Medical Research
Volume 7, Issue 4, 2025, Page 14-26
https://doi.org/10.6913/mrhk.070403

Medical Research

From Simulated Interaction to Intelligent Generation: A Generative Multi-Agent Virtual Patient Framework for Nursing Education with RAG-Based Safety Guardrails

Haiying SUI
Weifang University of Science and Technology, Email: 1181598128@qq.com

Abstract

In response to the growing mismatch between nursing workforce demand and constrained clinical teaching resources, this study proposes a Generative Multi-Agent Virtual Patient (GMVP) framework for high-fidelity nursing education. Grounded in situated learning, cognitive apprenticeship, and distributed cognition, GMVP employs a triadic agent architecture comprising narrative, physiological, and evaluator agents to reconstruct social interaction, physiological coherence, and formative assessment in virtual clinical environments. A design-based research methodology guides iterative development and classroom deployment aligned with outcome-based education standards. To address hallucination risks in high-stakes medical content, the system integrates retrieval-augmented generation with modular validation and physiological consistency checks. The framework supports scalable case generation, learning analytics, and equitable access to complex clinical training scenarios.

Keywords Nursing education; Generative multi-agent virtual patient; Triadic agent architecture; Retrieval-augmented generation; Design-based research

1 Introduction: Paradigm Reconstruction from "Simulated Interaction" to "Intelligent Generation"

1.1 The Practical Tensions of Nursing Education and the Institutional Logic of Digital Transformation

Examining the current state of nursing education in China and globally at the temporal juncture of 2026, the educational community is confronted with a sharp and pressing structural contradiction. On the one hand, the accelerating aging of society and the growing prevalence of chronic diseases have led to an explosive increase in demand from healthcare systems for high-quality nursing professionals equipped with advanced clinical decision-making capabilities. On the other hand, traditional clinical teaching resources are becoming increasingly depleted, while practical training models remain relatively lagging behind. According to national policy documents and strategic plans issued by relevant authorities, enhancing the level of medical degree education and strengthening the quality of clinical practice teaching have become

 Article History Received: August 5, 2025; Accpted: October 15, 2025; Published: December 15, 2025. *Medical Research*, ISSN 2664-0333 (print), ISSN 2664-0341 (online), a Quarterly, founded on 2019, published by Creative Publishing Co., Limited. Email: cpcl@cpcl.hk, https://mrhk.cc, https://cpcl.hk.

urgent tasks at the level of national strategy. In particular, as the long-term effects of population policies unfold and population aging deepens, the shortage of nursing human resources is manifested not only in quantitative terms but, more importantly, in qualitative dimensions. This includes higher expectations for nurses' abilities to manage complex nurse patient relationships, respond to emergency situations, and care for patients with multiple comorbidities.

For a long time, Standardized Patients (SPs) have been regarded by the global medical education community as the "gold standard" for clinical communication training and skills assessment. SPs refer to healthy individuals or patients who have undergone rigorous training and are able to consistently and realistically simulate patient symptoms, physical signs, and emotional responses. However, the large-scale adoption of the SP model has encountered substantial economic and managerial bottlenecks. Health economics studies indicate that the incremental cost-effectiveness ratio of SP-based assessments is significantly higher than that of traditional teaching approaches, and high-quality SP resources are extremely scarce, making it difficult to provide coverage for all nursing institutions, particularly application-oriented undergraduate colleges located outside major urban centers. Moreover, due to ethical and safety considerations, real clinical internships often restrict students' operational participation in critical care, infectious disease, or end-of-life nursing scenarios. As a result, students lack opportunities for "trial and error" when confronting extreme clinical situations. Although traditional mechanical simulators address issues of operational safety, their lack of emotional feedback and verbal interaction capabilities makes it difficult for students to cultivate deep empathy and sophisticated clinical reasoning when facing cold, inanimate models.

Early Virtual Patient (VP) technologies attempted to fill this gap through screen-based interaction. However, until around 2023, most mainstream VP systems were still built upon predefined decision trees or finite state machines. These "first-generation" virtual patients lacked genuine natural language understanding capabilities; their dialogues were limited to selecting from preset options, and they were unable to exhibit the emotional depth and dynamic physiological logic characteristic of real humans. When interacting with such systems, students often experienced a sense of rigidity and detachment, making it difficult to achieve immersion, which in turn substantially constrained the effectiveness of contextualized learning.

1.2 Limitations of Generative AI and the Intervention of Multi-Agent Systems

The rapid rise of Generative Artificial Intelligence, particularly large language models such as GPT-4 and Claude 3, has for the first time endowed computers with the ability to generate coherent, contextually relevant, and emotionally nuanced natural language. This development has opened new possibilities for medical simulation education. However, directly applying a single large language model to rigorous medical education scenarios entails inherent, and potentially critical, limitations, which are mainly reflected in three aspects.

The first is the risk of hallucination. Large language models are essentially statistical systems that predict the next token based on probability, which means that when dealing with serious medical issues such as drug dosages, pathophysiological mechanisms, or clinical guidelines, they may generate information that appears fluent but is in fact incorrect. In nursing education, such factual errors can not only mislead students but may also pose latent safety risks in future clinical practice.

The second limitation concerns the lack of role consistency. In long-term nursing processes, such as those spanning admission assessment, preoperative education, and discharge guidance, a single model often struggles to maintain consistency in patient personality traits and medical history. It may "forget" information from previous interactions or exhibit abrupt changes in character. This discontinuity severely undermines the coherence of scenario-based learning.

The third limitation lies in physiological disconnection. While large language models excel at processing text and semantics, they are not capable of independently simulating the complex physiological parameters of the human body, such as hemodynamic changes or acid base balance, in response to medical interventions. For example, after a student administers a diuretic, a single language model may be unable to accurately calculate the quantitative relationship between increased urine output and decreased blood pressure, thereby failing to support advanced clinical reasoning training.

It is against this backdrop that Multi-Agent Systems (MAS) have emerged and rapidly become a frontier topic in medical education technology. Unlike a single "all-purpose" model, MAS decomposes complex clinical simulation tasks into multiple specialized agent modules, each focusing on a specific function. These agents collaborate through standardized interfaces, collectively giving rise to complex system behaviors. On this basis, this paper formally proposes the concept of the "Generative Multi-Agent Virtual Patient (GMVP)," defining it as an intelligent system built upon a multi-agent architecture that integrates narrative generation, physiological computation, and instructional assessment. Such a system is capable of autonomously constructing clinical cases with high levels of social realism and physiological fidelity within virtual environments, while providing learners with personalized feedback. The introduction of this concept marks a paradigm shift in virtual patient technology, from script-based "simulated interaction" toward autonomously emergent "intelligent generation."

2 Theoretical Foundations: From Situated Learning to Cognitive Apprenticeship

This study is not a mere technological application; rather, it is deeply grounded in modern educational psychology. The design of the GMVP is not limited to simulating clinical scenarios but constitutes a reconstruction of the learning process itself. Its theoretical core primarily aligns with Situated Learning theory, Cognitive Apprenticeship, and Distributed Cognition.

2.1 Situated Learning and "Legitimate Peripheral Participation"

Situated Learning theory, proposed by Lave and Wenger, posits that learning is fundamentally a process of social participation and that knowledge is inherently contextualized, inseparable from the situations in which it is applied. In traditional classroom-based instruction, nursing knowledge is often abstracted and decontextualized, making it difficult for students to transfer theoretical understanding to complex clinical environments. The concept of Legitimate Peripheral Participation (LPP) introduced by Lave and Wenger emphasizes that novices begin by engaging in simple tasks at the periphery of a Community of Practice and gradually move toward the center as they acquire more complex skills.

Through the Narrative Agent, the GMVP constructs a high-fidelity virtual Community of Practice. Within this community, the virtual patient is not merely an object awaiting clinical operations but a "whole person" endowed with a social identity, such as a retired teacher or an anxious mother, along with family relationships and a broader social background. By generating complex social interactions, including family members' questioning or patients' fear, the narrative agent compels students to engage in legitimate peripheral participation within a virtual environment. Although the setting is safe and simulated, students are required to confront authentic social pressures and interpersonal conflicts. This design transforms learning from simple knowledge acquisition into a process of professional identity formation. Through repeated interactions with GMVPs that possess coherent memory and emotional continuity, students gradually shift from passive observers to professional nursing practitioners capable of independently managing complex nurse patient relationships.

2.2 Cognitive Apprenticeship and the Externalization of Tacit Knowledge

The Cognitive Apprenticeship model proposed by Collins and colleagues aims to teach cognitive skills by emulating traditional apprenticeship methods. Its core strategies include modeling, coaching, scaffolding, and reflection. In clinical nursing education, experts' cognitive processes, such as clinical reasoning and risk anticipation, are often tacit and therefore difficult for novices to observe directly.

Within the GMVP architecture, the Evaluator Agent functions as an "invisible mentor," effectively operationalizing the key components of cognitive apprenticeship. In terms of scaffolding, when students encounter impasses during patient interviews or deviate from appropriate clinical pathways, the system does not provide direct answers. Instead, the evaluator agent offers dynamic prompts, such as suggesting that the patient appears confused by a previous question and encouraging the student to adjust their communication approach. This form of support underpins the learner' s cognitive process and is gradually withdrawn as competence increases, in accordance with the principle of fading and the theory of the Zone of Proximal Development.

Reflection is facilitated through post-training analytic reports generated by the system. These reports go beyond procedural scores to reconstruct the student' s clinical reasoning path and compare it with that of an expert. This form of generative debriefing compels students to externalize their own thinking processes, enabling them to identify where critical cues were overlooked or incorrect judgments were made. In doing so, tacit knowledge is rendered explicit, fostering deeper cognitive development.

2.3 Distributed Cognition and Human–Machine Collaboration

Hutchins' theory of Distributed Cognition argues that cognitive processes are not confined to the human brain but are distributed across individuals, artifacts, and the environment. In contemporary clinical settings, nurses' decision-making is highly dependent on external tools such as patient monitors, electronic medical records, and laboratory reports.

Within the GMVP system, the Physiological Agent does more than simulate the patient' s body; it functions as a critical node within a distributed cognitive system. The real-time vital sign data it calculates and displays, including heart rate and oxygen saturation, serve as essential external representations that enable cognitive offloading. Students must learn to interpret these dynamically generated data and integrate them with the verbal and social information provided by the narrative agent. This deep integration of human, machine, and environment authentically reproduces the cognitive ecology of real clinical practice. The MAS architecture itself constitutes a technological instantiation of Distributed Cognition, wherein intelligence is distributed across multiple agents, and pedagogical effectiveness emerges from the interaction network linking these agents with the learner.

3 Conceptual Model: The Triadic Agent Architecture

To fully reproduce the multidimensional structure of clinical competence—cognition, skills, and affect—within a virtual environment, this study constructs a core *Triadic Agent Architecture*. Inspired by generative agent research at Stanford University and biomedical multi-agent simulation frameworks, this architecture consists of three core agents: the Narrative Agent, the Physiological Agent, and the Evaluator Agent. These agents collaborate through a shared blackboard system or message bus, enabling real-time data exchange and coordinated operation.

Within this architecture, the Narrative Agent functions as the "soul" of the virtual patient, responsible for all non-physiological social interactions, language generation, and emotional expression. Its core mechanisms include memory streams, role-playing, and intention recognition, and its primary educational value lies in providing a rich social context for learning and in cultivating communication and empathy skills.

The Physiological Agent serves as the "body" of the virtual patient. It is responsible for the real-time computation of vital signs and biochemical indicators based on formal medical rules. Its key mechanisms include differential equation solvers, pharmacokinetic and pharmacodynamic models, and mind–body coupling interfaces. Pedagogically, it provides objective clinical indicators and supports the training of clinical observation and intervention abilities.

The Evaluator Agent plays the role of an invisible "mentor." It continuously monitors interaction data and delivers feedback and assessment in real time. Its mechanisms include SBAR pattern matching, clinical reasoning path tracking, and knowledge graph validation. Its educational function is to provide scaffolding support and formative assessment, thereby promoting reflection and learning consolidation.

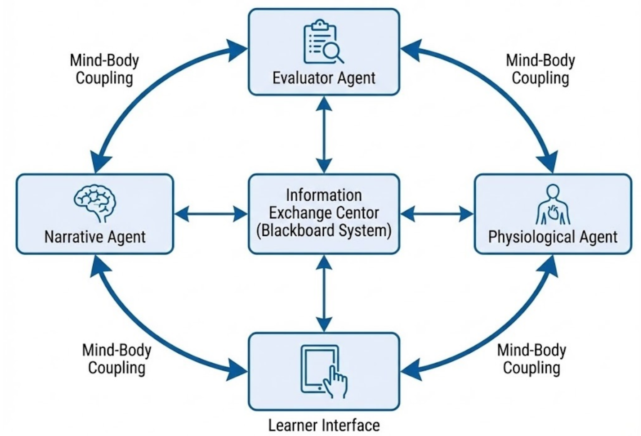

Figure 1: Conceptual Model of the Generative Multi-Agent Virtual Patient (GMVP) Ecosystem *This figure displays the core components of the GMVP system and their interaction relationships. Located in the center is the shared Information Exchange Center (Blackboard System). On the left is the "Narrative Agent" (Soul), responsible for processing social background and emotional memory; on the right is the "Physiological Agent" (Body), which calculates vital signs based on a rule engine. The two achieve "Mind-Body Coupling" via bidirectional arrows. Above is the "Evaluator Agent" (Invisible Tutor), which monitors interactions in real-time and generates feedback. Below, the learner (nursing student) interacts with the system through a multimodal interface.*

3.1 The Narrative Agent: Constructing Social Attributes and the Memory Stream

The Narrative Agent represents the "soul" of the virtual patient and is responsible for all non-physiological social interactions.

3.1.1 The Memory Stream Mechanism

Unlike traditional stateless chatbots, the Narrative Agent possesses an independent memory stream. This mechanism allows the agent to record observed events, assign importance scores to them, and retrieve relevant memories in subsequent interactions to guide behavior.

At the level of short-term memory, the agent recognizes immediate conversational context. If a student asks about allergies and repeats the same question shortly afterward, the Narrative Agent may display impatience, indicating that the question has already been answered. Such responses train students' listening and documentation skills.

At the level of long-term emotional memory, the agent accumulates affective experiences across time. If a student behaves coldly during the initial admission assessment—by avoiding eye contact or interrupting the patient—the agent encodes this as a negative emotional memory. When the same student later engages in preoperative education, the agent may retrieve this memory and respond with resistance, hostility, or refusal to cooperate. This cross-temporal causal feedback effectively simulates the real process of nurse–patient relationship formation and compels students to value the quality of every interaction.

3.1.2 Simulation of the Social Ecology

Leveraging the generative capacity of large language models, the Narrative Agent can expand a complete life history from limited seed information, such as age, occupation, living situation, and personality traits. More importantly, the multi-agent architecture allows multiple role agents—such as anxious family members or impatient attending physicians—to be instantiated simultaneously. This creates a miniature clinical social ecology within the virtual environment. Students must therefore manage not only patient communication but also family concerns and physician directives, which is essential for developing coordination and communication skills in complex interpersonal settings.

3.2 The Physiological Agent: A Rule-Based Life Engine

If the Narrative Agent constitutes the "soul," the Physiological Agent forms the "body" of the virtual patient.

3.2.1 Real-Time Physiological Computation

The Physiological Agent does not rely on language models to generate numerical values. Instead, it performs simulations based on rigorous physiological models, such as systems of differential equations or organ-agent frameworks. It computes vital signs, including heart rate, blood pressure, oxygen saturation, respiratory rate, and body temperature, as well as biochemical indicators such as blood gas values and electrolyte levels. This rule-based computation ensures strict medical logic and prevents physiological hallucinations. For example, when given the instruction to administer intravenous furosemide, the agent calculates specific urine output curves and electrolyte changes based on pharmacokinetic principles rather than producing arbitrary values.

3.2.2 The Mind–Body Coupling Feedback Loop

A key innovation of the architecture lies in the bidirectional, real-time coupling between the Narrative Agent and the Physiological Agent. This mechanism simulates neuroendocrine regulation in the human body.

When emotional states influence physiology, such as when the Narrative Agent experiences anxiety or fear due to a student' s communication, it sends a signal indicating sympathetic nervous system activation to the Physiological Agent. The Physiological Agent then adjusts physiological parameters accordingly, such as increasing heart rate or myocardial oxygen consumption, and may even trigger pathological events in susceptible patients.

Conversely, when physiological changes influence narrative expression, such as in cases of incorrect medication administration leading to hypoxia, the Physiological Agent computes the resulting physiological state and transmits it to the Narrative Agent. The Narrative Agent then modifies its language output to reflect distress, breathlessness, or fear, thereby signaling a clinical crisis to the learner.

3.3 The Evaluator Agent: The Invisible Mentor and Generative Debriefing

The Evaluator Agent functions as an all-seeing, invisible mentor that continuously monitors all interaction data between the student and the virtual patient, including dialogue text, operation logs, and response times.

3.3.1 Automated Assessment of the SBAR Communication Model

In scenarios such as nursing handovers or clinical reporting, the Evaluator Agent applies natural language processing techniques to assess whether student communication conforms to the SBAR framework. It evaluates clarity of situation description, completeness of background information, data-supported assessment, and appropriateness of recommendations, providing timely and structured feedback.

3.3.2 Clinical Reasoning Path Tracking

Beyond communication analysis, the Evaluator Agent reconstructs students' clinical reasoning paths by examining question sequences and action orders. By mapping these patterns, the system identifies unsafe or illogical decisions, flags high-risk behaviors, and highlights them in debriefing reports. This process makes students' reasoning explicit and reinforces patient safety awareness through reflective learning.

4 Research Methodology: Design-Based Research

To ensure that the GMVP model possesses both theoretical depth and practical relevance for addressing real instructional challenges, this study adopts the methodology of *Design-Based Research* (DBR). DBR emphasizes iterative cycles of design, implementation, analysis, and redesign within authentic educational contexts, with the explicit aim of bridging the gap between educational theory and practice. Unlike traditional experimental research, which seeks singular causal verification, this approach focuses on design optimization and theory generation within complex educational systems.

4.1 Research Phases and Iterative Process

Following the typical DBR cycle, this study is structured into three major phases, each aimed at continuously enhancing the instructional effectiveness of the GMVP system through iterative refinement.

4.1.1 Initial Analysis and Design Phase

The first phase focuses on foundational analysis and system design. A needs analysis is conducted based on Situated Learning theory to identify the persistent problem of disconnection between theoretical knowledge and practical application among nursing students prior to clinical internships. On this basis, the triadic architecture of the GMVP system is designed, and the interaction rules and data interfaces among the agents are formally defined. An initial prototype of the GMVP system is then developed, incorporating basic nursing scenarios related to hypertension and diabetes care.

4.1.2 Implementation Phase

In the second phase, the prototype system is deployed in an authentic instructional setting. The pilot implementation is conducted within the undergraduate nursing program at Weifang University of Science and Technology. This institution is selected because, as a representative application-oriented undergraduate college, it faces typical challenges such as large student cohorts and relatively constrained clinical resources, thereby offering a highly representative sample. The GMVP system is introduced as a supplementary practical training tool in the course *Internal Medicine Nursing*. After completing theoretical instruction, students engage in self-directed practice within the system during extracurricular time.

4.1.3 Evaluation and Revision Phase

The third phase centers on systematic evaluation and iterative revision. Multiple sources of data are collected, including students' system usage records, semi-structured interview feedback, and backend logs generated by the Evaluator Agent. These data are analyzed to diagnose system-level issues, such as limitations in the realism of emotional feedback and delays in physiological responses. Based on these findings, iterative optimization is conducted by fine-tuning agent prompts through prompt engineering, refining the physiological rule base, and adding targeted "hard-mode" scenarios. The revised system then enters the next design cycle.

4.2 Data Sources and Analytical Framework

A mixed-methods approach is employed for data collection to provide a comprehensive assessment of the effectiveness and limitations of the GMVP system.

Quantitative data include students' operational scores within the system, longitudinal changes in SBAR communication performance, time required to complete standardized scenarios, and the incidence rate of system-identified high-risk behaviors. These indicators are used to analyze students' skill acquisition trajectories.

Qualitative data are gathered through semi-structured interviews and focus group discussions. These data capture students' subjective experiences of the virtual patient's perceived realism, emotional resonance, and the effectiveness of system feedback. Particular attention is paid to students' psychological responses when confronted with virtual patient behaviors such as refusal or anger, in order to evaluate the development of affective competence.

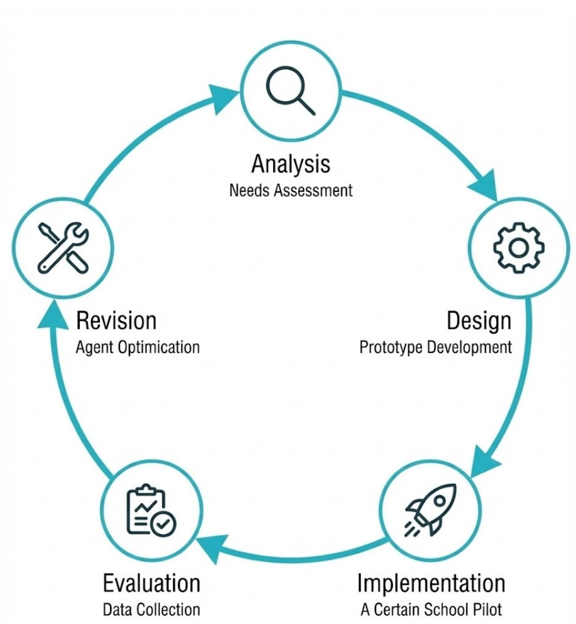

Figure 2: GMVP Iterative Development Process Based on Design-Based Research (DBR) *This figure illustrates the Design-Based Research (DBR) iterative cycle adopted in this study. Through repeated implementation and evaluation in authentic teaching contexts, the behavioral logic and pedagogical strategies of the intelligent agents are continuously optimized to ensure a deep integration of theory and practice.*

5 Localized Embedding in Nursing Education Contexts

The design of the GMVP system is not an abstract technological construct; rather, it is closely embedded within the concrete institutional context of nursing education in China. In particular, it is deeply integrated with the principles of *Outcome-Based Education* (OBE) and national educational standards.

5.1 Institutional Adaptation and the OBE Framework

The *National Standards for the Teaching Quality of Undergraduate Professional Programs in Nursing* issued by the Ministry of Education of China clearly stipulate that nursing education should cultivate students' clinical thinking, humanistic care, and lifelong learning abilities through diversified teaching approaches. At the same time, the OBE framework emphasizes a student-centered orientation, outcome-driven design, and continuous quality improvement. The GMVP system aligns closely with these institutional requirements and pedagogical principles in several key ways.

First, it adopts an outcome-based orientation. The system's design objectives are directly mapped to the core competency indicators specified in the national standards. For example, the interaction logic of the Narrative Agent corresponds to competencies in communication and collaboration, while the emergency simulation functions of the Physiological Agent align with competencies in clinical nursing practice.

Second, the system follows a backward design logic. Starting from the desired clinical competencies that students are expected to achieve, the complexity of disease conditions and the difficulty of interactions are designed in reverse. For instance, to cultivate critical thinking, the system incorporates complex cases characterized by ambiguous symptoms and contradictory signs, compelling students to engage in in-depth differential diagnosis and clinical judgment.

Third, the system supports continuous improvement. The large-scale data collected by the Evaluator Agent are not only used to assess individual student performance but also to identify common weaknesses across an entire cohort, such as frequent omissions in allergy history inquiry. These aggregated insights provide evidence-based support for instructors to adjust curriculum design and teaching strategies.

5.2 An Application Case at a Representative University

As a representative application-oriented undergraduate institution, a certain university demonstrated the strategic value of the GMVP system in addressing resource constraints and promoting educational equity after its adoption.

One notable advantage lies in the creation of an effectively unlimited case repository and the equalization of educational resources. Application-oriented institutions often face challenges such as limited case diversity and a shortage of typical clinical cases due to constraints in affiliated hospital resources. Supported by retrieval-augmented generation technology, the GMVP system can generate thousands of distinct patient profiles based on a single diagnosis, such as hypertension, with customizable attributes including gender, age, occupation, and family background. This capability enables the low-cost dissemination of high-quality teaching resources, allowing students in non-central regions to engage with rare or complex cases that would otherwise be inaccessible.

Another important feature is the simulation of dialects and cultural contexts. In response to China's diverse sociocultural landscape, the system incorporates dialect recognition and localized dietary habit simulation. In instructional practice at the representative university, the system generated virtual elderly patients speaking with regional accents and exhibiting locally prevalent dietary habits, such as high salt intake. During health education tasks, students were required to translate professional medical terminology into accessible, vernacular language and to provide practical recommendations grounded in local lifestyle practices. This context-sensitive training substantially enhanced students' communication competence and cross-cultural nursing abilities in real-world primary healthcare settings.

6 Technical Architecture Analysis: Anti-Hallucination Mechanisms and Data Insights

In the high-risk domain of medical education, ensuring the factual accuracy of AI-generated content is the lifeline of any intelligent system. Any erroneous description of drug dosage, contraindications, or

Table 1: Multidimensional Comparison of GMVP, Traditional SP, and Mannequins: Core Functionality and Cost-Benefit Analysis Matrix

Evaluation Dimension	**Traditional Standardized Patient (SP)**	**Early Mechanical Simulators (Mannequins)**	**Multi-Agent Virtual Patient (GMVP)**
Setup Cost	Medium (Recruitment & Training)	Very High (Expensive Hardware)	**Low (Software Development & Deployment)**
Operational Cost	Very High (Ongoing Wages, Management Fees)	Low (Maintenance Fees)	**Very Low (Compute Cost)**
Accessibility	Low (Appointment needed, Location constrained)	Medium (Lab constrained)	**Very High (24/7, Multi-terminal)**
Case Diversity	Low (Limited by Actor traits)	Low (Pre-programmed)	**Very High (Generative Infinite Case Library)**
Physiological Fidelity	Poor (Cannot simulate real vital sign changes)	Good (But lacks emotion)	**Good (Mind-Body Coupling Simulation)**
Emotional Interaction	**Excellent (Real Human Emotion)**	None	Good (Text & Voice-based Emotion)
Feedback Immediacy	Delayed (Post-session feedback)	Immediate (Data feedback)	**Immediate & Deep (Incl. Reasoning Analysis)**

emergency procedures may lead to serious instructional incidents and long-term safety risks. To address this challenge, the GMVP system introduces a rigorous safeguard architecture based on *Retrieval-Augmented Generation* (RAG) and modular checks and balances, thereby establishing a robust and multi-layered "guardrail mechanism."

6.1 RAG-Based Knowledge Anchoring and Anti-Hallucination Workflow

Hallucination represents the most critical risk of generative AI in medical applications. To systematically eliminate this risk, the GMVP system adopts a RAG-based dynamic verification workflow that tightly constrains free-text generation and anchors all medical outputs to authoritative knowledge sources.

The first step is intent recognition and interception. When a student raises a medically relevant question, such as inquiries about drug side effects, contraindications, or initiates a clinical operation, a front-end "gatekeeper" module performs real-time intent recognition. If the interaction is classified as involving serious medical knowledge, the system temporarily suspends the Narrative Agent' s unrestricted generative capability, preventing speculative or unverified responses at the source.

The second step involves authoritative knowledge retrieval. Once an interaction is flagged, the RAG module is activated to conduct semantic retrieval within a predefined *whitelisted knowledge base*. This knowledge base is strictly curated and limited to authoritative sources, including the Chinese Pharmacopoeia, the most recent national and professional clinical nursing practice guidelines, and peer-reviewed, officially approved textbooks. By restricting retrieval to vetted sources, the system ensures that all downstream generation is grounded in validated medical knowledge.

The third step consists of factual validation and physiological consistency checks. The retrieved authoritative text blocks are passed to a verification module that enforces grounding constraints on the Narrative Agent. The agent is required to generate responses exclusively based on the retrieved content. Furthermore, when a generated response involves specific physiological claims or numerical values, such as statements about dizziness implying elevated blood pressure, the output must be cross-validated by the Physiological Agent. Only if the Physiological Agent' s real-time calculations confirm that the patient' s simulated blood pressure exceeds the clinical threshold is the response permitted to be delivered. This dual-validation mechanism ensures both textual correctness and physiological coherence.

Empirical testing demonstrates that, in domains such as pharmacology and guideline adherence, the RAG-enhanced GMVP system achieves an accuracy rate exceeding 95 percent. This effectively eliminates factual hallucinations and ensures the scientific rigor and instructional safety of the generated content.

6.2 Modular Constraint Design and System-Level Data Insights

Beyond hallucination prevention, the modular architecture of GMVP enables systematic data insight generation for both instructional optimization and educational governance. Each agent operates within clearly defined functional boundaries, while interaction data are continuously logged and structured.

From a system perspective, all student interactions generate high-resolution learning traces, including dialogue sequences, decision timestamps, operational order, and physiological outcome changes. These data are aggregated and anonymized to form a multidimensional learning analytics dataset. By analyzing these data, the system can identify recurring error patterns, such as frequent omissions of allergy history inquiries or delayed emergency interventions, thereby revealing latent weaknesses in cohort-level clinical reasoning.

At the instructional level, these insights support evidence-based curriculum refinement. Instructors can use system-generated dashboards to examine distributions of high-risk behaviors, time-to-intervention metrics, and SBAR communication completeness, enabling targeted pedagogical interventions rather than relying on intuition or anecdotal experience.

At the system governance level, the modular logging and validation framework provides auditability and traceability. Every critical AI-generated medical statement can be traced back to its retrieval source, validation path, and physiological check result. This transparency is essential for regulatory compliance, ethical accountability, and future certification of AI-assisted medical education systems.

7 Conclusion and Future Directions

7.1 Research Conclusions

The integration of Multi-Agent Systems with generative virtual patients marks a paradigm shift in nursing education, moving from standardized simulation toward intelligent ecosystem-based simulation. The GMVP triadic agent architecture proposed in this study successfully reconstructs the complexity, social embeddedness, and physiological logic of clinical nursing practice within a virtual environment.

First, the Narrative Agent, through memory streams and social ecology simulation, addresses the long-standing limitations of traditional simulations that lack social context and emotional continuity, thereby providing a high-fidelity environment for situated learning.

Second, the Physiological Agent, through rule-based engines and mind–body coupling mechanisms, overcomes the physiological reasoning deficiencies of single large language models and enables objective data representation within a distributed cognition framework.

Third, the Evaluator Agent, through automated SBAR assessment and clinical reasoning path tracking, operationalizes the core mechanisms of cognitive apprenticeship, namely scaffolding and reflection, and facilitates the externalization of tacit knowledge.

For application-oriented undergraduate institutions in China, GMVP represents not merely a technological upgrade but a strategic opportunity to enhance educational equity and quality. By enabling high-quality clinical teaching resources to transcend geographical and economic constraints at minimal marginal cost, GMVP offers a viable technological pathway to alleviate the structural mismatch between nursing workforce supply and demand.

7.2 Research Limitations

Despite its demonstrated potential, this study is subject to several limitations.

One limitation concerns the "uncanny valley" effect in affective computing. Although the system performs well in textual and vocal emotional expression, substantial gaps remain in non-verbal communication. Current virtual avatars struggle to render micro-expressions and subtle body language, and overly realistic yet imperfect representations may induce discomfort among learners and disrupt empathy formation.

A second limitation involves technical latency and interaction fluency. The communication among multiple agents, combined with RAG retrieval and physiological computation, requires significant computational resources. Under suboptimal network conditions, response delays may occur, reducing immersion and the natural flow of interaction.

A third limitation relates to ethics and privacy. Training data generated by students include behavioral habits, communication styles, and even personal cognitive weaknesses, all of which constitute highly sensitive personal information. Ensuring robust privacy protection, preventing data leakage, and mitigating algorithmic bias while enabling personalized learning remains a challenge that must be addressed through combined institutional and technical measures.

7.3 Conditional Pathways for Future Advancement

In light of these findings, the deployment of GMVP technology should follow a strategy of conditional advancement rather than unrestrained expansion.

At the technological level, future research should focus on breakthroughs in multimodal models to enhance non-verbal emotional generation and on algorithmic optimization to reduce system latency.

At the institutional level, rigorous educational data ethics standards and algorithm auditing mechanisms should be established to ensure fairness, transparency, and accountability.

At the pedagogical ecosystem level, it is essential to explore a new paradigm of human–AI collaboration. In this ecosystem, AI systems undertake large-scale, standardized foundational training and broad coverage, while human educators are liberated from repetitive assessment tasks and can concentrate on guiding students through complex ethical dilemmas, delivering deep humanistic care, and perceiving subtle emotional cues beyond the reach of AI simulation. Only under the dual safeguards of robust technical guardrails and sustained humanistic commitment can GMVP truly become a powerful engine for enhancing the clinical competence of nursing professionals.

References

[1] World Health Organization. (2020). *State of the world's nursing 2020: Investing in education, jobs and leadership*. World Health Organization.

[2] Ministry of Education of the People's Republic of China. (2020). *Guidance on accelerating the innovative development of medical education.*

[3] Williams, B., & Song, J. J. (2016). Are simulated patients effective in facilitating development of clinical competence for healthcare students? A scoping review. *Advances in Simulation*, *1*(1), 1–9. https://doi.org/10.1186/s41077-016-0006-1

[4] Kononowicz, A. A., Woodham, L. A., Edelbring, S., Stathakarou, N., Davies, D., Saxena, N., ... & Zary, N. (2019). Virtual patient simulations in health professions education: Systematic review and

meta-analysis by the digital health education collaboration. *Journal of Medical Internet Research*, *21*(7), e14676. https://doi.org/10.2196/14676

[5] Thirunavukarasu, A. J., Ting, D. S. J., Elangovan, K., Gutierrez, L., Tan, T. F., & Ting, D. S. W. (2023). Large language models in medicine. *Nature Medicine*, *29*(8), 1930-1940. https://doi.org/10.1038/s41591-023-02448-8

[6] Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., ... & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys*, *55*(12), 1-38. https://doi.org/10.1145/3571730

[7] Park, J. S., O'Brien, J. C., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2023). Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (pp. 1-22). https://doi.org/10.1145/3586183.3606763

[8] Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge University Press.

[9] Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 453-494). Lawrence Erlbaum Associates.

[10] Hutchins, E. (1995). *Cognition in the wild*. MIT press.

[11] Anderson, T., & Shattuck, J. (2012). Design-based research: A decade of progress in education research? *Educational Researcher*, *41*(1), 16-25. https://doi.org/10.3102/0013189X11428813

[12] Spady, W. G. (1994). *Outcome-Based Education: Critical Issues and Answers*. American Association of School Administrators.

[13] Zhang, J., Li, Y., & Chen, J. (2022). The construction and application of an outcome-based education model in nursing education. *Nurse Education Today*, *119*, 105615.

[14] Lee, P., Bubeck, S., & Petro, J. (2023). Benefits, limits, and risks of GPT-4 as an AI aid for medicine. *New England Journal of Medicine*, *388*(13), 1233-1239. https://doi.org/10.1056/NEJMsr2214184

[15] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ...& Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems*, *33*, 9459-9470.

[16] Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine*, *19*(2), 98-100.

[17] De la Prieta, F., Rodríguez-González, S., Mathieu, P., & Bajo, J. (2019). Multi-agent systems for healthcare. *Journal of Medical Systems*, *43*(1), 1-2.

[18] Padilha, J. M., Machado, P. P., Ribeiro, A., Ramos, J., & Costa, P. (2019). Clinical virtual simulation in nursing education: An integrative review. *Nurse Education Today*, *79*, 165-172. https://doi.org/10.1016/j.nedt.2019.05.030

[19] Shorey, S., & Ng, E. D. (2019). The use of virtual reality simulation among nursing students and registered nurses: A systematic review. *Nurse Education Today*, *72*, 243-255.

[20] Liaw, S. Y., Wu, L. T., Chow, Y. L., Lim, S., & Tan, K. K. (2023). Artificial intelligence in nursing education: Opportunities and challenges. *Nurse Education in Practice*, *69*, 103626.

To Cite This Article Haiying SUI.(2025). From Simulated Interaction to Intelligent Generation: A Generative Multi-Agent Virtual Patient Framework for Nursing Education with RAG-Based Safety Guardrails. *Medical Research*, 7(4), 14-26. https://doi.org/10.6913/mrhk.070403

(Editor: Kai ZHENG, Email: cpcl@cpcl.hk)