

Systematic Review

Generative AI-Integrated Virtual Agents and Simulations in Health Professions Education: A Systematic Review

Xining (Ning) Wang¹, Andrew O'Malley^{1,*}, Alun Hughes¹ and Md Saifuddin Khalid²¹ School of Medicine, University of St Andrews, St Andrews KY16 9AJ, UK; xwang8@tcd.ie (X.W.); ah200@st-andrews.ac.uk (A.H.)² Department of Applied Mathematics and Computer Science Statistics and Data Analysis, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark; skhalid@dtu.dk

* Correspondence: aso2@st-andrews.ac.uk

Abstract

The rapid development of generative artificial intelligence (GenAI) is transforming both the health sector and health profession education, although AI-based systems have existed in these sectors for decades. GenAI-integrated virtual agents and simulations now play novel and critical roles in simulation-based education and are potential solutions to enhance the adaptability of health profession education. This systematic review was conducted using the PRISMA guidelines and explores how GenAI-integrated virtual agents and simulations are being applied in health profession education, with a particular focus on their educational impact, technical features and functionalities, and current limitations. This review aims to synthesize the pedagogical value and technological design of GenAI-integrated simulations and to inform health professionals and educators about the effective use, impact, and challenges of GenAI in health education simulations. A total of 16 papers were reviewed. Results show that GenAI-integrated virtual agents and simulations have potential to enhance clinical communication, diagnostic accuracy, multilingual interactions, and learner confidence for health profession education. Related theoretical, technological, and educational implications of generative AI-integrated virtual agents and simulations are discussed to inform future design and application. Limitations include insufficient educational effectiveness, response accuracy issues, and unresolved ethical and privacy concerns. Future studies should focus on long-term efficacy, ethical considerations, and optimizing AI-human collaboration in various health profession education contexts.

Keywords: generative AI; virtual agents; simulation; health professions education; systematic review



Academic Editor: Andrew Williams

Received: 9 March 2026

Revised: 9 June 2026

Accepted: 11 June 2026

Published: 18 June 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Simulations have become an indispensable tool in health profession education (Elendu et al., 2024). Numerous studies have demonstrated that simulation-based education utilizes scenarios and tools to replicate real-world clinical situations, providing learners with a safe and controlled environment in which to practice clinical skills, enhance decision-making abilities, develop teamwork competencies, and improve clinical outcomes (Diaz-Navarro et al., 2024; Elendu et al., 2024). Simulations are used across various disciplines in health science, including medicine, nursing, pharmacy, and public health, and they come in diverse forms, each tailored to specific educational goals (Kononowicz et al., 2019). The general effectiveness of traditional simulation methods in health education can broadly be divided into several categories: (a) high-fidelity whole-body mannequins, which emphasize

technical skill acquisition and crisis management; (b) standardized patient interactions, which target communication skills and clinical reasoning, including use of task trainers and partial mannequins; (c) models designed for learners to master specific procedures in a safe and controlled setting before practicing them in real-time scenarios; and (d) computer-based simulations, including using a series of digital tools and virtual patient simulations, to facilitate self-directed learning through interactive scenarios (Elendu et al., 2024).

Simulation-based education aims to bridge the gap between theoretical knowledge and real-world practice, which helps prepare healthcare professionals for the complexities of patient care and clinical communication (Sevdalis, 2015). However, there are certain limitations to applying traditional simulations in health education. Firstly, the high costs associated with high-fidelity manikins and the training and maintenance of standardized patients pose a substantial financial burden (Usheva et al., 2024). These high costs could make these tools inaccessible for many institutions, particularly those in resource-limited settings. Secondly, the scalability of traditional simulations is often constrained, as they require significant infrastructure, specialized personnel, and time-intensive setups, which are difficult to replicate on a larger scale for educational purposes (So et al., 2019). Thirdly, the lack of dynamic interactivity in pre-programmed virtual patients or simulators limits their ability to adapt to learners' actions in real-time (Tavarnesi et al., 2018), which may result in a less engaging and realistic learning experience. Finally, traditional simulations are restricted in providing personalized learning experiences, as they are not designed to cater to the diverse backgrounds, skill levels, and individual needs of learners (Usheva et al., 2024). These limitations of high-fidelity manikin-integrated simulations demonstrate the need for a new approach to increase the accessibility, adaptability, and effectiveness of simulation-based health education.

Generative artificial intelligence (GenAI) has developed rapidly in recent years, particularly in large language models (LLMs) and multimodal (e.g., text, audio, and video) virtual agents (Omirgaliyev et al., 2024). In health education, virtual agents usually encompass a wide scope of technologies, ranging from simple rule-based chatbots to advanced, interactive systems capable of simulating complex human conversations (Omirgaliyev et al., 2024). In health profession education, virtual agents can play various roles (e.g., simulated patients, health assistants, and counseling services). Early virtual agents relied on deterministic, rule-based decision trees with scripted dialog paths that restricted dynamic interaction (Krishnan, 2025). Recently, the advent of GenAI presented an ontological shift, which has fundamentally changed the nature of simulations (Floridi & Chiriatti, 2020). GenAI's underlying architecture often hides predictability that is beneficial for medical training. Mechanisms such as Reinforcement Learning from Human Feedback (RLHF), probability-based next-word prediction, and controlled temperature parameters naturally steer the AI toward producing highly standardized, textbook-like responses (Ouyang et al., 2022). For health profession education, this predictable typicality is advantageous. It allows institutions to efficiently and consistently generate quintessential clinical scenarios (e.g., standard patient histories and typical symptom presentations).

Despite these theoretical advantages, the pedagogical integration of GenAI agents in health profession education is underexplored. Current literature frequently highlights the technical capabilities of generative AI, yet there is a critical lack of evidence-based synthesis regarding how these systems influence learning outcomes, clinical skill acquisition, and professional identity formation in health profession education. This systematic review aims to provide a comprehensive overview of current research, outline the methodological procedures adopted in selecting and analyzing studies, and synthesize key findings and insights drawn from the reviewed literature.

2. Related Work

Recent developments in virtual agents utilize various GenAI models to create agents that go beyond simple chatbots (e.g., GPT-based systems, LLaMa, CLIP, and Whisper) (Casheekar et al., 2024). These can be natural, personalized, and complex interactions that enhance training and education (Bozkurt, 2023). In health profession education, these GIVAS serve as simulated patients, enabling learners to practice clinical skills (such as history-taking, diagnosis, and communication) through interactive dialogs (Pham et al., 2025) while also generating rich multimedia content like images, videos, and 3D models to create realistic virtual environments (Bozkurt, 2023; Preiksaitis & Rose, 2023). The integration of GenAI with extended reality (XR) technologies (including VR, AR, and MR) further enriches training by enabling immersive scenarios where GenAI dynamically responds to learners' actions and simulates complex clinical and emotional situations (Rossi et al., 2024). However, as GenAI's applications grow in health profession education, developing smart agents and simulations involves challenges alongside high computational demands (e.g., integrating NLP, computer vision, real-time rendering, and affective computing) and complex ethical issues (e.g., data privacy and bias) (Hirzle et al., 2023). Trust and psychological factors are crucial for user acceptance, and tailored guidelines are needed to ensure responsible AI use in sensitive fields like healthcare (Siala & Wang, 2022).

Furthermore, research on GIVAS's long-term efficacy and usability is limited, with an unclear understanding of optimizing GenAI for diverse contexts and ensuring access for underserved populations (Pham et al., 2025). Health profession education also lacks comprehensive, evidence-based guidance for integrating AI tools and virtual agents effectively into curricula, leaving educators reliant on fragmented studies, intuition, or vendor materials instead of systematic research (Zawacki-Richter et al., 2019). To date, no comprehensive synthesis exists to inform the instructional effectiveness, pedagogical assumptions, or learning impact of GIVAS in health profession education. Health profession education has decades of simulation research, but generative AI introduces a qualitatively different form of simulation (e.g., dynamic, adaptive, and dialogic), yet we lack a pedagogical synthesis of how this changes learning processes. Existing reviews have focused on technical capabilities or reported outcomes of AI systems, but have not examined how different configurations of generative AI virtual agents function pedagogically within simulation-based health profession education. This creates a gap between research and practice. Therefore, this systematic review aims to evaluate the application of GIVAS in health profession education, with a particular focus on their educational impact, technical features and functionalities, as well as current limitations. To achieve this goal, this study proposes the following research questions (RQs):

- RQ1. What is the current evidence on the effectiveness of GenAI-integrated virtual agents and simulations (GIVAS) in improving learning outcomes in health profession education?
- RQ2. What are the key technical features and functionalities of GenAI-integrated virtual agents and simulations (GIVAS) currently used in health profession education?
- RQ3. What are the reported challenges of GenAI-integrated multimodal simulations (GIVAS) and implications for health profession education?

3. Methodology

This study employs the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA, 2020) framework as the primary guide for all aspects of the search and review process (Page et al., 2021). A systematic search was conducted across four databases: *PubMed* (for covering biomedical and life sciences literature), *Scopus* (for covering multidisciplinary coverage, including social sciences and humanities), *Web of Science* (for

covering high-quality, citation-indexed research across disciplines), and *Google Scholar* (for complementary search, including gray literature and non-indexed sources) were used as the databases for identifying articles reporting empirical studies on GIVAS in health education.

The literature search covered studies published between January 2019 and February 2025. The year 2019 was selected as the starting point because the release of GPT-2 marked a significant milestone in the development of transformer-based generative language models and conversational AI systems (Radford et al., 2019). Next, significant advancements and increased adoption occurred following the release of models such as GPT-4 (2023) and GPT-3 (2020), which marked a turning point in the accessibility and application of GenAI. The final search was completed in February 2025, after which the review was conducted and the manuscript prepared. Figure 1 presents the PRISMA flow diagram with details of the study selection process, including identification, screening, eligibility, and inclusion stages.

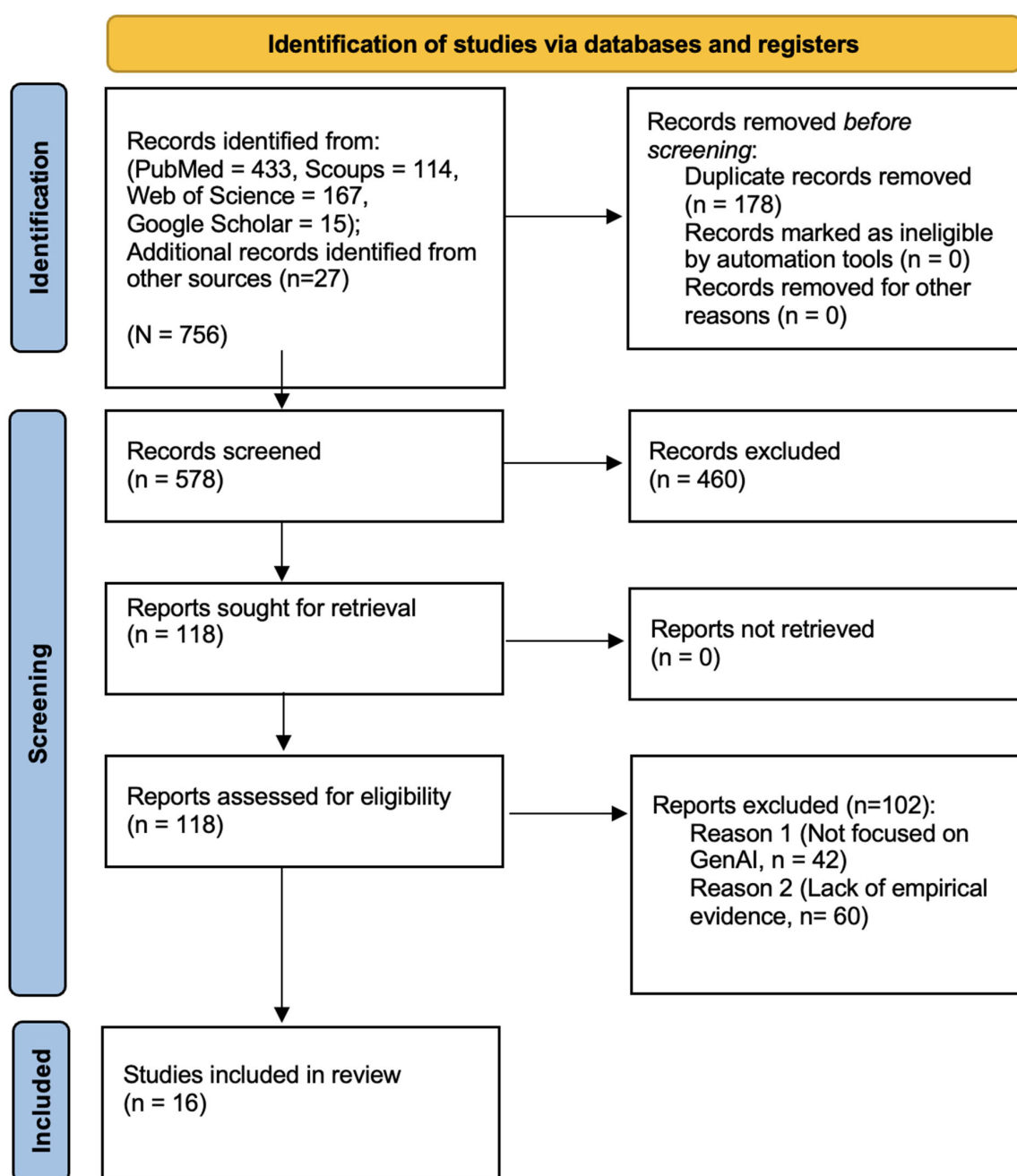


Figure 1. PRISMA flow chart.

Literature Search Strategy

The initial search involved identifying keywords related to the research topic (e.g., GenAI, Virtual Agent, and Health Education) and developing a list of synonyms and related terms for each concept. Boolean operators (AND, OR, NOT) and potential truncation (*) were used to refine the search. Identified keywords were combined to create the following search strings, which were used for literature retrieval from the databases:

("Generative AI" OR "GenAI" OR "Artificial Intelligence" OR "AI" OR "Large Language Models" OR "LLMs" OR "ChatGPT" OR "GPT-4" OR "Natural Language Processing" OR "NLP")

AND

("Virtual Agent" OR "Virtual Patient" OR "Simulated Patient " OR "SimPatient " OR "Virtual Assistant" OR "Conversational Agent" OR "Chatbot" OR "Digital Assistant" OR "Intelligent Tutoring System" OR "AI Tutor")

AND

("Health Professions Education" OR "Medical Education" OR "Clinical Education" OR "Healthcare Training" OR "Patient Education" OR "Health Literacy" OR "Nursing Education" OR "Public Health Education")

Criteria (Table 1)

Although the review primarily focused on peer-reviewed empirical studies, a small number of non-peer-reviewed sources (e.g., preprints from arXiv) were included where they provided relevant and timely evidence on emerging GenAI applications. This decision was made by the research team due to the rapidly evolving nature of GenAI technologies, where significant developments are often first reported outside traditional peer-reviewed outlets. All included studies were assessed by two researchers for relevance and methodological clarity prior to inclusion.

Table 1. Inclusion and exclusion criteria.

<i>Inclusion Criteria</i>	<i>Exclusion Criteria</i>
- Studies that are empirical studies. - Studies utilizing GenAI models. - GIVAS agents contribute to dialog generation, scenario adaptation, personalized feedback, or immersive learning experiences. - Studies that focus on the health sector or are related to health profession education (e.g., medical education, public health, nursing education, and mental health education). - Outcomes are measured using quantitative (e.g., test scores, performance metrics) or qualitative (e.g., student feedback, thematic analysis) methods. - Peer-reviewed journals, conference proceedings, and scholarly preprints with rigorous empirical methodology (e.g., arXiv, medRxiv, SSRN). - Studies that were published in English within the last six years (since 2019) to ensure relevance to the latest GenAI advancements.	- Studies that are commentary, opinion pieces, editorials, conceptual papers, or all types of reviews without empirical data. - Studies that use non-GenAI models. - Simulations or applications without virtual agents to generate interactive dialog or scenarios. - Studies outside the scope of health education. - Studies that are dissertations and theses. - Non-English publications or articles published more than six years ago.

3.1. Data Extraction and Synthesis

Data were extracted using a structured extraction form that captured study characteristics, AI system features, virtual agent roles, educational context, outcome measures, and reported limitations. The extraction form was developed deductively based on SPIDER-informed dimensions (see Section 3.2) and iteratively refined during the review process. A total of 756 studies appeared from the preliminary searching phase. After removing duplicated and non-qualifying studies ($n = 740$), a final set of 16 studies was selected for in-depth review (Table 2). The initial screening of all studies, including keywords, titles, abstracts, and full texts, was conducted by one researcher. Subsequently, a second researcher verified and approved the screening decisions to ensure consistency with the PRISMA guideline. Content Analysis and Situational Analysis, as shown in Figure 2, were utilized to synthesize and interpret data collected from the reviewed studies. Content Analysis facilitated the systematic identification and categorization of recurring themes related to GIVAS and health profession education (Rosen & Shoenberger, 2021). All reviewed studies were critically examined for methodological clarity, relevance to the research questions, and completeness of reported findings during the screening and data extraction process.

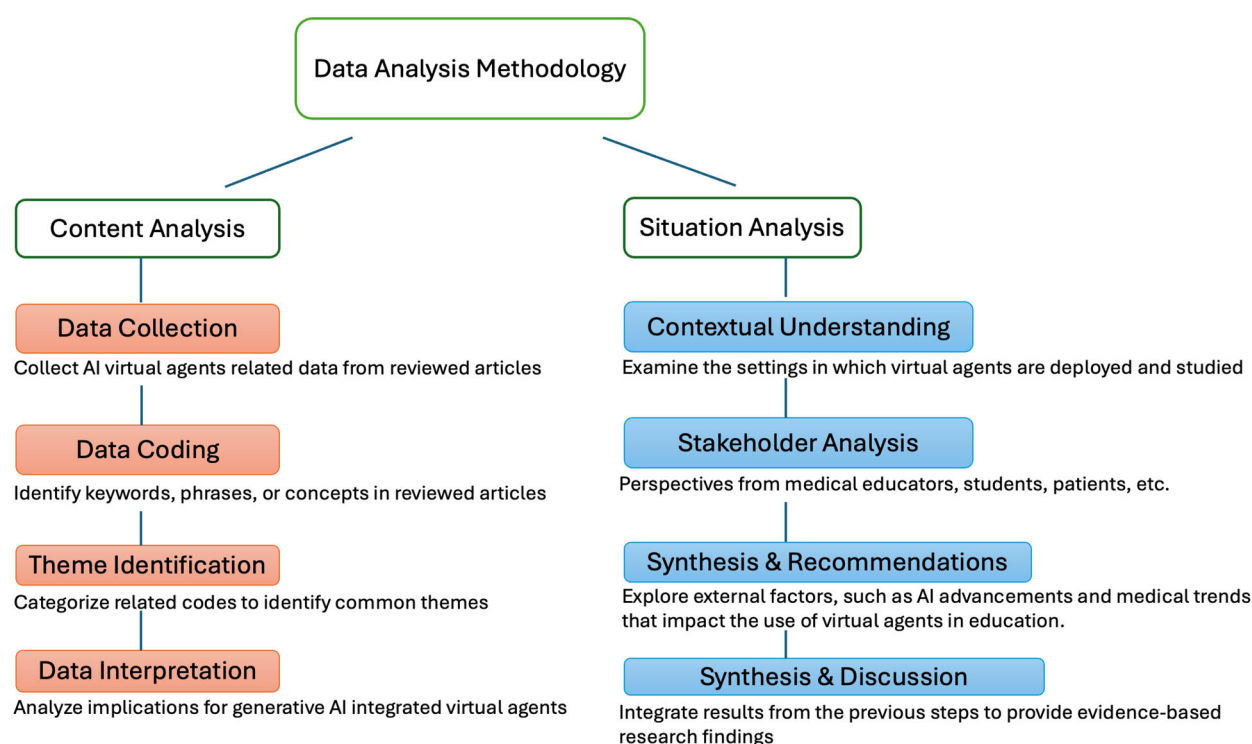


Figure 2. Methodology of data analysis (synchronize content analysis and situation analysis).

The PRISMA framework guided the identification, screening, eligibility assessment, and inclusion of studies. A PRISMA flow diagram illustrating the study selection process is provided in Figure 1. As the studies selected for review included diverse study designs (i.e., development studies, qualitative research, case demonstrations, and mixed-methods studies), a formal risk-of-bias assessment was not conducted. This systematic review focuses on descriptive and conceptual synthesis of study characteristics, technological configurations, and pedagogical implications, and it was therefore not prospectively registered in a systematic review registry.

Table 2. Summary of characteristics of reviewed studies.

ID	Authors & Year	Title	Study Type	Country/Cultural Context	Aim	Intelligence/AI System	Virtual Agent Type and Pedagogical Role	Specialty	Sample Size	Target Population	Conclusion Summary
1	(Yan & Alterovitz, 2024)	A General-purpose AI Avatar in Healthcare	Conceptual & development study	USA	To develop and evaluate a general-purpose AI avatar, enhancing the interaction and engagement with patients.	LLM A system based on ChatGPT-3.5 and enhanced through prompt engineering	Text-based virtual agent as doctor	Healthcare	No human sample	Patients seeking medical advice	AI could improve patient engagement and diagnostic relevance
2	(Badawy et al., 2025)	A pilot study of generative AI video for patient communication in radiology and nuclear medicine	Mixed methods	Australia/English Thai translation	To evaluate the effectiveness of communication skills training and to explore the potential of AI in enhancing such training.	HeyGen	Video-based AI avatar of a medical physicist	Language translation of personalized patient information in Radiology and	N = 13	Thai-speaking medical physicists and postgraduate students	GenAI can be an effective tool for personalized, multilingual patient communication, specifically using the Thai language.
3	(Chen et al., 2025)	Customizing Generated Signs and Voices of AI Avatars: Deaf-Centric Mixed-Reality Design for Deaf-Hearing Communication	Qualitative Study	USA/Interpretation	Interactive communication and collaborative learning between learners of mixed hearing and signing abilities	DeepMotion for realistic avatar creation. Unreal MetaHuman for hyper-realistic facial expressions. Apple Vision Pro for mixed reality visualization.	Embodied AI avatar (sign & voice generation with mixed reality overlay) AI-based sign–speech interpreter	Deaf hearing communication	N = 15	Deaf and hard of hearing individuals	The design can be interpreted to enhance face-to-face communication for hard hearing people
4	(Ba et al., 2024)	Enhancing clinical skills in pediatric trainees: a comparative study of ChatGPT-assisted and traditional teaching methods	Experimental Study	China	To evaluate the effect of ChatGPT-assisted teaching of clinical skills to pediatric interns	ChatGPT 4 by OpenAI	LLM-based conversational AI tutor	Pediatric training	N = 77	Pediatric trainees/students	ChatGPT-assisted instruction significantly enhances clinical skills, particularly in patient communication and clinical judgment
5	(Greca et al., 2024)	Enhancing therapeutic engagement in Mental Health through Virtual Reality and Generative AI: a co-creation approach to trust building	Qualitative study	Italy	Enhancing patient trust and engagement in mental health treatment with VR and GenAI.	FLUX.1-Schnell and StableFast3D Unity (custom VR system)	3D GenAI avatars	Mental health	NA	Mental health patients	The co-creation approach of VR and GenAI can enhance patient agency, emotional anchoring, and overall therapeutic outcomes.
6	(Gutiérrez Maquilón et al., 2024)	Integrating GPT-Based AI into Virtual Patients to Facilitate Communication Training Among Medical First Responders	Mixed-methods study	Austria	To investigate the usability and effectiveness of ChatGPT-based generative voice agents in simulating verbal interactions with virtual patients during emergency scenarios	ChatGPT (GPT-3.5 Turbo) by OpenAI, ElevenLabs for text-to-speech (TTS) voice synthesis	Generative voice agent Text-to-speech, AI-driven virtual patient (voice-based)	Emergency medicine and triage training	N = 24	Medical first responders	It can enhance communication training for first responders through realistic verbal interactions
7	(Sevgi et al., 2024)	Medical education with large language models in ophthalmology: custom instructions and enhanced retrieval capabilities	Case demonstration	UK	To investigate how customized GPTs can enhance ophthalmology education	Custom GPT 4 by OpenAI	LLM-based conversational AI tutor	Ophthalmology	No human samples	Medical students and practicing ophthalmologists	Custom GPTs significantly enhance ophthalmology education
8	(Lv et al., 2025)	Multimodal Metaverse Healthcare: A Collaborative Representation and Adaptive Fusion Approach for Generative AI-Driven Diagnosis	Study of development and evaluation	China/Poland/India/Korea/USA	Using multimodal natural language understanding in metaverse healthcare environments with text, audio, and video features.	Multimodal deep learning framework	3D virtual avatars with multimodal generative AI Backend diagnostic modeling & decision support	General healthcare and diagnostics	N = 2199 video clips and 93 human reviewers	Patients in critical care	It may help people effectively enhance diagnostic accuracy and patient interaction
9	(Contreras et al., 2024)	Revolutionising Faculty Development and Continuing Medical Education Through AI-Generated Videos	Qualitative study	Switzerland	To have objective feature assessment and learner feedback to guide the adoption of effective AI video generation tools	HeyGen, Synthesia, Colossian, and HourOne	AI-generated video avatars as an instructor	Continuing medical education	n = 25	Medical educators and learners	AI-generated videos can be a viable alternative to traditionally produced educational videos, but ethical disclosures are needed.
10	(Chu & Goodell, 2024)	Synthetic Patients: Simulating Difficult Conversations with Multimodal Generative AI for Medical Education	Development/ system design study	USA	To explore the use of AI patients to simulate sensitive conversations in medical training	Custom GPT-4 (OpenAI), Midjourney, Stable Diffusion, ElevenLabs, and HeyGen	Multimodal LLM-based conversational agent Synthetic patient (AI-driven virtual patient)	Communication skills training in healthcare	NA	Medical students and healthcare trainees	The AI agent can offer high-fidelity, scalable simulations for training medical professionals in difficult conversations
11	(Samala & Rawas, 2024)	Generative AI as Virtual Healthcare Assistant for Enhancing Patient Care Quality	Quantitative study	Indonesia/ Lebanon	To investigate ChatGPT's effectiveness in enhancing patient care	ChatGPT (GPT-3.5 by OpenAI) with TensorFlow	Conversational AI chatbot virtual healthcare assistant	Chronic disease management	500 simulated or unspecified patient–AI interactions	General patients	ChatGPT effectively enhances patient care by providing accurate medical advice

Table 2. Cont.

ID	Authors & Year	Title	Study Type	Country/Cultural Context	Aim	Intelligence/AI System	Virtual Agent Type and Pedagogical Role	Specialty	Sample Size	Target Population	Conclusion Summary
12	(Abi-Rafeh et al., 2023)	Complications Following Facelift and Neck Lift: Implementation and Assessment of Large Language Model and Artificial Intelligence (ChatGPT) Performance Across 16 Simulated Patient Presentations	Mixed-methods evaluation	Canada/USA	To study how the integration of GPT-based AI in a mixed reality (MR)–VP could support communication training	Large language model GPT-3.5 Turbo by OpenAI	Text-to-speech virtual agent Patient-facing triage & guidance agent	Plastic surgery and postoperative care	N = 16	Postoperative patients	AI generates differential diagnoses and red-flag warnings for postoperative complications, but tends to overestimate urgency
13	(Chheang et al., 2024)	Towards Anatomy Education with Generative AI-based Virtual Assistants in Immersive Virtual Reality Environments	Study of development and evaluation	USA	To present a VR environment designed to support human anatomy education using generative AI	GPT-3.5 by OpenAI Speech-to-text/text-to-speech (Azure)	LLM-based conversational agent Embodied AI virtual assistant in VR	Anatomy education	N = 16	University students with medical anatomy knowledge	AI-embodied virtual assistants can provide interactive, personalized learning experiences in VR
14	(Sardesai et al., 2024)	Utilizing Generative Conversational Artificial Intelligence to Create Simulated Patient Encounters: A Pilot Study for Anaesthesia Training	Study of development and evaluation	UK	To evaluate a ‘no-code’ generative AI solution to create 2D and 3D virtual avatars	Convai & ChatGPT(3.5)	AI virtual patient (2D & 3D) for communication & consent training	Anaesthesia training	N = 15	Anaesthetic students	Students reported notable increases in their confidence levels. Students who used the resources outperformed their counterparts in clinical skills assessments
15	(Mittenentzwei et al., 2024)	AI-Assisted Character Design in Medical Storytelling with Stable Diffusion	Development/case study	Germany/Norway	Presenting semi-automated character design pipeline using Stable Diffusion for creating virtual patients	Stable Diffusion Digital Humans Leonardo AI	Text-to-image generative AI	Health education	No human samples	General public	It can enhance medical storytelling by creating realistic, data-driven narratives.
16	(Mool et al., 2024)	Using Generative AI to Simulate Patient History-Taking in a Problem-Based Learning Tutorial: A Mixed-Methods Study	Mixed-methods study	USA	To explore how voice-to-voice interaction with a 3D avatar affects medical students’ learning	ConvAI within Unreal Game Engine 5.0 and Metahuman	LLM-based conversational agent (voice-enabled) Virtual agent-type AI-driven virtual patient (3D avatar)	Medical history-taking Medical problem-based learning	N = 26	Medical students	It can improve students’ history-taking skills and engagement

3.2. Pedagogy-Based Analytical Framework—SPIDER

To examine how GIVAS influence learning and pedagogy, this study adopted a pedagogy-oriented analytical framework adapted from the Sample, Phenomenon of Interest, Design, Evaluation, and Research type (SPIDER) approach (Methley et al., 2014). The SPIDER framework enables a structured synthesis of how different AI agent configurations function within educational designs and support specific learning processes. Specifically, the framework guided analysis across five dimensions: (1) learner context (e.g., medical students, trainees, and clinicians), (2) pedagogical phenomenon of interest (e.g., communication skills, clinical reasoning, and engagement), (3) educational design and agent role (e.g., AI as simulated patient, tutor, or instructional avatar), (4) learning mechanisms (e.g., experiential practice, scaffolding, repetition, and immersion), and (5) type of evidence reported (e.g., pilot study, mixed-methods, and usability evaluation). This framework aims to guide data analysis based on pedagogical perspectives by linking AI agent configurations to learning contexts, educational designs, and underlying learning mechanisms.

4. Data Analysis

Situational Analysis was employed to map the social, technological, and institutional contexts within which these empirical studies occurred, highlighting the interplay between GenAI, virtual agents in multimodal settings (including immersive XR simulations), educational cases, and clinical practice (Clarke, 2005). This combined method was chosen to present a comprehensive understanding of the themes and contextual factors refined from the listed literature (Figure 2).

5. Findings

An overview of the 16 empirical papers is summarized in Table 2. Collectively, the reviewed studies demonstrate a wide-ranging application of generative GIVAS in health profession education across diverse geographical and cultural contexts. For instance, most studies were from the USA ($n = 6$), accounting for approximately 37.5% of the reviewed papers. Others are from China ($n = 2$), the UK ($n = 2$), and individual cases from Australia, Austria, Switzerland, and Italy. Several studies adopted cross-cultural or international perspectives, including multilingual communication and culturally adaptive virtual agents. This geographical distribution suggests that current research activity is concentrated in a limited number of countries, particularly the United States, which reflects a greater availability of technical resources and institutional support for AI-enabled educational innovation.

At the same time, the presence of cross-cultural studies indicates growing global interest in the pedagogical potential of GIVAS, despite uneven research capacity across regions. Table 2 (column: Name/Platform of AI system) shows that a substantial number of studies ($n = 10$) employ ChatGPT (versions 3.5 or 4) as the foundational language model, fine-tuned or adapted to simulate natural, context-aware patient dialogs. Meanwhile, a subset of papers integrates multimodal features with other tools, including ElevenLabs (used for voice synthesis), HeyGen and Synthesia (for AI-generated avatars and video), and combinations of commercial text-to-speech services. Several systems were developed using custom or hybrid platforms, combining generative language models with visual design tools (e.g., Midjourney or institutional simulation environments, with the aim of enhancing immersion and interaction fidelity).

Beyond technical configuration, several studies explicitly aligned GIVAS with existing medical curricula or assessment practices, including: ChatGPT-assisted teaching of clinical skills in pediatrics (ID 4); simulation-based communication training for first responders (ID 6); curriculum-embedded anatomy education in immersive VR (ID 13); and problem-based learning tutorials using AI-driven virtual patients (ID 16). This distribution highlights

the centrality of ChatGPT in experimental AI applications in health profession education, while also indicating the growing ecosystem of multimodal generative tools. Increasingly experimental applications were developed by GenAI technology (especially OpenAI), which led to the emergence of a broader ecosystem of multimodal generative tools supporting various health educational designs. Virtual agents assumed a variety of forms, including AI avatars ($n = 6$), conversational chatbots ($n = 4$), generative voice agents ($n = 3$), and multimodal or embodied agents deployed in immersive environments such as virtual or mixed reality ($n = 3$). This reflects the versatile role of GIVAS in supporting responsive, personalized, and context-sensitive learning or communication experiences (e.g., virtual patients, teaching assistant, speech interpreter, and virtual agents embodied in VR). These pedagogical roles were associated with distinct educational purposes. For example, simulated patient roles primarily supported experiential communication and history-taking practice, whereas AI tutors, AI assistants, or instructional avatar-based roles were oriented toward content delivery, scaffolding, and guided learning.

The educational applications spanned a broad range of specialties, including mental health, communication support for deaf and hard-of-hearing individuals, clinical and triage skills training, pharmacy and oncology education, radiology, medical translation, AI ethics, and curriculum development. Across these domains, GIVAS were most frequently used to enhance learners’ communication skills, engagement, and affective competencies, often through emotionally responsive and culturally diverse virtual patient interactions. Many studies prioritized demographic variability (e.g., age, ethnicity, regional accents, and literacy levels), which suggested an emerging emphasis on inclusivity and cultural sensitivity in AI-powered educational design. Target populations were similarly heterogeneous, encompassing medical students and postgraduate trainees ($n = 5$), patients (general, pediatric, or mental health populations; $n = 4$), healthcare professionals including radiologists and medical physicists ($n = 3$), and broader communities such as educators or the general public. This trend indicates that current empirical research on GIVAS has focused primarily on early-stage adoption, feasibility, and pedagogical alignment. It can lay the groundwork for future studies to examine learning mechanisms and longer-term educational outcomes in greater depth.

The findings of this research are systematically summarized, including the key themes (categories), authors, and keywords from the reviewed studies (Table 3). It provides an overview of the number of papers contributing to each categorized theme, which allows for an organized presentation of the distribution of research focus across the field.

Table 3. Categories, studies, and keywords of reviewed studies.

Category	Studies	Keywords
C1—Strengthening Health Profession Education	Mittenentzwei et al. (2024) Ba et al. (2024) Gutiérrez Maquilón et al. (2024) Contreras et al. (2024) Sardesai et al. (2024) Chu and Goodell (2024) Chheang et al. (2024) Mool et al. (2024)	Enhancing students’ perceived self-efficacy and confidence; promoting authentic learning environments; AI-driven storytelling Developing empathy and communication skills; tailoring education to learner needs; providing alternative pathways in health education
C2—Potential to Improve Clinical Skills	Ba et al. (2024) Samala and Rawas (2024) Lv et al. (2025) Chheang et al. (2024) Sardesai et al. (2024) Mool et al. (2024)	Enhancing diagnostic accuracy and treatment recommendations; clinical decision-making efficiency; supporting clinical skill acquisition and automation reliability

Table 3. *Cont.*

Category	Studies	Keywords
C3—Enhanced Engagement and Human-like Interaction	Yan and Alterovitz (2024) Chen et al. (2025) Greca et al. (2024) Mool et al. (2024) Chu and Goodell (2024) Sardesai et al. (2024)	Patient engagement; human-like interactions to improve relational dynamics; building attributable trust between users and AI systems
C4—Communication Skills Improvement	Badawy et al. (2025) Mittenentzwei et al. (2024) Chen et al. (2025) Ba et al. (2024) Greca et al. (2024) Samala and Rawas (2024) Gutiérrez Maquilón et al. (2024) Chu and Goodell (2024) Mool et al. (2024)	Bridging human–AI communication gaps; improving time coordination in dialogues Improving medical students’ communication; enhancing patient-centered communication in clinical care
C5—Personalized Learning and Tailored Learning Environments	Badawy et al. (2025) Mittenentzwei et al. (2024) Chen et al. (2025) Ba et al. (2024) Chu and Goodell (2024) Mool et al. (2024)	Creating personalized patient information materials; individualized learning pathways; using personality representation to tailor learning or interaction content
C6—Others	Abi-Rafeh et al. (2023) Chen et al. (2025) Greca et al. (2024) Sardesai et al. (2024)	Emotional display modulation; AI-driven empathy simulation; emotional anchoring to build trust; promoting patient agency and autonomy; realistic training scenarios; AI conversational accuracy range (77–86%)

Note: C = Categorization.

Next, Table 4 highlights four aspects of GIVAS, which could enhance health education learning outcomes (**RQ1**): (1) enhanced educational and clinical skills, (2) engagement, (3) communication, and (4) personalized learning. Also, evidence of advanced NLP technology, AI-embedded multilingual support, and multimodal interaction to diversify versatile learning experiences were summarized to illustrate key features and functionalities of GIVAS (**RQ2**). With respect to the challenges of GIVAS (**RQ3**), many virtual agents are unable to fully replicate human interactions, often encountering difficulties in understanding accented or rapid speech, and potentially exhibiting biases in their applications.

Table 4. Summary of key findings related to research questions.

RQs	Content Analysis (Initial Themes)	Situation Analysis ¹ (Institutional and Social Context)	Situation Analysis ² (Technological Context)	Key Findings
RQ1: Learning Outcomes	- Enhanced educational and clinical skills - Engagement and human-like Interaction - Communication - Personalized learning	Institutions explored AI to supplement skill-based training, and this integration was selective and still experimental. Learners reported increased confidence, skill readiness, and engagement with AI tools. Virtual agents provided personalized support, promoting inclusivity and learner motivation.	Multimodal AI agents (e.g., ChatGPT, VR) enhanced practical skills (e.g., communication and interaction), but not theoretical knowledge. AI tools such as MetaHuman and HeyGen can simulate human traits but struggle with latency, body language, and emotional expression.	GIVAS can improve non-technical skills, offer immersive simulations, and provide personalized learning experiences in health education.

Table 4. *Cont.*

RQs	Content Analysis (Initial Themes)	Situation Analysis ¹ (Institutional and Social Context)	Situation Analysis ² (Technological Context)	Key Findings
RQ2: Features and Functionalities	- Advanced NLP technology - AI-embedded Multilingual Support - Multimodal interaction	Research teams were interested in NLP but faced challenges integrating it into standardized learning systems. Virtual agents supported varied formats (e.g., voice, video, text); learners can benefit from broader access. Institutions may consider accessibility and fairness in deploying voice-interactive systems.	Adaptive AI offers tailored content, multilingual support, and dynamic feedback (e.g., voice/text/video). Conversational agents fostered engagement but also sounded robotic or overly formal, reducing trust.	GIVAS used NLP, multilingual capabilities, and multimodal interfaces to create more diverse and versatile learning environments.
RQ3: Challenges of GIVAS	- Human interaction gaps - Speech recognition challenges - Potential bias	Virtual agents lacked empathy, natural dialog, and human unpredictability. Trust in GIVAS was affected by perceived unfairness, hallucinations, and a lack of transparency.	AI hallucinations, static retrieval, and unexplainability hindered reliability; speech-to-text was still inconsistent.	Virtual agents could not fully replicate human interactions. Virtual agents also faced issues with accented/rapid speech and may exhibit biases.

Note: The Situation Analysis ¹ refers to the situation within an institutional and social context, and Situation Analysis ² refers to the situation within a technological context (as noted in Table 4).

5.1. Strengthening Health Profession Education and Clinical Skills (RQ1)

Twelve out of 16 studies (Table 2, Category C1 and C2) demonstrated that the use of GIVAS in the health sector presents significant implications for health profession education. Studies emphasized the role of AI in enhancing participants' communication (e.g., patient engagement and history-taking) by having consistent, adaptive feedback and safe spaces for practice (Mool et al., 2024; Yan & Alterovitz, 2024). AI-driven platforms can promote authentic learning environments that replicate real-world clinical complexity, often through immersive simulations and AI-driven storytelling, which help contextualize learning in emotionally and ethically nuanced scenarios (Mittenentzwei et al., 2024). These environments not only foster clinical reasoning but also support the development of empathy and emotional communication skills, which are critical components of professional identity formation. Furthermore, AI systems enable tailored educational experiences by adapting content and pacing to individual learner needs, thus providing alternative and flexible pathways within health education that can accommodate diverse backgrounds, learning preferences, and career trajectories (Chu & Goodell, 2024).

In addition, six out of 16 studies suggested that the use of GIVAS can potentially improve clinical skills by enhancing diagnostic accuracy and the relevance of treatment recommendations (Ba et al., 2024; Lv et al., 2025). By providing real-time data analysis and support for decision-making, AI systems contribute to greater clinical decision-making efficiency, allowing practitioners to make informed choices more quickly and confidently (Samala & Rawas, 2024). Moreover, several studies highlight the role of AI in supporting clinical skill acquisition, offering learners structured, repeatable, and feedback-rich environments to develop procedural and cognitive competencies (Mool et al., 2024). As automation becomes more reliable, AI tools also assist in standardizing routine clinical tasks, enabling healthcare professionals to focus more on complex decision-making and patient-centered care (Sardesai et al., 2024).

5.1.1. Enhanced Engagement and Human-like Interaction (Table 3, Category C3)

Seven out of the 16 papers reported on the findings of enhanced engagement and human-like interaction. Studies indicated that GIVAS have significant potential to enhance both patient and student engagement in health profession education (Mool et al., 2024; Yan & Alterovitz, 2024). GIVAS with high-fidelity can improve training in doctor–patient

communication and support users from diverse cultural backgrounds in multilingual interactions/translations (Badawy et al., 2025). Virtual agents can simulate human-like interactions and contribute to the relational dynamics between users and AI systems (Yan & Alterovitz, 2024). For example, the study by Yan and Alterovitz (2024) designed personality traits (e.g., empathy, humor, and trustworthiness) for virtual agents. It allowed users to choose the medical specialty and personal characteristics of the AI doctor based on their preferences (Table 2, Category C6). This made the user feel like they were interacting with a “doctor with a personality” rather than a cold machine. As a result, users began referring to the virtual agent as “he” or “she” instead of “it,” which reflected a shift toward perceiving the AI as a human-like entity that had gained the user’s trust (Yan & Alterovitz, 2024).

Furthermore, medical students can interact with virtual patients repeatedly to role-play and practice communication skills with patients, family members, and doctors, without the fear of making mistakes or facing real-life consequences. Systems can provide conversation summaries and performance checklists to help students receive immediate feedback and improve their communication abilities (Mool et al., 2024). This indicated that GIVAS can generate transparent and understandable AI behaviors that strengthened users’ confidence and acceptance of these technologies. GIVAS also enhanced emotional engagement and built trust, particularly in mental health contexts, and when developing clinical judgment (Greca et al., 2024). For example, in VR settings that had realistic and multisensory stimuli (e.g., visual, auditory, and tactile dimensions), interaction with virtual agents enhanced patients’ immersion and sense of presence. These highly personalized dynamic stimuli (e.g., realistic fear-inducing objects used in exposure therapy) elicited authentic psychological and physiological responses, thereby increasing emotional engagement in therapy (Greca et al., 2024). For students’ education, GIVAS provided realistic training scenarios (Table 2, Category C6), especially in VR environments, where realistic and adaptable patient simulations help develop history-taking and storytelling abilities. These tools specifically enhanced engagement by delivering rich, quality, and data-driven narratives (Mittenentzwei et al., 2024). However, in the study of Lv et al. (2025), two different groups of participants’ interactive skills (e.g., maintaining relationships and facial expressions) did not present significant differences.

5.1.2. Communication Skills in Simulations (Table 3, Category C4)

Unlike most of the literature stating that virtual patients powered by GenAI have the potential to provide interactive and realistic scenarios for medical students to practice clinical skills, 10 out of the 16 included studies were more multi-dimensional and critical. For example, Gutiérrez Maquilón et al. (2024) indicated that using ElevenLabs AI Voice Tools can simulate realistic human voices whilst mimicking sounds such as suffering, breathing difficulties, and groans. In their study, participants rated the AI-generated virtual patient at a moderately positive level, yet the latency (3 s) of communicating with AI reduced the interactivity and user experience (Gutiérrez Maquilón et al., 2024). If latency is too high, users may become impatient or perceive the system as “choppy”. In the study of Contreras et al. (2024), learners believed that AI-generated agents and videos were high-quality, engaging, and well-paced for learning, whereas deficits have been found in facial expressions, monotonous voice tones, and restricted body language. AI avatars still lack full naturalness, and this may affect users’ experience (Contreras et al., 2024).

Mool et al. (2024) identified that whilst it is an accepted norm that human beings are not perfect in their skills of language, tone, expression, and body language, an ideal human representation is conversely expected from a simulated human role. However, the authors did not elaborate on what precisely constitutes this ideal representation within the study. Yan and Alterovitz (2024) revealed that virtual agents in health have been

increasingly used to support communication with empathy and personalization, e.g., using NLP to analyze users' queries and respond to them in a human-like manner (Table 2, Category C6). Other studies explored the challenge of bridging human–AI communication gaps and aimed to create more seamless and intuitive interactions between users and AI systems (Chu & Goodell, 2024; Greca et al., 2024). Their key focus was on improving time coordination in dialogs, such as enhancing turn-taking, minimizing response delays, and aligning conversational pacing with human expectations (Lv et al., 2025; Mool et al., 2024; Sardesai et al., 2024). These improvements contribute to smoother interactions and a more natural user experience.

5.1.3. Personalized Learning and Tailored Learning Environments (Table 3, Category C5)

Six out of 16 studies indicated that virtual agents could adapt to the individual needs of learners, offering tailored content and feedback based on their performance and progress within realistic scenarios (Sardesai et al., 2024). For instance, Sevgi et al. (2024) indicated that virtual agents had the flexibility to help learners by answering their questions. Learners find AI-generated educational content, such as virtual instructor videos, to be engaging and effective. For example, learners rated AI-generated videos as high-quality and well-paced for learning, with many not even realizing the content was AI-generated (Contreras et al., 2024). The presence of virtual instructors and the *interactive nature of GIVAS* can positively impact learners' motivation and attitudes, and further lead to better engagement and learning outcomes. At the same time, AI-powered systems with virtual agents can offer efficient and cost-effective solutions for producing educational content at scale (Contreras et al., 2024), facilitating rapid development of diverse instructional materials and simulations.

However, issues were identified regarding complexity in the delivery of medical terminology. For instance, certain translated phrases lacked naturalness or the language was overly formal, and some awkward phrasing disrupted the fluidity of spoken communication (Badawy et al., 2025). In health profession education, GenAI-powered personalized learning is not only about “flexible teaching.” It requires accurate knowledge to train students' professional competencies, and it is crucial because it directly impacts the quality of future communication between students and patients, and may even affect clinical safety.

5.2. Technical Features and Functionalities (RQ2)

Most studies ($n = 16$) found that GIVAS leveraged *advanced NLP technology* to enable diverse conversational abilities (Badawy et al., 2025; Greca et al., 2024). In doing this, virtual agents can understand and respond to natural language queries, simulate nuanced clinical dialogs, and adapt their responses based on the learner's input and contextual cues. Also, GIVAS uses natural language processing and context-aware communication to enable realistic, interactive patient simulations that promote learners' personal and team work (Ba et al., 2024; Greca et al., 2024). For example, GIVAS improved students' clinical communication skills in problem-based learning settings while supporting scalable, personalized learning (Mool et al., 2024).

Additionally, *AI agents can offer multilingual support* (i.e., by using GenAI's functionality in translation and interpretation), which supported communication in multiple languages to cater to users' diverse backgrounds (Badawy et al., 2025). This combination of features enhanced accessibility, engagement, and the overall learning experience for students and healthcare professionals (Chen et al., 2025). GIVAS utilize *multimodal interaction to diversify learning experiences*. In the studies reviewed, there were both voice- and text-based interactions for users to choose from, which indicated flexibility for learners based on their preferences or situational needs (Gutiérrez Maquilón et al., 2024; Mool et al.,

2024). To enhance the understanding of complex medical concepts in health, these GIVAS incorporate visual aids such as images, diagrams, and videos, making abstract or intricate topics more accessible (Badawy et al., 2025). Furthermore, some GIVAS were integrated with XR technologies to provide highly immersive experiences, such as virtual surgeries or detailed exploration, which can allow learners to practice and visualize medical procedures in a safe and controlled environment (Greca et al., 2024; Lv et al., 2025). This multimodal approach enriches the educational experience by catering to diverse learning styles and fostering deeper engagement.

Three out of 16 studies used *GenAI to create virtual patients* that simulated realistic patient–clinician interactions, which could be scalable as *low-cost educational tools* (e.g., using the OpenAI API) (Abi-Rafeh et al., 2023; Mool et al., 2024; Sardesai et al., 2024). Simulated teaching environments and standardized patients used to replicate real-life clinical scenarios are often costly; hence, these GIVAS are a cost-effective alternative for scalable clinical training and can be customized to provide feedback on clinician performance. In addition, VR environments with GIVAS can support interactive learning, particularly in complex subjects such as general healthcare, diagnostics, and anatomy, by providing immersive, three-dimensional simulations that allow learners to engage without risk (Chheang et al., 2024; Lv et al., 2025). These systems enable students to visualize internal body structures, perform virtual diagnostic procedures, and interact with virtual patients in realistic clinical scenarios. This approach can enhance spatial understanding, procedural knowledge, and clinical reasoning.

Moreover, these systems facilitate verbal communication and *adapt to different cognitive complexities*, enhancing engagement and understanding. Tools like ChatGPT, Google Cloud’s Dialogflow, HeyGen, Synthesia, Unity, Stable Diffusion, and DeepMotion were employed to develop GIVAS with added quizzes, educational content, and further curriculum design in student assessments (Chheang et al., 2024; Contreras et al., 2024). GIVAS were used as *a tutoring system* that provided dynamic scenario generation, real-time adaptability, and personalized instruction, addressing challenges like scalability and integration of domain-specific knowledge. These models can also assist in summarizing the literature and generating new ideas for assignments (Gutiérrez Maquilón et al., 2024; Lv et al., 2025).

Other studies utilized 2D avatars *to simulate patient encounters*. AI models, such as GPT-3.5-turbo, GPT-4.0, GPT-4o, and other LLMs, empowered GIVAS with dynamic and context-aware interactions. However, the choice of AI platform significantly influences the quality, depth, and adaptability of the simulations (Gutiérrez Maquilón et al., 2024), as different platforms vary in their capacity for customization and accessibility. For instance, in the study of Mittenentzwei et al. (2024), the virtual agent helped educators and learners tailor the content to specific medical specialties, scenarios, and learning objectives. This adaptability makes the tools more inclusive and user-friendly, which can cater to a wide range of learners with varying levels of expertise. However, GIVAS need to ensure the accuracy of medical content, address ethical concerns related to AI use in healthcare education, and mitigate potential biases in AI-generated materials (Lv et al., 2025).

5.3. Challenges of GIVAS (RQ3)

Although many research findings appeared positive and promising, several limitations and challenges were also identified. For example, although virtual agents can offer human-like responses and use text-to-speech for realism, they *cannot fully replicate real human interactions*, and their answers can be longer and overly formal (Sardesai et al., 2024). Conversations with AI virtual agents were affected by a limited understanding of vocabulary, faulty voice recognition, poor error handling, repetitive interactions, and lack of conversational diversity and accuracy (Gutiérrez Maquilón et al., 2024). Also, AI-generated

content may include *potential biases*, misleading responses, and a lack of transparency regarding data sources (Chheang et al., 2024). Because LLMs operate like “black boxes” (often referred to as the unexplainability and incomprehensibility of AI systems), it is difficult to trace responses (Byrnes & Robinson, 2024). LLMs may also produce incorrect outputs (hallucinations), yet uploading customized content and designing user warnings can mitigate this problem (Sardesai et al., 2024). On the other hand, issues with speech-to-text functionality (e.g., *difficulty understanding accents or fast speech*) also affected user experience and correctness scores (Chheang et al., 2024).

The characters assigned to AI agents (avatars) can ensure consistency for interactions, and introducing some unpredictability mimics real patient behavior, as well as enhancing reusability (Sardesai et al., 2024). In terms of virtual agents in video format, the quality and accuracy of *input content are critical* for the effectiveness of AI-generated video. Poor data can undermine educational goals despite strong visuals. Furthermore, the lack of a full preview mode requires repeated video generation that can increase wait times (Contreras et al., 2024). Issues such as *response delays* and *lack of naturalness in interactions* in AI-generated content can affect the learning experience.

Inaccurate interpretations or incomplete retrieval of guideline information by GIVAS can result in severe consequences (Zhang & Zhang, 2023). For example, the correctness of clinical decisions made by a GIVAS can vary based on document structure and content placement (Sevgi et al., 2024). Another example is that errors made by a virtual agent like EyeTeacher could reinforce incorrect knowledge in learners, and it may similarly then fail to address participants’ weaknesses when providing feedback. Associated risks were also identified with EyeAssistant, specifically automation bias, where users placed excessive trust in AI-generated responses and reduced their own critical judgment (Sevgi et al., 2024). For some GenAI LLMs (e.g., GPT-3.5), limitations include a *restricted number of document uploads*, which narrows the scope of information available to be summarized, and reference to static content retrieval that may miss the latest research developments (Abi-Rafeh et al., 2023).

Furthermore, *data privacy and security* were major concerns (Badawy et al., 2025; Chu & Goodell, 2024; Sevgi et al., 2024). Some studies employed OpenAI and HeyGen resources that lacked a full understanding of how data was used, stored, or shared, leading to issues of transparency and consent (Contreras et al., 2024). Similarly, studies indicate that the virtual agent itself is not a legal entity and is not accountable; it remains unclear who is responsible for errors or harm caused by virtual agents (Abi-Rafeh et al., 2023; Contreras et al., 2024). Notably, there is no universal framework for evaluating the safety, efficacy, and ethical use of GIVAS in healthcare and related education. Existing laws may not adequately address the unique challenges posed by GIVAS.

6. Discussion

The results revealed that GIVAS are increasingly used in the health profession education sector, and they have the potential to improve the experience of learning and training. GIVAS appear to mediate health profession education through the following pedagogical mechanisms: (1) experiential rehearsal in safe simulation environments, (2) scaffolded dialogic feedback that supports progressive competence development, and (3) enhanced social presence that fosters professional identity formation.

6.1. Theoretical Implications

The above mechanisms can be further related to established learning theories that many of the reviewed studies have not noted. Firstly, the use of simulated, interactive environments aligns closely with experiential learning theory (Kolb, 1984), where learners experience knowledge being transferred by experience through the “Socratic” conversa-

tion with GIVAS. For example, because the “patient” is a virtual AI agent, learners can immediately retry the clinical scenario with a different approach to see how the outcome changes. Secondly, the research findings broadly align with cognitive load theory, as GIVAS may help reduce or add cognitive load depending on the design and functionality of the GIVAS. For example, the use of GIVAS as a tutor or guide to provide learners support in understanding clinical documents and by generating scenarios can reduce cognitive load. However, research shows that instructional techniques that help novices (e.g., step-by-step guidance) can actually hinder experts, which is called the expertise reversal effect (Kalyuga, 2007). Reliance on AI-generated responses may also limit opportunities for deeper cognitive processing and independent reasoning, particularly if learners adopt a passive interaction style. The foundations of social constructivism also reveal that learners may get used to the social roles of these virtual agents (i.e., virtual patients), even if they know they are chatbots, which depends on the extent to which users perceive GIVAS in their interactions as “real social individuals” (Vygotsky & Cole, 1978).

In health education, if the humanization of virtual agents was developed sufficiently (e.g., by developing humanistic voice tone, facial expressions, and movements), students would practice active learning during engagement with the agent that could enhance their diagnostic reasoning and clinical communication skills. Moreover, social presence theory further reinforces that multimodal methods such as voice, facial expressions, and gestures can significantly improve users’ trust and willingness to engage with virtual agents (Lowenthal, 2010; Ochs et al., 2022). Multimodal communication methods (e.g., such as the expression of semantics, voice, facial emotions, and gestures) can enhance the perceived presence of GIVAS. In the context of health profession education, the quality of interaction between learners or patients and virtual agents directly influences learning outcomes and behavior change. Users are more likely to trust a medical AI that can have a natural conversation (high social presence) rather than a mechanical question-answering system (low social presence) (Simbach, 2024). This, in turn, can increase users’ trust and willingness to engage. Such increased engagement is particularly valuable in the education domain, as it supports more active participation in health management, greater receptiveness to educational content, and improved communication in patient care.

6.2. Technological Implications

The technological development of GIVAS implies that the existing GenAI models have made certain breakthroughs in perception capabilities, memory architecture, ethical design, and cross-environment adaptability. Newer advancements in audio generation, such as WaveNet, VALL-E, and AudioLM, have enabled virtual agents to synthesize realistic speech, music, and sound effects, allowing for richer, multimodal interactions. For instance, virtual patients capable of expressing emotions like anxiety, anger, or pain create authentic scenarios for learners to practice professional communication skills. Moreover, AI-driven clinical reasoning simulations can present dynamically complex cases. Future breakthroughs in GIVAS may include integration of voice, vision, haptic, and even olfactory sensors to achieve holographic interaction (AlShaghroud et al., 2023). However, this is an emerging technology with limited functionality in complex modalities (e.g., a combination of audio, video, and images), and seamlessly integrating all these modalities into a responsive, real-time virtual agent is still technically and computationally demanding (Soliman et al., 2024). Furthermore, challenges such as ensuring data privacy, maintaining content accuracy, and addressing ethical concerns remain significant barriers to broader adoption (Bouderhem, 2024). This study contributes to emerging research on GIVAS as health educational tools in simulation-based education by summarizing relevant implications and addressing gaps in ethical and practical implementation.

6.3. Educational Implications

GIVAS in health education has the potential to help improve teaching efficiency and systematic standardization (Xu et al., 2025). A significant finding across the reviewed studies (e.g., Ba et al., 2024; Sardesai et al., 2024; Mool et al., 2024) is the high level of student-reported satisfaction and self-efficacy after having training with GIVAS (standardly designed training). However, we identify a critical evidence gap: there is a lack of objective clinical accuracy data to cross-reference these subjective feelings. This raises the risk of overconfidence bias, where the helpful and cooperative nature of GenAI may mislead students into a fake sense of readiness. But in the field of health and medicine, the essence of much education and training is in facing complex and dynamic individuals, who require experience, intuition, and ethical judgment. Findings did not disclose enough of GIVAS's assistance on ethics reflection and theoretical knowledge acquisition. GIVAS may facilitate the acquisition of foundational knowledge and standardized clinical skills (Srinivasa et al., 2022), and yet it might be difficult to safeguard learners' authentic ethical perceptions and self-efficacy, and further, prevent the 'mechanization' of human empathy during these AI interactions.

6.4. Limitations

Many studies and developments of GIVAS presented their efforts on pursuing certain scales of LLM input, rather than pursuing accuracy and credibility. The application of GIVAS in health profession education is in the emerging stages: Most of the reviewed studies can be classified as pilot studies or development-focused projects with aims focused on examining the technical feasibility or conceptual promise of GenAI systems and virtual agents. Many studies lack theoretical frameworks, rigorous evaluation, longitudinal follow-up, and stakeholder involvement. When using GIVAS in health profession education, several studies claimed alignment with curricular goals, yet only a minority demonstrated integration into existing medical education frameworks. For educators and students, most GIVAS available are still in text form (Ba et al., 2024; Sevgi et al., 2024). Whilst a small number of experiments have developed multimodal GIVAS with avatar images embedded in virtual reality environments, most of their development is currently in the pilot study stage (Ochs et al., 2022). Some of the virtual agents seem to have the ability to interact emotionally with users, yet these are not reported as being satisfactory for the educational and training needs of the users. The main problems experienced are the struggle of virtual agents dealing with complex queries and the potential for biased or inaccurate responses, as well as ethical concerns and data privacy issues (Sai et al., 2024).

In this review, although the findings suggest that GIVAS hold considerable promise for enhancing learning and training in health profession education, the current evidence base remains preliminary. Many of the included studies are small-scale, pilot, or development-focused, with limited longitudinal evaluation. The geographical distribution of studies is relatively narrow (most studies are from developed countries), with evidence concentrated in specific regions and educational contexts. This limits the generalizability of the findings across diverse healthcare systems and learner populations.

7. Conclusions

This systematic review examined current empirical studies ($n = 16$) of generative intelligent virtual agents and simulations (GIVAS) in health profession education. GIVAS use advanced natural language processing (NLP) and multimodal interaction technologies, transcending the deterministic, rule-based logic of traditional simulations, introducing a probabilistic and conversational paradigm that reflects the predictability of real clinical practice, which is beneficial for medical education. This study contributes a pedagogy-oriented synthesis that explicates the learning mechanisms through which GIVAS supports education,

linking agent roles, educational designs, and evidence types across studies. The findings suggest that GIVAS exhibit promising potential to enhance clinical skills, engagement, communication, and personalized learning. However, challenges such as human interaction gaps, speech recognition issues, and potential biases highlight the need for ongoing ethical and technical advancements to ensure effective, inclusive, and trustworthy implementation.

Author Contributions: Conceptualization, A.O., M.S.K. and A.H.; Methodology, M.S.K. and X.W.; Validation, X.W., A.O., M.S.K. and A.H.; Formal analysis, X.W. and A.O.; Investigation, X.W., A.O., M.S.K. and A.H.; Writing—original draft preparation, X.W.; Writing—review and editing, X.W., A.O., M.S.K. and A.H. All authors have read and agreed to the published version of the manuscript.

Funding: Owen Silver Legacy—administered by School of Medicine, University of St Andrews; Danish Data Science Academy Visit Grant.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Abi-Rafeh, J., Hanna, S., Bassiri-Tehrani, B., Kazan, R., & Nahai, F. (2023). Complications following facelift and neck lift: Implementation and assessment of large language model and artificial intelligence (ChatGPT) performance across 16 simulated patient presentations. *Aesthetic Plastic Surgery*, 47(6), 2407–2414. [CrossRef] [PubMed]
- AlShaghrouh, S., AlShuwaier, A., & AlRakaf, L. (2023, July 23–28). *Artificially intelligent and interactive 3D hologram*. International Conference on Human-Computer Interaction, Copenhagen, Denmark.
- Ba, H., Zhang, L., & Yi, Z. (2024). Enhancing clinical skills in pediatric trainees: A comparative study of ChatGPT-assisted and traditional teaching methods. *BMC Medical Education*, 24(1), 558. [CrossRef] [PubMed]
- Badawy, M. K., Kharmwan, K., & Carrion, D. (2025). A pilot study of generative AI video for patient communication in radiology and nuclear medicine. *Health and Technology*, 15(2), 395–404. [CrossRef]
- Bouderhem, R. (2024). Shaping the future of AI in healthcare through ethics and governance. *Humanities and Social Sciences Communications*, 11(1), 416. [CrossRef]
- Bozkurt, A. (2023). Unleashing the potential of generative AI, conversational agents and chatbots in educational praxis: A systematic review and bibliometric analysis of GenAI in education. *Open Praxis*, 15(4), 261–270. [CrossRef]
- Byrnes, J., & Robinson, M. (2024). Transparency and authority concerns with using AI to make ethical recommendations in clinical settings. *Nursing Ethics*, 32(6), 1749–1760. [CrossRef] [PubMed]
- Casheekar, A., Lahiri, A., Rath, K., Prabhakar, K. S., & Srinivasan, K. (2024). A contemporary review on chatbots, AI-powered virtual conversational agents, ChatGPT: Applications, open challenges and future research directions. *Computer Science Review*, 52, 100632. [CrossRef]
- Chen, S., Cheng, H., Su, S., Patterson, S., Kushalnagar, R., Huang, Y., & Wang, Q. (2025). Customizing generated signs and voices of AI avatars: Deaf-centric mixed-reality design for deaf-hearing communication. *Proceedings of the ACM on Human-Computer Interaction*, 9(2), CSCW055. [CrossRef]
- Chheang, V., Sharmin, S., Márquez-Hernández, R., Patel, M., Rajasekaran, D., Caulfield, G., Kiafar, B., Li, J., Kullu, P., & Barmaki, R. L. (2024). Towards anatomy education with generative AI-based virtual assistants in immersive virtual reality environments. In *2024 IEEE international conference on artificial intelligence and extended and virtual reality (AIxVR)*. IEEE.
- Chu, S. N., & Goodell, A. J. (2024). Synthetic patients: Simulating difficult conversations with multimodal generative AI for medical education. *arXiv*, arXiv:2405.19941. [CrossRef]
- Clarke, A. (2005). *Situational analysis*. SAGE Publications, Inc. [CrossRef]
- Contreras, I., Hossfeld, S., de Boer, K., Wiedler, J. T., & Ghidinelli, M. (2024). Revolutionising faculty development and continuing medical education through AI-generated videos. *Journal of CME*, 13(1), 2434322. [CrossRef] [PubMed]
- Diaz-Navarro, C., Armstrong, R., Charnetski, M., Freeman, K. J., Koh, S., Reedy, G., Smitten, J., Ingrassia, P. L., Matos, F. M., & Issenberg, B. (2024). Global consensus statement on simulation-based practice in healthcare. *Advances in Simulation*, 9(1), 19. [CrossRef] [PubMed]
- Elendu, C., Amaechi, D. C., Okatta, A. U., Amaechi, E. C., Elendu, T. C., Ezech, C. P., & Elendu, I. D. (2024). The impact of simulation-based training in medical education: A review. *Medicine*, 103(27), e38813. Available online: https://journals.lww.com/md-journal/fulltext/2024/07050/the_impact_of_simulation_based_training_in_medical.22.aspx (accessed on 25 January 2026). [CrossRef] [PubMed]

- Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines*, 30(4), 681–694. [CrossRef]
- Greca, A. D., Amaro, I., Barra, P., Rosapepe, E., & Tortora, G. (2024). Enhancing therapeutic engagement in mental health through virtual reality and generative AI: A co-creation approach to trust building. In *2024 IEEE international conference on bioinformatics and biomedicine (BIBM)*. IEEE. [CrossRef]
- Gutiérrez Maquilón, R., Uhl, J., Schrom-Feiertag, H., & Tscheligi, M. (2024). Integrating GPT-based AI into virtual patients to facilitate communication training among medical first responders: Usability study of mixed reality simulation. *JMIR Formative Research*, 8, e58623. [CrossRef] [PubMed]
- Hirzle, T., Müller, F., Draxler, F., Schmitz, M., Knierim, P., & Hornbæk, K. (2023). When XR and AI meet—A scoping review on extended reality and artificial intelligence. In *Proceedings of the 2023 CHI conference on human factors in computing systems*. Association for Computing Machinery. [CrossRef]
- Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review*, 19(4), 509–539. [CrossRef]
- Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development*. FT Press.
- Kononowicz, A. A., Woodham, L. A., Edelbring, S., Stathakourou, N., Davies, D., Saxena, N., Tudor Car, L., Carlstedt-Duke, J., Car, J., & 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. [CrossRef] [PubMed]
- Krishnan, N. (2025). AI agents: Evolution, architecture, and real-world applications. *arXiv*, arXiv:2503.12687.
- Lowenthal, P. R. (2010). Social presence. In *Social computing: Concepts, methodologies, tools, and applications* (pp. 129–136). IGI Global.
- Lv, J., Slowik, A., Rani, S., Kim, B.-G., Chen, C.-M., Kumari, S., Li, K., Lyu, X., & Jiang, H. (2025). Multimodal metaverse healthcare: A collaborative representation and adaptive fusion approach for generative AI-driven diagnosis. *Research*, 8, 0616. [CrossRef] [PubMed]
- Methley, A. M., Campbell, S., Chew-Graham, C., McNally, R., & Cheraghi-Sohi, S. (2014). PICO, PICOS and SPIDER: A comparison study of specificity and sensitivity in three search tools for qualitative systematic reviews. *BMC Health Services Research*, 14, 579. [CrossRef] [PubMed]
- Mittenentzwei, S., Garrison, L. A., Budich, B., Lawonn, K., Dockhorn, A., Preim, B., & Meuschke, M. (2024). *AI-assisted character design in medical storytelling with stable diffusion* (SSRN 4772811). SSRN. [CrossRef]
- Mool, A., Schmid, J., Johnston, T., Thomas, W., Fenner, E., Lu, K., Gandhi, R., Western, A., Seabold, B., & Smith, K. (2024). Using generative AI to simulate patient history-taking in a problem-based learning tutorial: A mixed-methods study. *Technology, Knowledge and Learning*. [CrossRef]
- Ochs, M., Bousquet, J., Pergandi, J.-M., & Blache, P. (2022). Multimodal behavioral cues analysis of the sense of presence and social presence during a social interaction with a virtual patient. *Frontiers in Computer Science*, 4, 746804. [CrossRef]
- Omigraliyev, R., Kenzhe, D., & Mirambekov, S. (2024). Simulating life: The application of generative agents in virtual environments. In *2024 IEEE AITU: Digital generation*. IEEE.
- Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., & Ray, A. (2022). Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 35, 27730–27744. [CrossRef]
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., . . . Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. [CrossRef] [PubMed]
- Pham, T. D., Karunaratne, N., Exintaris, B., Liu, D., Lay, T., Yuriev, E., & Lim, A. (2025). The impact of generative AI on health professional education: A systematic review in the context of student learning. *Medical Education*, 59(12), 1280–1289. [CrossRef] [PubMed]
- Preiksaitis, C., & Rose, C. (2023). Opportunities, challenges, and future directions of generative artificial intelligence in medical education: Scoping review. *JMIR Medical Education*, 9, e48785. [CrossRef] [PubMed]
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). *Language models are unsupervised multitask learners*. Available online: https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf (accessed on 6 January 2026).
- Rosen, N. L., & Shoenberger, N. A. (2021). “Words speak louder than actions”: The connection between gendered language and bullying behavior. *Open Journal of Social Sciences*, 9(8), 197–214. [CrossRef]
- Rossi, M., Ciletti, M., Melchiorre, L., & Toto, G. A. (2024). The impact of generative artificial intelligence (GenAI) on education: A review of the potential, the risks and the role of immersive technologies. *Education Sciences & Society-Open Access*, 15(2), 400–415. [CrossRef]
- Sai, S., Gaur, A., Sai, R., Chamola, V., Guizani, M., & Rodrigues, J. J. P. C. (2024). Generative AI for transformative healthcare: A comprehensive study of emerging models, applications, case studies, and limitations. *IEEE Access*, 12, 31078–31106. [CrossRef]
- Samala, A. D., & Rawas, S. (2024). Generative AI as virtual healthcare assistant for enhancing patient care quality. *International Journal of Online & Biomedical Engineering*, 20(5), 174–187. [CrossRef]
- Sardesai, N., Russo, P., Martin, J., & Sardesai, A. (2024). Utilizing generative conversational artificial intelligence to create simulated patient encounters: A pilot study for anaesthesia training. *Postgraduate Medical Journal*, 100(1182), 237–241. [CrossRef] [PubMed]

- Sevdalis, N. (2015). Simulation and learning in healthcare: Moving the field forward. *BMJ Simulation and Technology Enhanced Learning*, 1(1), 1. [\[CrossRef\]](#) [\[PubMed\]](#)
- Sevgi, M., Antaki, F., & Keane, P. A. (2024). Medical education with large language models in ophthalmology: Custom instructions and enhanced retrieval capabilities. *British Journal of Ophthalmology*, 108(10), 1354–1361. [\[CrossRef\]](#) [\[PubMed\]](#)
- Siala, H., & Wang, Y. (2022). SHIFTing artificial intelligence to be responsible in healthcare: A systematic review. *Social Science & Medicine*, 296, 114782. [\[CrossRef\]](#) [\[PubMed\]](#)
- Simbach, L. V. (2024). *Designing for a better user experience: The effects of visual appearance, gender, and context on perceptions of trust, perceived ease of use, empathy, customer satisfaction, competence, social presence, and intention to use*. University of Twente.
- So, H. Y., Chen, P. P., Wong, G. K. C., & Chan, T. T. N. (2019). Simulation in medical education. *Journal of the Royal College of Physicians of Edinburgh*, 49(1), 52–57. [\[CrossRef\]](#) [\[PubMed\]](#)
- Soliman, M. M., Ahmed, E., Darwish, A., & Hassanien, A. E. (2024). Artificial intelligence powered metaverse: Analysis, challenges and future perspectives. *Artificial Intelligence Review*, 57(2), 36. [\[CrossRef\]](#)
- Srinivasa, K., Kurni, M., & Saritha, K. (2022). Harnessing the power of AI to education. In *Learning, teaching, and assessment methods for contemporary learners: Pedagogy for the digital generation* (pp. 311–342). Springer.
- Tavarnesi, G., Laus, A., Mazza, R., Ambrosini, L., Catenazzi, N., Vanini, S., & Tuggener, D. (2018). Learning with virtual patients in medical education. In *CEUR workshop proceedings*. SUPSI.
- Usheva, N., Bliznakova, K., & Grancharov, D. (2024). Barriers to the application of simulation technologies in the education of health care students. *European Journal of Public Health*, 34, ckae144-1889. [\[CrossRef\]](#)
- Vygotsky, L. S., & Cole, M. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Xu, Q., Wu, Y., Zheng, H., Yan, H., Wu, H., Qian, Y., Wu, Y., & Liu, B. (2025). Standardization in artificial general intelligence model for education. *Computer Standards & Interfaces*, 94, 104006. [\[CrossRef\]](#)
- Yan, N., & Alterovitz, G. (2024). A general-purpose AI avatar in healthcare. *arXiv*, arXiv:2401.12981. [\[CrossRef\]](#)
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—Where are the educators? *International Journal of Educational Technology in Higher Education*, 16(1), 39. [\[CrossRef\]](#)
- Zhang, J., & Zhang, Z.-M. (2023). Ethics and governance of trustworthy medical artificial intelligence. *BMC Medical Informatics and Decision Making*, 23(1), 7. [\[CrossRef\]](#) [\[PubMed\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.