International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Gamification And Summarization Framework For Effective Learning For Medical Students

Shrinidhi Deshpande[1], Ravindra Mule[2], Himani Kale[3], Dr. Minakshi Atre[4], Tanvi Limaye[5]

[1,2,3,4,5]Dept. of Artificial Intelligence & Data Science Pune Vidyarthi Griha's College of Engineering, Technology & Management Pune, India

Abstract

Medical education involves vast study of variety of reference books. This necessitates precise information retrieval and excellent retention of extensive, complex concepts. This work is focused to enhance the learning of medical students with AI powered summarization and effective concept understanding with gamification. The authors examine previous research on the integration of AI technologies and gamification techniques in medical learning environment. The article compares and summarizes the AI-driven retrieval methods and game-based engagement strategies, helping to improve knowledge acquisition and retention of complex medical concepts. A thorough literature evaluation involves research on Large Language Models (LLM), medical domain-specific retrieval model structures, and gamified, interactive learning environments. The accuracy, contextual grounding, and adaptability of Retrieval-Augmented Generation (RAG) systems are investigated. The impact of gamification approaches on learner engagement and retention is crucial; hence a comparative analysis is carried out using parameters such as domain customization, retrieval performance, and interactive features. The research showed that RAG enhanced contextual accuracy, reducing irrelevant information in medical learning tools. Gamification enhanced engagement, nevertheless its integration with adaptive retrieval systems was limited in previous research. Across the reviewed work, most of the research focused either on RAG-based retrieval systems or LLM-based content generation with less focus on interactive or gamified learning. It highlighted a gap in designing adaptive, learner-centric medical RAG systems with effective knowledge retention. The review of 28 research paper motivated the authors to build Gamification and Summarization Framework for Effective Learning for Medical Students.

Keywords: Artificial Intelligence, Interactive learning environments, Information Retrieval, language model, gamification, Retrieval-Augmented Generation

1. INTRODUCTION

Medical education is a knowledge-intensive and dynamic sector in which learners must not only acquire huge amounts of theoretical information but also retain and apply it in clinical settings [1]. With the rapid evolution of medical knowledge, textbooks and reference materials are constantly updated, making it difficult for students to remain current [1]. Lectures, reading, and taking notes are examples of traditional learning methods that frequently lack personalization [2]. Furthermore, they usually fall short in terms of successfully engaging students [3]. In recent years, the incorporation of artificial intelligence (AI) into

education has created new opportunities for adaptive and data-driven learning systems. A promising technique for producing accurate and contextually rich replies from massive document collections is RAG architectures [4]. To increase retrieval accuracy and contextual relevance, these designs combine domain-specific information sources with LLMs [5]. Such systems can facilitate accurate and effective information retrieval in the medical field by bridging the gap between static textbook content and dynamic, real-time learning needs. Few studies have thoroughly examined the relationship between RAG systems and gamification in the context of medical pedagogy, despite the increased interest in AI-driven teaching. Studies that are now available frequently concentrate on discrete implementations and lack a unified framework for assessing their technological design, ethical considerations, and pedagogical impact. By combining recent research in these areas, highlighting important design trends, evaluation metrics, and educational implications, this literature survey seeks to close that gap. The shortcomings of traditional learning resources, which usually fall short in sustaining student interest and accommodating individual demands, serve as the impetus for our study. Reviewing recent research on intelligent educational aids and examining cutting-edge methods for document retrieval and gamification are the main contributions of this work. This paper's remaining sections are arranged as follows: A review of the literature on RAG designs, gamification techniques, and current AI-based learning platforms is provided in **Section 2**. The research gaps and research objectives are analyzed in **Section 3**. The results, restrictions, and possible future study directions are covered in **Section 4**. The paper's main conclusions and suggestions for real-world application are provided from **Section 5** onwards**.**

2. LITERATURE SURVEY

There has been a notable increase in the use of cutting-edge technology in medical education and healthcare, with research concentrating on several important areas, such as simulation, gamification, and LLMs. Based on extensive searches in sources such as arXiv and PubMed, a thorough 2025 evaluation emphasized the enormous potential of LLMs in clinical support and medical education [1]. Increasing reliability for medical applications, especially through Retrieval-Augmented Generation (RAG), is a common goal in the development of these models. To guarantee reliable information retrieval for medical specialists, RAG is a fundamental methodology in the development of specialized medical question-answering systems, where it is coupled with advanced rapid engineering [2]. RAG systems are scalable and contextually rich thanks to a two-step approach that indexes medical articles into vector databases and instantly retrieves pertinent data to respond to user inquiries [18]. RAG's performance has been benchmarked and optimized through additional research. For example, the MEDRAG toolset showed that RAG may greatly improve performance on medical benchmarks such as MIRAGE [6]. It has been demonstrated that innovations like Medical Graph RAG outperform conventional RAG and GraphRAG, setting new standards by offering more understandable and fact-based responses [12]. From developing conversation engines for Traditional Chinese Medicine that lessen dependency on external sources [11] to reducing hallucinations in LLMs by comparing RAG-enhanced outputs to human-generated medical templates, where models such as GPT-4 demonstrated high accuracy [10], the usefulness of RAG has been demonstrated in a variety of applications. RAG has been used to increase frontline health professionals' capacity in education, specifically, but a 2023 study found that there is still opportunity for improvement, as only 35% of responses were deemed thorough despite their high accuracy [25]. 47 studies on RAG chatbots in education, mostly from 2024, were found in a larger survey, suggesting that interest in the topic is expanding quickly [3].

Research looks toward adding specific knowledge sources to LLMs in addition to RAG. For instance, by integrating medical textbooks, the LLM-AMT framework enhances medical competency. Accuracy increases range from 11.6% to 16.6% employing a Knowledge Self-Refiner and a hybrid retriever [7]. Similarly, it has been demonstrated that using Clinical Practice Guidelines (CPGs) greatly enhances LLM-based clinical decision assistance [8]. Compared to general models, custom-built models such as EyeGPT, which was created for ophthalmology, use RAG and fine-tuning on data to obtain greater empathy, better performance, and less hallucinations [19]. Another important topic of research is the general usefulness of models such as ChatGPT. On medical competency tests such as the USMLE, GPT-4 has proven to be significantly more capable than its predecessors [21]. ChatGPT has demonstrated efficacy in clinical practice by producing clinical notes, radiological reports, and discharge summaries, facilitating decision-making and enhancing productivity [22]. Although these models make medical knowledge more accessible, they run the danger of weakening critical thinking abilities when used incorrectly [23]. To maximize their advantages in medical research and education while reducing hazards, strict regulations and ethical considerations are also required [24].

The application of gamification to improve clinical reasoning and health education is another well-known trend. Serious games are the most widely used gamification method, according to a 2025 scoping review, with most of the research consisting of pilot or cross-sectional studies conducted in the US [4]. The most utilized features in healthcare apps, such as points (70%), social interactions (55%), leaderboards (40%), and progress status (40%), have been the subject of research aimed at determining the key factors of successful gamification [28]. Although it also identified problems such as over-reliance and regional limitations, the Delphi technique has been used to prioritize important components for medical educators [15]. To boost motivation through psychological components, a 2023 systematic review delineated a research agenda for gamification in healthcare, with an emphasis on chronic diseases, health education, and preventative medicine [26]. While hybrid models employing Q-learning and deep reinforcement learning can mimic educational environments to increase student engagement and task completion rates [14], the incorporation of AI into serious games presents opportunities for flexible game design and improved patient contact [16]. Lastly, simulation is still a vital component of contemporary medical education. It offers a safe, regulated setting where students can practice handling risky or unusual clinical situations without endangering patients [27]. With benefits including ongoing learning access, personalized experiences, and increased cost-effectiveness, AI-driven healthcare simulation is becoming increasingly popular [17].

Table I groups the 28 papers according to their main areas of interest, which are Simulation & Gamification, LLM applications, and RAG, in order to compile the findings from the evaluated literature. The methodological focus, shared restrictions, and quantifiable performance or learning indicators relevant to medical education are examined for each group.

SUMMARY OF MEDICAL EDUCATION STUDIES

Category	Number of papers	Key Focus / Methods	Common Limitations	Observed / Potential Metrics
RAG-focused	9	Retrieval-Augmented Generation, vector	Low personalization, limited interactivity, evaluation gaps	Accuracy, relevance, response time

		DBs, hybrid retrievers		
LLM applications in medical contexts	11	GPT-based QA, summarization, fine-tuning	Hallucinations, inconsistent evaluation	Correctness, coherence, reasoning accuracy
Simulation & Gamification	8	Serious games, reinforcement learning, VR/AR, interactive modules	Limited technical depth, narrow scope, insufficient scalability	Engagement, adaptability, knowledge gain

3. RESEARCH GAPS AND RESEARCH OBJECTIVES

The integration of AI, particularly RAG into medical education has shown a lot of promise. The absence of tailored learning paths, the lack of integration between information retrieval and learner engagement, and the requirement for a standardized framework to assess their combined efficacy are the main causes of these gaps. In order to overcome these limitations, this study suggests and assesses a new, integrated platform. This study will be guided by the following research gaps and the goals that go along with them:

RESEARCH GAPS AND OBJECTIVES

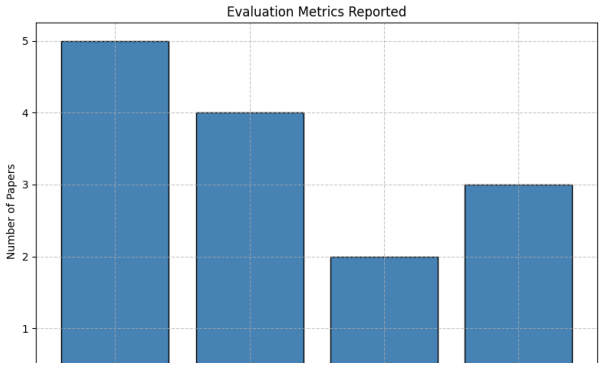
Sr. No.	Research Gaps	Research Objectives
1.	Most existing AI-based medical education tools either focus on information retrieval (RAG/LLMs) or gamification independently, with minimal integration between the two.	To design and evaluate an integrated framework that combines Retrieval-Augmented Generation with gamification elements to deliver both precise, contextually relevant medical information and sustained learner engagement.
2.	Existing RAG systems in medical education often prioritize factual accuracy but provide limited adaptability to individual learner profiles, prior knowledge, or preferred learning styles. As a result, the learning experience remains largely static and does not optimize knowledge delivery for diverse student needs.	To develop and evaluate a personalized Retrieval-Augmented Generation framework that adapts content delivery, difficulty level, and gamification elements based on each learner's progress, performance, and engagement patterns.
3.	There is no standardized set of evaluation metrics for measuring the combined effectiveness of retrieval accuracy and gamified engagement in medical education, making cross-comparison between studies difficult.	To define and implement a standardized evaluation framework that measures both retrieval performance (accuracy, contextual relevance, hallucination rate) and gamification effectiveness (engagement rate, retention improvement) in medical learning platforms.

4. ANALYSIS

The 28 papers are divided into three categories by this analysis: Retrieval-Augmented Generation (RAG), LLM applications in medical contexts, and Simulation & Gamification. Through representative visualizations, the section highlights trends while examining methodological patterns, evaluation techniques, and significant focal areas across various categories.

RAG

Fig. 1. Evaluation Metrics reported in RAG studies



The main objective of RAG studies is to evaluate the effectiveness of retrieval-augmented generation methods.

Figure 1 shows the most commonly utilized metric was Accuracy/F1 score, which was followed by BLEU/ROUGE/METEOR, Recall@K/MMR, and Human Evaluation Scores. According to this pattern, researchers use quantitative, repeatable metrics such as F1-score and Recall@K to evaluate the effectiveness of retrieval-augmented generation techniques, but they also include some qualitative human assessment to capture user-centered or practical performance aspects.

LLMs in Medical

While RAG studies concentrate on enhancing retrieval and response quality, LLM applications in medical contexts prioritize domain-specific tasks and real-world applications.

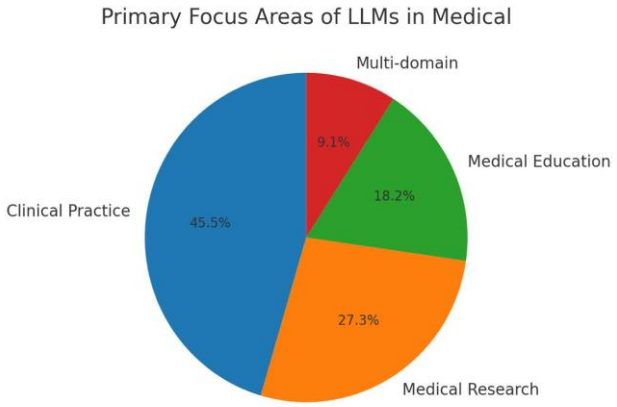


Fig. 2. Primary Focus Areas of LLMs in

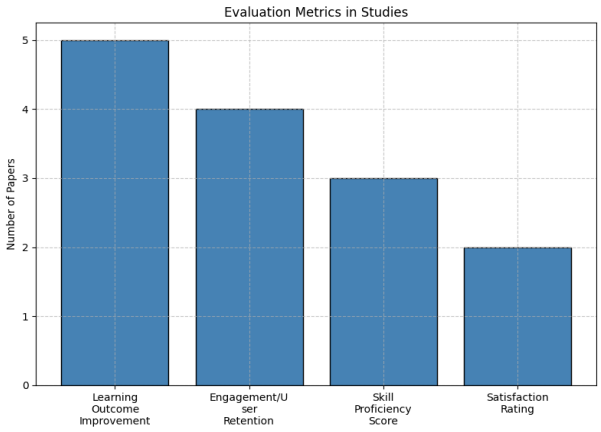


Fig. 4. Evaluation Metrics Reported in Gamification and Simulation studies

Figure 2 depicts the main areas of concentration for investigations using LLMs in the medical field. Clinical practice is represented by the most prominent portion, indicating that most of the research is focused on these applications. Notably, specialized systems like EyeGPT have demonstrated how the range of AI-driven teaching aids can be increased by customizing LLMs for certain therapeutic domains, such as ophthalmology [12]. Medical research, medical education, and multi-domain account for 27.3%, 18.2%, and 9.1% of the total, respectively, indicating that these topics have received moderate attention in the literature.

Gamification and Simulation

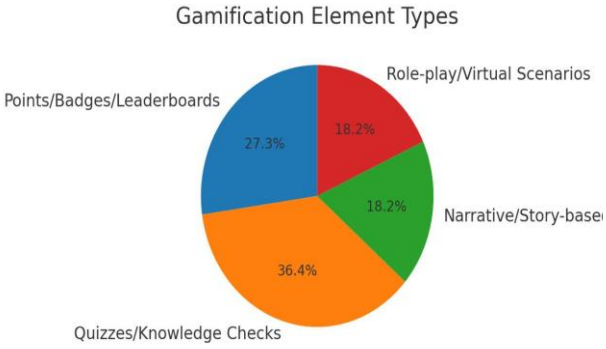


Fig. 3. Gamification elements employed

Gamification and simulation research emphasizes interactive learning experiences and learner outcomes rather than model-centric performance.

Figure 3 shows the kinds of gamification components used in simulation-based research. In contrast to points/badges/leaderboards, narratives/story-based games, and role-play/virtual scenarios, which are less prevalent and reflect a selective use of gamification tactics in training and educational simulations, quizzes/knowledge checks make up the biggest proportion, demonstrating their widespread use. The most

often utilized components were points and quizzes, which demonstrated a focus on engagement and motivation tactics.

Figure 4 represents the evaluation metrics reported in gamification and simulation studies. The greatest improvement in learning outcomes suggests that most research focuses on improving students' knowledge or abilities. User retention and engagement follow, highlighting how crucial it is to maintain participants' active participation. While satisfaction is the least reported metric, indicating that it is viewed as a secondary outcome rather than the main goal, skill proficiency is the third most assessed metric, emphasizing the evaluation of practical capabilities.

Overall, the literature shows increasing complexity in retrieval and fine-tuning methods, a variety of healthcare applications for LLMs, and creative ways for gamification and simulation. A brief overview of methodological trends, assessment techniques, and important areas of concentration in each of these categories is given by the selected visualizations.

5. CONCLUSION

The reviewed literature emphasizes how LLMs and RAG frameworks, which provide improved accuracy and contextual relevance in information retrieval, are rapidly advancing in medical education. Most current systems, however, lack the motivating and dynamic elements required for long-term student involvement. Although gamification has demonstrated potential for boosting engagement, there is yet little integration of gamification with AI-driven, domain-specific knowledge systems. To optimize knowledge acquisition and retention in the medical field, it is evident that solutions that successfully blend accurate, adaptive retrieval mechanisms with captivating, personalized learning experiences are required.

6. DISCUSSIONS

Future advancements in AI-powered medical education tools might investigate further in-depth customization by adjusting to a student's chosen learning modes in addition to their present level of expertise. Including multimodal learning materials, like annotated medical imagery, might improve understanding and retention even more. We can expect that advances in RAG architectures will improve retrieval accuracy and reduce hallucinations, though their reliability in high-stakes medical learning remains to be validated [28]. Gamification elements may develop to incorporate variable difficulty levels, real-time collaborative challenges, and immersive chat-based simulations; yet, their enduring effects on knowledge retention and clinical performance remain uncertain. Future research routes will also probably be influenced by ethical issues pertaining to explainability, bias in AI models, and data privacy. Overall, even if the prospects are promising, further empirical research is necessary to ascertain how well these advances can be applied in an actual medical education context.

7. LIMITATIONS

There are several restrictions in the reviewed literature that limit the current development of AI-driven, gamified medical education systems. While few studies have integrated Retrieval-Augmented Generation with gamification, most focus on learner engagement or retrieval accuracy separately. The reliability of reported results has been reduced by the lack of empirical validation of such systems in real-life classrooms or clinical training settings. Because there are no established standards for assessing gamified medical learning resources, cross-study comparisons are difficult to conduct. Bias, hallucination, and privacy are among the ethical and safety issues that are recognized but not always addressed in practice. Since many

studies concentrate on certain medical specializations, domain coverage is limited, which restricts the generalizability of the findings. Additionally, the long-term educational impact is questionable because the majority of evaluations measure engagement and retention over short timeframes.

REFERENCES

1. J. Vrdoljak, Z. Boban, M. Vilović, M. Kumrić, and J. Božić, "A Review of Large Language Models in Medical Education, Clinical Decision Support, and Healthcare Administration," Healthcare 2025, Vol. 13, Page 603, vol. 13, no. 6, p. 603, Mar. 2025, doi: 10.3390/HEALTHCARE13060603.
2. H. Ahmed Tahir Supervisors, T. Panula Mohammad Feli Alabi Rasheed Omobolaji, and H. Ahmed Tahir, "Development of a Medical Question-Answering System Using Large Language Models with Retrieval Augmented Generation and Prompt Engineering of a Medical Question-Answering System Using Large Language Models with Retrieval Augmented Generation and Prompt Engineering," 2025.
3. J. Swacha and M. Gracel, "Retrieval-Augmented Generation (RAG) Chatbots for Education: A Survey of Applications," Apr. 01, 2025, Multidisciplinary Digital Publishing Institute (MDPI). doi: 10.3390/app15084234.
4. C. Y. Lee, C. H. Lee, H. Y. Lai, P. J. Chen, M. M. Chen, and S. Y. Yau, "Emerging trends in gamification for clinical reasoning education: a scoping review," BMC Med Educ, vol. 25, no. 1, pp. 1–14, Dec. 2025, doi: 10.1186/S12909-025-07044-7/TABLES/2.
5. M. Sevgi, F. Antaki, and P. A. Keane, "Medical education with large language models in ophthalmology: custom instructions and enhanced retrieval capabilities," British Journal of Ophthalmology, vol. 108, no. 10, pp. 1354–1361, Oct. 2024, doi: 10.1136/BJO-2023-325046.
6. G. Xiong, Q. Jin, Z. Lu, and A. Zhang, "Benchmarking Retrieval-Augmented Generation for Medicine," Proceedings of the Annual Meeting of the Association for Computational Linguistics, pp. 6233–6251, 2024, doi: 10.18653/v1/2024.findings-acl.372.
7. Y. Wang, X. Ma, and W. Chen, "Augmenting Black-box LLMs with Medical Textbooks for Biomedical Question Answering," EMNLP 2024 - 2024 Conference on Empirical Methods in Natural Language Processing, Findings of EMNLP 2024, pp. 1754–1770, 2024, doi: 10.18653/v1/2024.findings-emnlp.95.
8. D. Oniani et al., "Enhancing Large Language Models for Clinical Decision Support by Incorporating Clinical Practice Guidelines," in 2024 IEEE 12th International Conference on Healthcare Informatics (ICHI), IEEE, Jun. 2024, pp. 694–702. doi: 10.1109/ICHI61247.2024.00111.
9. Y. Xie et al., "A Preliminary Study of o1 in Medicine: Are We Closer to an AI Doctor?," Sep. 2024, Accessed: Jul. 17, 2025. [Online]. Available: https://arxiv.org/pdf/2409.15277
10. A. Li, R. Shrestha, T. Jegatheeswaran, H. O. Chan, C. Hong, and R. Joshi, "Mitigating Hallucinations in Large Language Models: A Comparative Study of RAG-enhanced vs. Human-Generated Medical Templates," medRxiv, p. 2024.09.27.24314506, Sep. 2024, doi: 10.1101/2024.09.27.24314506.
11. Y. Liu, "The Application of RAG Technology in Traditional Chinese Medicine," pp. 402–408, Sep. 2024, doi: 10.2991/978-94-6463-512-6_43.

12. J. Wu et al., "Medical Graph RAG: Towards Safe Medical Large Language Model via Graph Retrieval-Augmented Generation," Oct. 2024, [Online]. Available: http://arxiv.org/abs/2408.04187
13. J. Hale, S. Alexander, S. T. Wright, and K. Gilliland, "Generative AI in Undergraduate Medical Education: A Rapid Review," J Med Educ Curric Dev, vol. 11, Jan. 2024, doi: 10.1177/23821205241266697.
14. B. ROHEEM, "Gamified Learning Platforms Using Reinforcement Learning," 2024, Accessed: Jul. 19, 2025. [Online]. Available: https://www.researchgate.net/profile/Elijah-William-2/publication/390866634_Gamified_Learning_Platforms_Using_Reinforcement_Learning/links/6800c33bded43315572a7bbe/Gamified-Learning-Platforms-Using-Reinforcement-Learning.pdf
15. Y.-F. Wang, Y.-F. Hsu, K.-T. Fang, and L.-T. Kuo, "Medical Education Online Gamification in medical education: identifying and prioritizing key elements through Delphi method Gamification in medical education: identifying and prioritizing key elements through Delphi method," 2024, doi: 10.1080/10872981.2024.2302231.
16. D. Tolks, J. J. Schmidt, and S. Kuhn, "The Role of AI in Serious Games and Gamification for Health: Scoping Review.," JMIR Serious Games, vol. 12, no. 1, p. e48258, Jan. 2024, doi: 10.2196/48258.
17. A. Hamilton, "Artificial Intelligence and Healthcare Simulation: The Shifting Landscape of Medical Education," 2024, doi: 10.7759/cureus.59747.
18. M. Špeletić, M. Špeletić, M. Protić, and S. Nikolić, "EXPLORING RAG IN MEDICAL QUESTION ANSWERING: INTEGRATING LLMS AND VECTOR DATABASES," UNITECH – SELECTED PAPERS, 2024, doi: 10.70456/gajm2853.
19. X. Chen et al., "EyeGPT for Patient Inquiries and Medical Education: Development and Validation of an Ophthalmology Large Language Model," J Med Internet Res, vol. 26, no. 1, p. e60063, Dec. 2024, doi: 10.2196/60063.
20. S. S. Manathunga and Y. A. Illangasekara, "Retrieval Augmented Generation and Representative Vector Summarization for large unstructured textual data in Medical Education," Aug. 2023, Accessed: Jul. 17, 2025. [Online]. Available: https://arxiv.org/pdf/2308.00479
21. H. Nori, N. King, S. M. Mckinney, D. Carignan, E. Horvitz, and M. 2 Openai, "Capabilities of GPT-4 on Medical Challenge Problems," Mar. 2023, Accessed: Jul. 17, 2025. [Online]. Available: https://arxiv.org/pdf/2303.13375
22. J. Liu, C. Wang, and S. Liu, "Utility of ChatGPT in Clinical Practice," J Med Internet Res, vol. 25, 2023, doi: 10.2196/48568.
23. J. Clusmann et al., "The future landscape of large language models in medicine," Communications Medicine, vol. 3, no. 1, Dec. 2023, doi: 10.1038/S43856-023-00370-1.
24. T. Dave, S. A. Athaluri, and S. Singh, "ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations," Front Artif Intell, vol. 6, 2023, doi: 10.3389/FRAI.2023.1169595.
25. Y. Al Ghadban et al., "Transforming Healthcare Education: Harnessing Large Language Models for Frontline Health Worker Capacity Building using Retrieval-Augmented Generation," Dec. 17, 2023. doi: 10.1101/2023.12.15.23300009.

26. A. D. Gajardo Sánchez, L. R. Murillo-Zamorano, J. López-Sánchez, and C. Bueno-Muñoz, "Gamification in Health Care Management: Systematic Review of the Literature and Research Agenda," Sage Open, vol. 13, no. 4, Oct. 2023, doi: 10.1177/21582440231218834.
27. H. Y. So, P. P. Chen, G. K. C. Wong, and T. T. N. Chan, "Simulation in medical education," Mar. 01, 2019, Royal College of Physicians of Edinburgh. doi: 10.4997/JRCPE.2019.112.
28. R. Garett and S. D. Young, "Health Care Gamification: A Study of Game Mechanics and Elements," Technology, Knowledge and Learning, vol. 24, no. 3, pp. 341–353, Sep. 2019, doi: 10.1007/s10758-018-9353-4.