



AlPatient: Simulating Patients with EHRs and LLM Powered Agentic Workflow





Huizi Yu, MS, Jiayan Zhou, PhD, Lingyao Li, PhD, Themistocles L. Assimes, MD, Danielle S. Bitterman, MD, Xin Ma, PhD⁵, Lizhou Fan, PhD

MOTIVATION & METHOD

- Simulated patient systems play a crucial role in medical training and evaluation.
- Challenges with current simulated patient systems include limited intelligence, lack of diverse patient profiles, and trustworthiness concerns.
- We developed an LLM-powered simulated patient system AlPatient incorporating the AlPatient Knowledge Graph (KG) and Reasoning RAG agentic workflow.

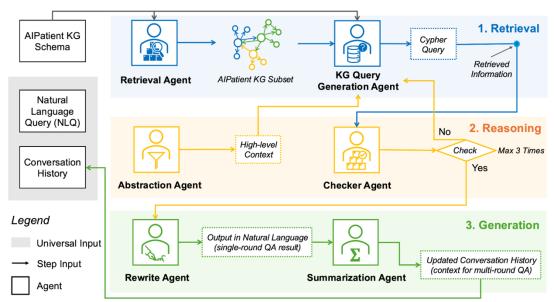


Figure 1. Reasoning RAG agentic workflow is the AlPatient system's processing backbone, comprising three key stages: retrieval, reasoning, and generation.

EVALUATION FRAMEWORK

Performance aspect	Evaluation dimension	Evaluation by	Metrics
Effectiveness	Knowledgebase validity (NER)	Medical doctors	F1
	QA accuracy (conversation)	Researchers	Accuracy
	Readability	Algorithm	Flesch Reading Ease, Flesch-Kincaid Grade Level
Trustworthiness	Robustness (system)	Researchers	Accuracy, ANOVA
	Stability (personality)	Researchers	Accuracy, ANOVA

RESULTS

- Knowledgebase Validity: GPT-4 Turbo achieved the highest F1 score (0.89)
- QA Accuracy: The full agent setup achieved 94.15% accuracy
- Readability: AlPatient responses had a median Flesch Reading Ease of 77.23 and Flesch-Kincaid Grade Level of 5.6, ensuring accessibility.
- Robustness & Stability: No significant accuracy loss due to paraphrased inputs or personality variations, confirming system reliability.

CONCLUSION

- The multi-agent AI framework significantly improves simulated patient realism and intelligence.
- Reasoning RAG enhances accuracy, reliability, and patient interaction fidelity.
- Al-generated patient simulations can support medical education, clinical decision-making, and model evaluation.