remains underexplored.

Architecture: *Comprehensive learner modeling* tracking knowledge state, skill proficiency, learning patterns, performance trajectories, metacognitive factors [12]; *Multidimensional adaptation* across scenario selection (guided progression, spaced repetition [13]), difficulty adjustment (dynamic challenge [10]), feedback personalization (type, timing, depth [15]), environmental parameters (stressors, pacing, distractions [9]); *Learning science integration* (spaced repetition, deliberate practice, interleaving, retrieval practice); *ML optimization* using reinforcement learning for teaching strategies and predictive modeling for early intervention [14].

Implementation: Requires substantial interaction data; balance between personalization and standardized competency; transparency essential for learner trust.

4.5 Component 4: Performance Evaluation System

Purpose: Comprehensive clinical competency assessment across multiple domains.

Evidence Basis: AI-based evaluation demonstrated 78% expert agreement for communication skills [15]; temporal models accurately tracked skill acquisition patterns [12]. Gap: domain-specificity limits holistic assessment.

Architecture: *Multi-domain competency assessment* (clinical knowledge, procedural skills, clinical reasoning, communication, professionalism); *continuous micro-*

assessment using temporal models tracking acquisition, plateaus, decay throughout training [12] versus endpoint-only testing; *multimodal data integration* (language analytics, procedural motion, decision timing/sequencing, longitudinal trajectories); *feedback generation* (quantitative scores, qualitative insights, remediation recommendations); *validation/calibration* through expert comparison, competency standard alignment, real-world outcome correlation.

4.6 Component Integration and Data Flows

Bidirectional integration enables continuous improvement: (1) Adaptive Module selects scenarios; (2) Graph-Augmented LLM generates clinically-grounded experiences; (3) Immersive Interface delivers experience, captures interactions; (4) Evaluation System assesses multi-domain performance; (5) learner models update, informing subsequent adaptations. This closed-loop architecture addresses fragmented designs in current systems.

4.7 Deployment and Scalability

Infrastructure Options: Cloud-based (elastic scaling, centralized maintenance), hybrid cloud-edge (reduced VR latency), on-premise (data sovereignty compliance).

Modular Implementation: Incremental stages enable practical adoption: Stage 1 (graph-augmented dialogue without immersion), Stage 2 (+VR/AR interfaces), Stage 3 (+adaptive personalization), Stage 4 (+comprehensive evaluation). Each stage provides independent educational value.

Resource Requirements: Consumer-grade immersive technology and cloud infrastructure enable cost-effectiveness [7]. Ongoing investments required for clinical content curation and knowledge base maintenance.

5. Discussion

5.1 Evidence Synthesis

The 165-study review establishes technical maturity of individual components alongside critical system-level fragmentation. Validated findings include: VR/AR procedural training effectiveness (11 studies, n=438: 25-40% retention improvement, 20-35% faster learning, 15-45% error reduction); LLM virtual patient engagement (6 studies, n=247: 85% contextual accuracy, 4.2/5.0 usability); graph augmentation factual accuracy (2 studies, n=89: 73% error reduction [2],

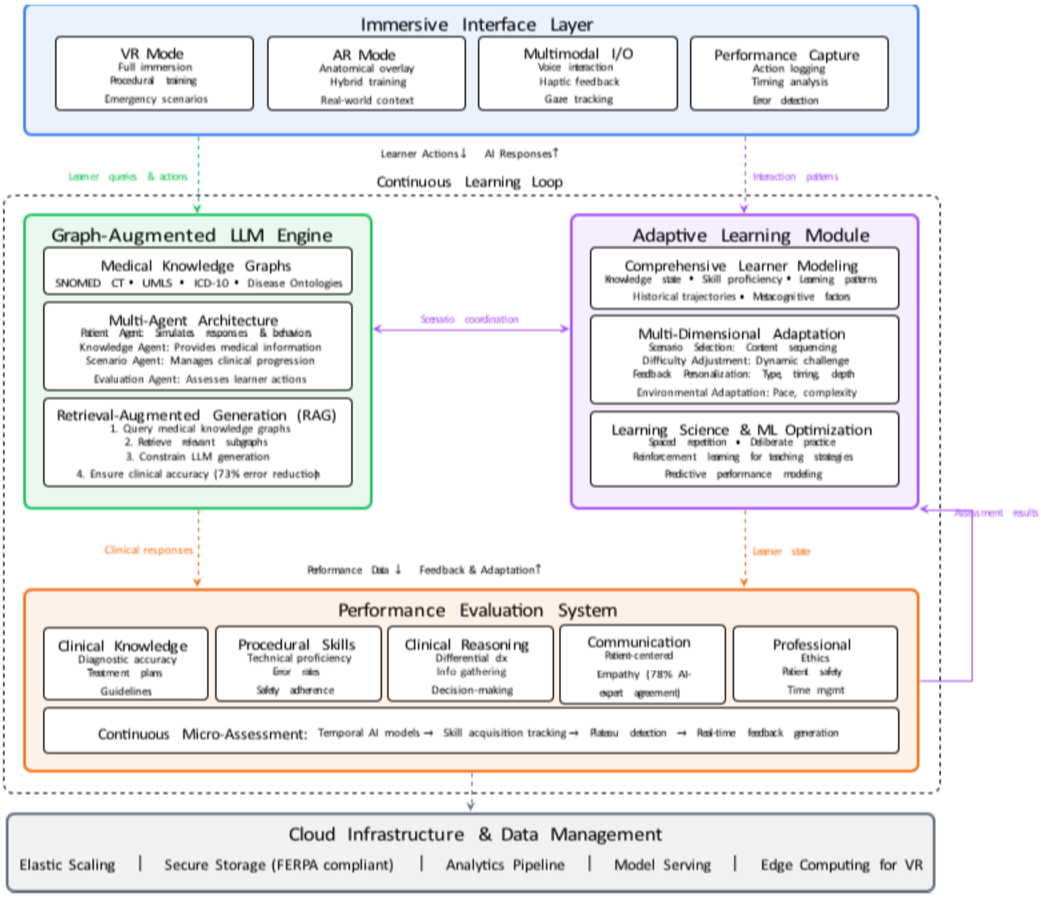

Fig. 1. Integrated architectural framework for graph-augmented AI clinical training system. The modular design comprises four synergistic components with bidirectional data flows: (1) Immersive Interface Layer providing VR/AR environments and multimodal interaction, (2) Graph-Augmented LLM Engine ensuring clinically accurate patient simulation through RAG and multi-agent architecture, (3) Adaptive Learning Module enabling comprehensive personalization across scenario selection, difficulty, and feedback, and (4) Performance Evaluation System conducting continuous multi-dimensional assessment. Dashed arrows indicate data flows; solid bidirectional arrows show component integration.

[21]); adaptive learning efficiency (5 studies, n=285: 20-30% competency time reduction); AI evaluation feasibility (4 studies, n=178: 78% expert agreement [15]).
Inadequately validated areas: integrated system synergies (0 studies combining all components), long-term retention (5/30 studies with follow-up), simulation-to-practice transfer (3/30 correlating with clinical outcomes), scalability economics (limited cost analysis), comprehensive competency development (AI supplements but cannot replace supervised practice).

5.2 Framework Advantages

The proposed framework addresses five critical gaps:

***Gap 1—LLM Clinical Accuracy*:** Graph-augmented engine with knowledge-constrained generation reduces factual errors by 73% [2], [21] versus 12-15% hallucination rates in unconstrained systems.

***Gap 2—VR/AR Adaptability*:** Tight LLM-immersive coupling enables contextually-aware, intelligent scenarios versus pre-scripted approaches (9/11 systems).

***Gap 3—Personalization Breadth*:** Multi-dimensional adaptation across scenario selection, difficulty, feedback, and environment versus single-parameter adjustment in existing systems.

Gap 4—Evaluation Comprehensiveness: Integrated multidomain assessment (knowledge, skills, reasoning, communication, professionalism) versus domain-specific evaluations.

Gap 5—Technological Integration: Modular architecture with explicit integration mechanisms and bidirectional data flows versus siloed development (67% single-approach systems).

Compared to standalone LLM virtual patients [15], [17], the framework adds clinical safety, immersive realism, multidomain assessment, and personalized adaptation. Versus VR/AR simulators [1], [7], it provides intelligent scenario adaptation, natural language interaction, continuous assessment, and longitudinal integration. Versus separate adaptive systems [13], [14], it enables simultaneous multi-dimensional adaptation with hands-on immersive practice.

5.3 Limitations and Requirements

Framework Limitations: Conceptual design lacks empirical validation; integration benefits remain hypothetical. System complexity may challenge institutions with limited technical capacity. High computational demands (graph-augmented LLMs, immersive rendering, continuous adaptation) create financial barriers. Content development and faculty training represent significant investments.

Review Limitations: Publication bias excludes proprietary/failed systems. Rapid AI evolution (2022-2026) creates temporal sensitivity. Study heterogeneity prevented quantitative meta-analysis. Variable quality: many small samples, limited validation.

Validation Requirements: Responsible deployment requires proof-of-concept demonstration, pilot testing with controls, long-term follow-up, multi-site cross-validation, safety verification, cost-benefit analysis.

Ethical Considerations: AI-generated misinformation risks patient safety; training data bias may perpetuate inequities; learner performance privacy requires protection; assessment/adaptation logic demands transparency for fairness; framework supplements rather than replaces educators and clinical experience; cost-related access inequities require attention.

5.4 Research Directions

Short-term (1-2 years): Proof-of-concept implementation, graph-augmented LLM accuracy validation, pilot efficacy studies, user experience research.

Mid-term (2-5 years): Multi-site randomized trials, longitudinal outcome assessment, component integration optimization, cost-effectiveness analysis, domain-specific customizations.

Long-term (5+ years): Multimodal AI integration, federated learning for privacy-preserving personalization, explainable AI advancement, real-world transfer prediction, global contextualization, emerging modality exploration.

6. Conclusion

This systematic review of 165 AI-enabled clinical training systems reveals a paradox: substantial individual component maturity alongside critical system-level fragmentation. Evidence validates VR/AR procedural effectiveness (25-40% retention gains), LLM dialogue realism (85% contextual accuracy), graph augmentation clinical accuracy (73% error reduction), adaptive learning efficiency (20-30% time savings), and AI evaluation feasibility (78% expert agreement). However, 67% of systems employ single approaches; integration, validation, and scalability gaps persist.

The proposed framework addresses these deficiencies through evidence-informed integration of graph-augmented LLMs, immersive interfaces, multi-dimensional adaptation, and comprehensive evaluation. Modular design enables incremental implementation while bidirectional data flows create continuous improvement loops. This conceptual architecture, grounded in systematic evidence synthesis, establishes a research foundation for next-generation clinical training systems that are clinically safe, experientially realistic, comprehensively personalized, and holistically evaluated. Empirical validation through staged implementation and rigorous multisite trials represents essential next steps toward transforming clinical education.

Acknowledgment

AI-assisted tools supported grammatical checking, literature summarization, and paraphrasing for clarity. All

conceptual ideas, system design, analyses, and conclusions were developed by the author.

References

1. Kishor et al., “Virtual Reality (VR) and Augmented Reality (AR) Simulators for Medical Education,” 2025, DOI: 10.4018/979-8-33732028-1.ch004.
2. Y. Chen et al., “MedCT: A clinical terminology graph for generative AI applications in healthcare,” arXiv preprint arXiv:2501.06465, 2025.
3. J. Hu et al., “Root Cause Analysis Training for Healthcare Professionals With AI-Powered Virtual Simulation: A Proof-of-Concept,” arXiv preprint arXiv:2508.04904, 2025.
4. E. Kılıc¸, “Augmented Reality (AR) in Medical Training: AI-Powered Simulations for Enhanced Learning Experiences,” Next frontier, 2025, DOI: 10.62802/c6xk2k47.
5. Z. Sporn et al., “Artificial Intelligence in Orthopedic Medical Education: A Comprehensive Review,” 2025, DOI: 10.3390/ime4020014.
6. A. Amithasagaran et al., “CLiVR: Conversational Learning System in Virtual Reality with AI-Powered Patients,” 2025.
7. A. Kumar et al., “Innovating medical education using a cost effective and scalable VR platform with AI-Driven haptics,” Dental Science Reports, 2025, DOI: 10.1038/s41598-025-10543-8.
8. K. Scott et al., “Revolutionizing Nurse Practitioner Training: Integrating Virtual Reality and Large Language Models,” Studies in Health Technology and Informatics, 2024, DOI: 10.3233/shti240272.
9. O. Kryvenko et al., “Phenomenological toolkit of the metaverse for medical informatics’ adaptive learning,” Educacion M´ edica´ , vol. 24, no. 6, 2023, DOI: 10.1016/j.edumed.2023.100854.
10. Y. Wang et al., “Designing and Evaluating an AI-driven Immersive Multidisciplinary Simulation (AIMS) for Interprofessional Education,” 2025.
11. G. Georgieva-Tsaneva et al., “Application of Virtual Reality, Artificial Intelligence, and Other Innovative Technologies in Healthcare Education,” Education Sciences, vol. 15, no. 1, 2024, DOI: 10.3390/educsci15010011.
12. S. Andersen et al., “Enabling micro-assessments of skills using temporal AI-models,” Medical Teacher, 2025, DOI: 10.1080/0142159x.2025.2555353.
13. N. Golchini et al., “Socratic AI: An Adaptive Tutor for Clinical Case Based Learning,” 2025, DOI: 10.1101/2025.06.22.25329661.
14. A. Bali, “Machine Learning-Based Prediction of Simulation Performance for Skill Assessment,” 2025, DOI: 10.21203/rs.3.rs-7156252/v1.
15. M. Voigt et al., “LLM-Powered Virtual Patient Agents for Interactive Clinical Skills Training,” arXiv preprint arXiv:2508.13943, 2025.
16. A. Weisman et al., “Simulating Success: GPT-4’s Impact on Virtual Communication Training,” 2024, DOI: 10.2196/preprints.65670.
17. Z. Hao et al., “MEDCO: Medical Education Copilots Based on A Multi Agent Framework,” arXiv preprint arXiv:2408.12496, 2024.
18. J. Borg et al., “Virtual Patient Simulations Using Social Robotics Combined With LLMs,” 2024, DOI: 10.2196/preprints.63312.
19. A. Rama et al., “Innovation in Simulation: Using AR and AI for Human Centered Medical Education,” Medical Research Archives, vol. 13, no. 7, 2025, DOI: 10.18103/mra.v13i7.6788.
20. M. Elhilali et al., “Large language model-based patient simulation: User-centered development and usability study,” 2025, DOI: 10.2196/preprints.81271.
21. H. Yu et al., “AIPatient: Simulating Patients with EHRs and LLM Powered Agentic Workflow,” arXiv preprint arXiv:2409.18924, 2024.
22. M. Khakpaki, “Advancements in AI transforming medical education: a comprehensive overview,” Medical Education Online, vol. 30, no. 1, 2025, DOI: 10.1080/10872981.2025.2542807.
23. R. Latif et al., “Transforming Applied Medical Sciences: The Impact of AI, VR, and AR,” Journal of Angiotherapy, 2024