Multimodal Interaction through Embodied Agents: An Intelligent Assistant for Immersive Virtual Environments*

Muhammad Saad[1,*], Muhammad Saeed[1], Fedwa Laamarti[1] and A.El Saddik[1]

[1]Multimedia Communications Research Laboratory (MCRLab), University of Ottawa, Ottawa, Ontario, Canada

Abstract

We present the design and implementation of ZapAura, a virtual reality web-based platform based on Hubs, designed to enhance industries such as education and healthcare through an AI-driven embodied agent in immersive environments. Our system introduces a knowledge-grounded embodied agent capable of engaging users through natural spoken dialogue, enriched with context-aware animations and emotionally aligned expressions. Furthermore, we deployed the pre-trained Llama 3.2 model within a Retrieval-Augmented Generation (RAG) setup using dataset collected from multiple domains such as computer vision, multimedia, haptics, and AI ethics. This design and implementation enable the virtual avatar to generate factually grounded responses based on real-time vector searches. The system supports both text and speech input. Body animations are rendered using a Mixamo and Blender pipeline and are dynamically mapped to conversational content, ensuring realistic interactions. We conducted a preliminary user study (n=8) evaluating the embodied agent's emotional expressiveness across three scenarios. Participants consistently rated the embodied agent as emotionally congruent and responsive, with positive rating across the three evaluation categories.

Keywords

Multimodal Interaction, Embodied Conversational Agent (ECA), Retrieval-Augmented Generation (RAG), Virtual Reality (VR), Metaverse

1. Introduction

The integration of Embodied conversational agents within virtual environments represents a transformative development in how users engage with digital content, particularly in immersive learning, remote collaboration, and metaverse applications. Web based Platforms such as Hubs [1] gain accessibility, extensibility, and cross-platform support; the demand for intelligent, responsive agents within these spaces becomes evident. Traditional virtual reality (VR) applications are not particularly rich in personalized interaction mechanisms, relying instead on pre-scripted content, static dialogs, or limited user-driven navigation models [2]. To address this gap, the recent development of intelligent agents capable of natural language processing, dialogue generation, and dynamic emotional expression has been enabled by advances in generative artificial intelligence (AI), avatar embodiment, and speech technologies [3, 4, 5].

Virtual assistants driven by large language models (LLMs) such as GPT [6], Gemini [7], Llama [8], Claude [9] and DeepSeek [10] have shown significant potential in adaptive and content-aware support across domains, including healthcare [3], education [11], and simulation [12]. However, these applications are often limited to desktop applications or text-based interfaces within proprietary ecosystems. In contrast, metaverse-based platforms, such as Hubs, offer a flexible and accessible framework for developing embodied AI agents. Still, the use of a fully interactive avatar, particularly capable of supporting real-time speech-based interaction and expressive animations remains under development in such environments.

Proceedings of the ACM CHI 2026 Workshop on Shaping Future Human Connections: Social Augmentation through XR Technologies, April 13, 2026, Barcelona, Spain

*This work was conducted at the Multimedia Communications Research Laboratory (MCRLab), University of Ottawa.

*Corresponding author.

msaad104@uottawa.ca (M. Saad); muhammadsaeedicp@gmail.com (M. Saeed); flaamart@uottawa.ca (F. Laamarti); elsaddik@uottawa.ca (A.El Saddik)

0009-0005-9846-8991 (M. Saad); 0009-0003-4462-9863 (M. Saeed); 0000-0002-0338-9264 (F. Laamarti); 0000-0002-7690-8547 (A.El Saddik)

To address these challenges, we present an extended and customized version of Hubs [1], named ZapAura, that integrates a real-time embodied conversational agent allowing users to engage in dynamic, speech-based interactions with AI avatars. The platform enables dynamic, speech-based interaction with an agent representing a domain expert that operates as an interactive entity within the virtual space. As it is not just limited to the person talking, the embodied agent in ZapAura can serve specific roles, such as a teacher avatar interacting with multiple student avatars, which are regular user avatars that do not contain AI capability [13] by embodied conversational agent, we refer to an AI-driven avatar that is not limited to visual representation alone, but participates directly in interaction through speech, facial, and body animation. Inspired by the recent work in anatomy education and personalized learning that demonstrates the efficiency of an embodied AI assistant in virtual spaces [13]. Our system supports a wide range of applications from intelligent NPCs and a virtual platform to personalized VR onboarding and collaborative teamwork.

The main contributions of this work are as follows:

- We present the design and implementation of an embodied conversational agent for immersive virtual environments, enabling real-time speech-driven interaction with synchronized facial and body animations within a social VR setting.
- We introduce a modular design approach that integrates a knowledge-grounded language model with expressive avatar embodiment to support multimodal interaction.
- We construct a domain-specific dataset from more than 370 PDF documents covering computer vision, haptics, multimedia, and AI ethics, and implement a RAG pipeline to ground agent responses in factual knowledge.

2. Background

2.1. Embodied Conversational Agents in Extended Reality

Virtual reality (VR) is becoming more lifelike, and a big reason for that is embodied conversational agents (ECAs). These are avatars that behave a lot like real people they talk, use body language, and show expressions. This makes them especially useful in education. In one study, Chheang et al. [11] used AI-powered avatars in a VR anatomy course and noticed that the students were more focused and performed better compared to those taught the usual way. Recent LLM advancements enable dynamic dialogue through platforms like Milo [14] for XR deployments, social VR agents with context-aware responses [15], and CUIfy [16] for embedding speech-based agents in XR. Embodied conversational agents (ECAs) have been shown to support learning by reducing cognitive load through embodied AR assistance [17] and by eliciting higher levels of social presence compared to text-only agents [18].

2.2. Retrieval-Augmented Generation for Educational Applications

Large language models face challenges with hallucinations and outdated information in educational contexts [3]. Retrieval-Augmented Generation (RAG) addresses this by grounding LLM outputs in domain-specific knowledge sources, improving factual accuracy and contextual relevance [19]. RAG systems retrieve relevant documents from external knowledge bases and incorporate them into generation, enabling dynamic knowledge updates without retraining. Recent Agentic RAG developments introduce autonomous decision-making capabilities for managing retrieval strategies and iteratively refining contextual understanding, while RAG architectures provide knowledge-domain guardrails that limit responses to desired domains, preventing off-topic or inaccurate outputs—particularly critical for conversational educational agents requiring both factual accuracy and natural dialogue [20].

2.3. Voice Synthesis and Multimodal Embodiment

The integration of voice synthesis technologies has become increasingly important for creating realistic and engaging virtual agents. Modern text-to-speech (TTS) systems and voice cloning technologies

enable the synthesis of human-like audio that conveys emotional nuances, speaker identity, and contextual prosody. In virtual reality and metaverse applications, voice cloning enhances immersion by allowing agents to communicate with natural, expressive speech patterns that can be personalized to match specific individuals.

Speech-driven animation systems synchronize avatar lip movements and facial expressions with synthesized speech, creating more believable interactions. Wav2Lip [21], a GAN-based model, has become a foundational approach for producing realistic lip-sync animations that closely match spoken words or sounds, enhancing the authenticity of virtual characters across various languages, accents, and speech patterns. Recent advances have improved upon these foundations, with systems achieving higher visual quality and more natural synchronization. Multi-modal embodiment systems that integrate visual, auditory, and behavioral cues have been shown to be particularly effective in educational and training applications, where realistic communication is essential for learner engagement.

2.4. Social Presence and Collaboration in Metaverse Platforms

Social presence, defined as moment-by-moment awareness of co-presence with another entity (social actor) accompanied by engagement [22], is critical for effective social VR applications. Embodied avatars foster greater social presence than non-embodied communication, with avatar realism—encompassing appearance and behavioral fidelity—significantly influencing co-presence and interpersonal connection [23]. Metaverse platforms like Hubs have enabled accessible virtual worlds for education and collaboration, with Yano [24] demonstrating web-based social VR's potential for distance and face-to-face teaching through student behavior analysis in virtual spaces.

Comparative studies reveal substantial variations in metaverse educational experiences. Puente-Fernández et al. [25] examined learning and classroom quality across Hubs, Meta Horizon Workrooms, Spatial, and Arthur, finding that while learning outcomes remained consistent, factors like avatars, spatial arrangement, mobility, and functionalities significantly influenced concentration, usability, presence, and learning. International collaborative education has also leveraged these platforms, with Hansen et al. [26] conducting an eight-week course across Finland, Norway, and the United States, comparing 2D video with 3D VR collaboration. Their findings showed full-body avatars on Spatial.io enabled more natural classroom-like interactions than Hubs, while emphasizing the need for detailed planning and strategies to reduce technology learning curves.

2.5. Ethical Considerations in AI Avatar Systems

The deployment of realistic AI avatars, particularly those that replicate specific individuals, raises important ethical considerations regarding consent, privacy, and potential misuse [2]. While powerful for creating engaging virtual experiences, voice cloning technologies require careful attention to authorization and consent protocols. Methuku and Myakala [27] also stressed that clear consent and strong data protection should be baked into any system that uses AI avatars. At ZapAura, this is a priority. We've implemented a secure, role-based access control system where only administrators have permission to modify the embodied agent configuration.

3. Methodology

This section presents the design and implementation of a realistic embodied conversational agent that functions as a domain expert in an immersive environment. Figure 1 shows the end-to-end processing pipeline from user input to synchronized audiovisual output. The system uses a modular architecture that separates input handling, knowledge retrieval, response generation, speech synthesis, and avatar animations, enabling runtime flexibility and efficient processing.

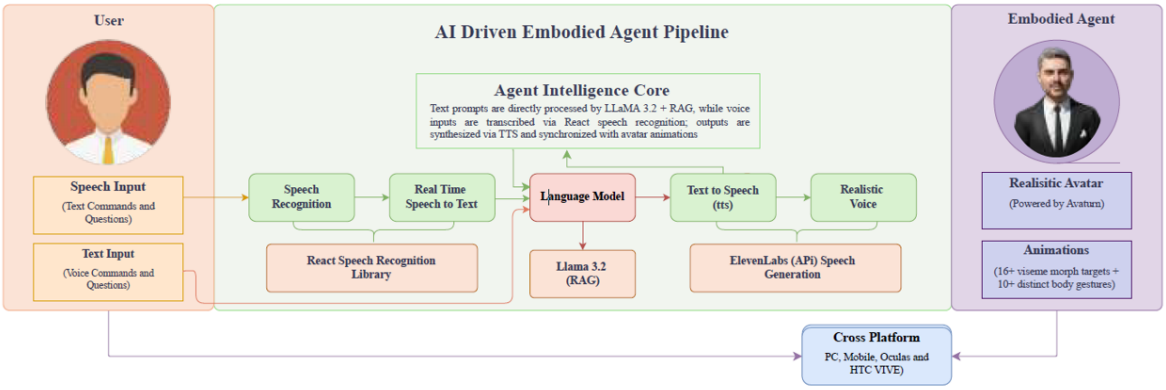

Figure 1: System architecture of the embodied agent in ZapAura. This diagram illustrates the bidirectional information flow between the user and agent. Input via voice or text is processed through React Speech Recognition and then semantically analyzed by a Llama 3.2 model with RAG. Response generation utilizes ElevenLabs for speech synthesis, while Rhubarb's viseme mapping and Mixamo animations create synchronized facial expressions and full-body movements. The architecture supports multimodal interaction across desktop and VR environments, enabling real-time, immersive communication between physical and virtual entities.

3.1. User Input Processing

3.1.1. Speech Input

To enable seamless and dynamic interaction between the user and the avatar, we integrate the open-source React Speech Recognition library, which provides a wrapper over the browser's native Web Speech API to convert speech to text. The system continuously listens to user speech and transcribes it into text in real time, enabling voice-based interaction without requiring typed input. The library supports multiple languages, can adapt to various accents, and maintains reasonable performance in noisy environments by leveraging a browser-based speech recognition engine. Once the speech input is transcribed, the resulting text is forwarded to downstream components such as the Llama 3.2 model and other dialog handling modules for further processing and response generation.

3.1.2. Text Input

In addition to speech input, the system supports direct text-based interaction through a dedicated user interface integrated into the ZapAura environment. This component includes a responsive text field that accepts keyboard input and submits queries via the Enter key or a clickable button. Such functionality addresses use cases where speech input may be impractical, e.g. in noisy settings or when users prefer text communication.

3.2. Knowledge Grounding with Retrieval-Augmented Generation

The agent core implements a retrieval-augmented generation (RAG) pipeline combining neural retrieval with LLM generation to produce factually grounded responses, addressing hallucination issues in standalone LLM systems. We construct a domain-specific dataset from more than 370 PDF documents (approximately 2.3 million tokens) that cover research papers, lecture notes, and teaching materials as primary sources, supplemented by foundational books and key papers in computer vision, haptics, multimedia, and ethical AI. Documents are preprocessed and segmented into 512-token chunks with 128-token overlap, preserving contextual information across boundaries. Each chunk maintains the source metadata for citation traceability and source type classification, yielding 8,947 indexed chunks in a FAISS vector database.

At query time, all-MiniLM-L6-v2 generates 384-dimensional embeddings for queries q and chunks d with 15 ms average latency. Cosine similarity matches the query against indexed chunks:

$$\text{sim}(q, d) = \frac{\mathbf{q} \cdot \mathbf{d}}{\|\mathbf{q}\| \, \|\mathbf{d}\|} \tag{1}$$

The top-5 chunks are retrieved with a minimum similarity threshold of 0.65. Retrieved chunks are concatenated by similarity and provided to Llama 3.2 (3B parameters) with instructions to ground responses in evidence, cite sources, and avoid speculation. The model generates response sequence $\mathbf{y} = (y_1, y_2, \ldots, y_T)$ where:

$$P(\mathbf{y} \mid \mathbf{x}, d) = \prod_{t=1}^{T} P(y_t \mid y_{<t}, \mathbf{x}, d) \tag{2}$$

We configure temperature 0.5 for balanced diversity and factual consistency with 256-token maximum length. Generated responses include inline citations and are forwarded to speech synthesis and avatar animation for synchronized multimodal delivery. When retrieved contexts fall below the similarity threshold, the model produces a conservative fallback response.

3.3. Speech Processing

Once our fine-tuned Llama model generates a response in text form, the system proceeds to convert the text to speech using ElevenLabs, a high-fidelity text-to-speech service. It offers API-based access to advanced speech synthesis models capable of producing highly realistic human-like speech across multiple voices and languages. We programmatically passed a textual response from our language model to ElevenLabs API, specifying a custom voice profile associated with an instructional persona. This voice ID ensures that the avatar consistently speaks in a recognizable and contextually appropriate tone, contributing to the continuity of the interaction experience.

One of ElevenLabs' key properties is its ability to adjust the voice's tone, emotion, and speaking style based on the text it receives. This feature makes the interaction more realistic by delivering responses in a more natural tone and with appropriate emotional expressions.

3.4. Avatar rendering and animations

The conversational agent is represented using a 3D avatar generated through the Avaturn online platform and downloaded in GLB format. The avatar is imported into Blender and converted to FBX format, as Mixamo supports FBX-based rigs for applying and retargeting body animations. To enable speech-driven facial motion, the avatar's predefined viseme morph targets provided by Avaturn are utilized, corresponding to standard phonetic mouth shapes (e.g., silence, PP, FF, TH, DD, KK, CH, SS, NN, RR, AA, E, I, O, U, and AH). During runtime, speech generated through the ElevenLabs text-to-speech pipeline is accompanied by phoneme-level timing information, which is mapped to these visemes and rendered at 30 fps using cosine interpolation.

The avatar operates within the system as a coordinated combination of facial animation, body gestures, and synthesized speech. Figure 2 shows representative snapshots of body animations captured during live interaction in the immersive environment. The figure illustrates commonly used gesture states, including idle-talking animations during neutral conversational flow and disappointment body animations used to express subdued or negative responses. During interaction, the system transitions between these states based on the conversational context and detected emotional cues through keyword-based pattern matching. Happiness-related cues (e.g., "happy," "excited," "great") trigger celebratory gestures and smiling expressions; anger or frustration cues (e.g., "angry," "frustrated," "annoyed") produce tense body language; neutral inputs default to idle talking animations with default facial expressions, while disappointment-related cues (e.g., "sad," "sorry," "disappointed") trigger restrained body movements and lowered expressive intensity, and so on. These animations visually reinforce conversational intent and emotional tone, enabling the avatar to respond in a manner that remains consistent with both speech content and interaction context.

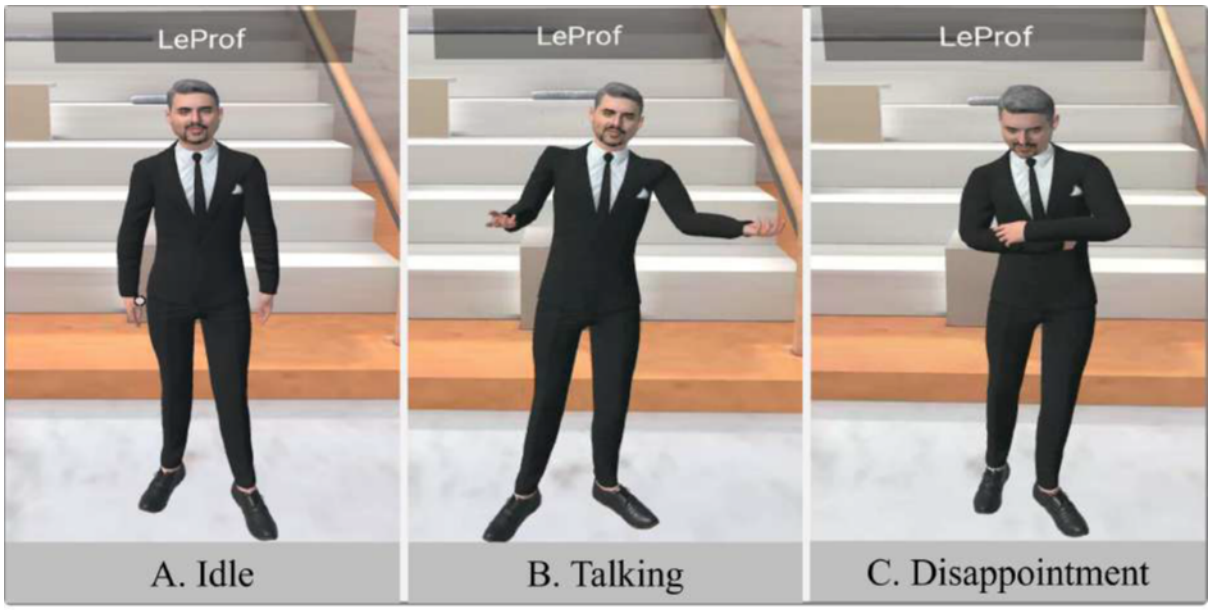

Figure 2: Snapshots of the Virtual Avatar body gestures during user interaction in the immersive environment, demonstrating context-appropriate physical expressions in response to user prompts and questions.

4. Results and Discussions

4.1. System Configuration

To ensure smooth deployment and scalability, the ZapAura setup leverages Docker, which provides a consistent runtime environment, following the architecture adopted by the Hubs core team. The system is fully containerized to isolate and manage core components such as Hubs (client-side), Spoke (web-based 3D scene editor), Reticulum (backend server for networking), and Dialog (a WebRTC SFU server for media streaming), ensuring that each service can run independently while seamlessly communicating within the virtual platform. For the development of the AI-driven agent, we used a pre-trained Llama3.2 language model in RAG environment locally on a workstation equipped with 32 GB RAM and an NVIDIA RTX Ada GPU with 20 GB VRAM, enabling efficient handling of large model parameters. Additionally, resource allocation is carefully optimized for GPU-intensive tasks, such as embedding generation and inference, while CPU-bound services, including the RESTful API and container orchestration, run concurrently on dedicated cores. This setup not only supports low-latency real-time interactions but also maintains modular scalability for future enhancements.

4.2. Participant Insights

To evaluate the expressiveness and contextual alignment of the embodied agent, eight participants completed a structured, scenario-based evaluation within the ZapAura environment. The survey included three emotion-centric scenarios: happy, sad, and surprised. Each scenario was presented as a first-person statement (e.g. “I just completed my comprehensive exam”), which participants gave verbally to the avatar. Participants then observed and evaluated the avatar’s multimodal response, including verbal content, tone of voice, facial expressions, and body language, measuring how convincingly the avatar responded with emotional congruence.

In the Happy scenario, 50% of participants strongly agreed that the avatar’s verbal response, tone, and gestures appropriately conveyed joy, with 37.5% finding facial expressions uplifting. The animated response was perceived as natural and celebratory, particularly through synchronized smiling and arm gestures. In the Sad scenario, 62.5% strongly agreed that tone matched the mood of disappointment, while 50% found facial expressions and body language effectively conveyed sadness. Most participants agreed the avatar made them feel understood, fostering human-like support. In the Surprised scenario,

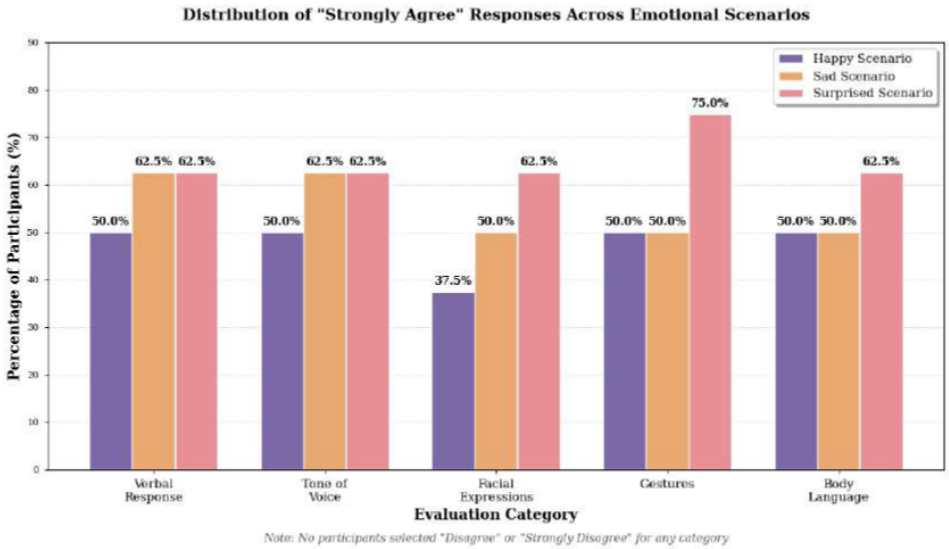

Figure 3: Distribution of "Strongly Agree" responses across three emotional scenarios (n=8). The Surprised scenario achieved the highest agreement, particularly in gesture appropriateness (75%). No participants selected "Disagree" or "Strongly Disagree" for any category, indicating consistent perceived realism.

75% strongly agreed that gestures appropriately expressed surprise, and 62.5% reported that facial expressions and speech delivery helped them process the scenario emotionally. Hand movements and raised vocal tone contributed to a convincing performance. Notably, no participants selected "Disagree" or "Strongly Disagree" for any category, suggesting consistent perceived realism. The distribution of "Strongly Agree" responses is illustrated in Figure 3.

Each item was rated on a 5-point Likert scale ranging from "Strongly Disagree" to "Strongly Agree". Some participants also commented that the avatar was "emotionally expressive" and "surprisingly responsive."

5. Conclusions

In this work, we presented ZapAura, a metaverse application built on the Hubs platform. We integrate a real-time embodied conversational agent to facilitate immersive, multimodal interaction in virtual environments. Our system combines speech recognition, expressive avatar animation, and a RAG architecture for contextual language understanding, enabling users to engage with a domain-specific virtual assistant capable of realistic human-like dialogue. We provide this domain specific assistance through the pre-trained Llama 3.2 model which we integrated within a RAG setup for processing queries across computer vision, multimedia, haptics, and AI ethics domains. Our system design supports both VR and web browser. Through a user study, we observed enhanced engagement and personalized support, when interacting with the embodied agent compared to standard, non-interactive virtual environments. These results highlight the potential of embodied conversational agents as educational and collaborative agents in virtual environments.

Declaration on Generative AI

The authors have not used any Generative AI tools.

References

[1] Hubs Foundation, Hubs community edition documentation, 2025. https://docs.hubsfoundation.org/. Open-source virtual collaboration platform.

[2] B. Preim, P. Saalfeld, A survey of virtual human anatomy education systems, Computers & Graphics 71 (2018) 132–153.

[3] M. Sallam, Chatgpt utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns, in: Healthcare, volume 11, MDPI, 2023, p. 887.

[4] G. Caldarini, S. Jaf, K. McGarry, A literature survey of recent advances in chatbots, Information 13 (2022) 41.

[5] Y. Chen, S. Jensen, L. J. Albert, S. Gupta, T. Lee, Artificial intelligence (ai) student assistants in the classroom: Designing chatbots to support student success, Information Systems Frontiers 25 (2023) 161–182.

[6] J. Achiam, S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F. L. Aleman, D. Almeida, J. Altenschmidt, S. Altman, S. Anadkat, et al., GPT-4 technical report, arXiv preprint arXiv:2303.08774 (2023).

[7] Gemini Team, R. Anil, S. Borgeaud, J.-B. Alayrac, J. Yu, R. Soricut, J. Schalkwyk, A. M. Dai, A. Hauth, K. Millican, et al., Gemini: a family of highly capable multimodal models, arXiv preprint arXiv:2312.11805 (2023).

[8] H. Touvron, T. Lavril, G. Izacard, X. Martinet, M.-A. Lachaux, T. Lacroix, B. Rozière, N. Goyal, E. Hambro, F. Azhar, et al., LLaMA: Open and efficient foundation language models, arXiv preprint arXiv:2302.13971 (2023).

[9] Anthropic, Models overview – claude documentation, 2026. https://platform.claude.com/docs/en/about-claude/models/overview. Accessed: 2026-01-26.

[10] A. Liu, B. Feng, B. Xue, B. Wang, B. Wu, C. Lu, C. Zhao, C. Deng, C. Zhang, C. Ruan, et al., DeepSeek-v3 technical report, arXiv preprint arXiv:2412.19437 (2024).

[11] V. Chheang, S. Sharmin, R. Márquez-Hernández, M. Patel, D. Rajasekaran, G. Caulfield, B. Kiafar, J. Li, P. Kullu, R. L. Barmaki, 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, 2024, pp. 21–30.

[12] K. Cheng, Q. Guo, Y. He, Y. Lu, S. Gu, H. Wu, Exploring the potential of GPT-4 in biomedical engineering: the dawn of a new era, Annals of Biomedical Engineering 51 (2023) 1645–1653.

[13] P. Saalfeld, A. Schmeier, W. D'Hanis, H.-J. Rothkötter, B. Preim, Student and teacher meet in a shared virtual reality: a one-on-one tutoring system for anatomy education, arXiv preprint arXiv:2011.07926 (2020).

[14] A. Shoa, D. Friedman, Milo: an LLM-based virtual human open-source platform for extended reality, Frontiers in Virtual Reality 6 (2025) 1555173.

[15] H. Wan, J. Zhang, A. A. Suria, B. Yao, D. Wang, Y. Coady, M. Prpa, Building LLM-based AI agents in social virtual reality, in: Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, 2024, pp. 1–7.

[16] K. B. Buldu, S. Özdel, K. H. C. Lau, M. Wang, D. Saad, S. Schönborn, A. Boch, E. Kasneci, E. Bozkir, Cuify the XR: An open-source package to embed LLM-powered conversational agents in XR, in: 2025 IEEE International Conference on Artificial Intelligence and eXtended and Virtual Reality (AIxVR), IEEE, 2025, pp. 192–197.

[17] C. M. De Melo, K. Kim, N. Norouzi, G. Bruder, G. Welch, Reducing cognitive load and improving warfighter problem solving with intelligent virtual assistants, Frontiers in Psychology 11 (2020) 554706.

[18] Y. Yoo, Y. Sun, O. Shaikh, A. S. Won, Comparing text-only and virtual reality-embodied conversational AI agents for interpersonal skills training, in: Companion Publication of the 2025 Conference on Computer-Supported Cooperative Work and Social Computing, 2025, pp. 194–198.

[19] Y. Gao, Y. Xiong, X. Gao, K. Jia, J. Pan, Y. Bi, Y. Dai, J. Sun, H. Wang, H. Wang, Retrieval-augmented generation for large language models: A survey, arXiv preprint arXiv:2312.10997 2 (2023).

[20] Z. Li, Z. Wang, W. Wang, K. Hung, H. Xie, F. L. Wang, Retrieval-augmented generation for

educational application: A systematic survey, Computers and Education: Artificial Intelligence (2025) 100417.
[21] K. R. Prajwal, R. Mukhopadhyay, V. P. Namboodiri, C. V. Jawahar, A lip sync expert is all you need for speech to lip generation in the wild, in: Proceedings of the 28th ACM International Conference on Multimedia, 2020, pp. 484–492.
[22] R. Skarbez, F. P. Brooks, Jr., M. C. Whitton, Immersion and coherence: Research agenda and early results, IEEE Transactions on Visualization and Computer Graphics 27 (2020) 3839–3850.
[23] M. E. Latoschik, D. Roth, D. Gall, J. Achenbach, T. Waltemate, M. Botsch, The effect of avatar realism in immersive social virtual realities, in: Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology, 2017, pp. 1–10.
[24] K. Yano, A platform for analyzing students' behavior in virtual spaces on mozilla hubs, in: International Conference on Immersive Learning, Springer, 2023, pp. 209–219.
[25] V. Uribe, P. Figueroa, V. Gomez, The influence of metaverse environment design on the quality of experience in virtual reality classes: a comparative study, in: Frontiers in Education, volume 9, Frontiers Media SA, 2024, p. 1451859.
[26] R. Tulaskar, A. Opel, M. Turunen, I. J. Erdal, A. K. Gulbrandsen, Through the looking glass: Exploring the metaverse for teaching media education in real-time international collaboration, Norsk Medietidsskrift 31 (2025) 1–19.
[27] V. Methuku, P. K. Myakala, Digital doppelgangers: Ethical and societal implications of pre-mortem AI clones, arXiv preprint arXiv:2502.21248 (2025).