



Extended Abstract—Immersive Education with Historical Characters: Conversational MetaHuman Based on Large Language Model, Speech Recognition and Generation

Victor Yuan, Alan Miller and Iain Oliver

University of St Andrews, St Andrews, Scotland

vy1@st-andrews.ac.uk

alan.miller@st-andrews.ac.uk

iao@st-andrews.ac.uk

Abstract. With the advancement of technologies in computer graphics, large language models, and speech processing, the development of interactive characters for historical education has become increasingly feasible. This work-in-progress paper presents a functional prototype that integrates frameworks such as Unreal Engine 5, MetaHuman and Convai to develop an interactive system for engaging with historical figures. This system provides an accessible solution for educational institutions, particularly those with limited resources, by minimizing development efforts while maintaining a high level of interactivity, immersion, and customizability. Users can engage in conversational interactions with historically accurate figures, who respond in real-time with facial animations, body language, and natural speech. These characters are designed to generate meaningful dialogue based on databases of well-documented historical knowledge to enhance educational experiences and effectiveness. This approach offers a novel method for cultural and historical education, providing an immersive interface that allows users to engage with history in a more personal and immediate manner. The paper also includes a discussion of potential future enhancements to improve user engagement and learning outcomes. This work demonstrates the potential of integrating contemporary technologies to develop dynamic, accessible, and impactful tools for historical education, paving the way for more engaging and effective methods of experiencing history.

Keywords: Immersive Learning, Interactive System, Virtual Character, Virtual Reality.

1 Introduction

Institutions such as galleries, museums, libraries, and archives house vast collections of cultural and historical records, which serve as valuable educational resources for preserving past cultures and heritage. To keep these resources relevant and engaging, educators have increasingly experimented with immersive learning approaches, particularly Virtual Reality [1]. A key aspect of learning about past cultures is the presence of human figures, the primary contributors and custodians of historical societies and artifacts. Gamage and Ennis suggested that incorporating virtual characters in educational environments significantly enhances engagement and knowledge retention, reaffirming the potential of this approach [2].

However, achieving convincing realism in interactions with virtual human characters remains a major challenge [3]. Historically, efforts to create these characters in VR were constrained by limitations in computational efficiency, most notably in computer graphics. These constraints often led to an undesired effect, causing discomfort in observers and diminishing the sense of immersion [4]. Early virtual environment development primarily focused on improving photorealism to address such issues and reduce the unsettling impact.

In recent years, advancements in computer graphics have significantly improved visual fidelity, making virtual characters appear more lifelike. Currently, the focus has shifted toward creating more emotionally resonant experiences through integration of motion and speech processing. The breakthroughs and widespread adoption of artificial intelligence, particularly in large language models, have enabled more complex and spontaneous

interactions between users and virtual characters, enriching the overall immersive experience. Nevertheless, achieving realistic social interaction requires more than convincing dialogue and photorealism. Elements such as body language, facial expressions, and behavior must also be effectively integrated [5].

Joan and Caecilia emphasized the importance of refining virtual character responses in VR [6], particularly ensuring their emotional expressions align with user interactions. Achieving this level of sophistication typically requires a highly intensive and costly system. This paper explores an integrated solution with existing services and frameworks for designing interactive and contextual NPCs (non-player characters) that aim to create effective connections with users. By enhancing the interactions between users and virtual characters, this research seeks to demonstrate how virtual environments can foster higher engagement, creating new possibilities for future cultural and heritage education.

This work brings together concepts from multiple disciplines, including computer graphics, artificial intelligence, human-computer interaction, and educational technology. By integrating MetaHuman for photorealistic character modeling with natural language processing and speech synthesis, the system aims to create a more interactive and immersive historical learning experience. This approach explores the potential of real-time, AI-driven dialogue to enhance engagement and provide a platform for users to interact with historical figures in a conversational manner. This approach seeks to complement traditional methods of historical education by offering an alternative way to engage with historical narratives in an interactive setting.

This manuscript builds on the 3D production assets introduced in a conference paper presented at the iLRN 2024 conference [7]. The potential ideas discussed at the conference are further explored and expanded in this paper, with the goal of advancing the development of virtual environments for educational and cultural preservation purposes.

2 Related Work

In the 1970s, Mori observed the Uncanny Valley effect [8], where humans tend to have negative responses to humanoid figures that are not highly realistic. Acceptance improves as human likeness becomes more accurate. Current research suggests that the Uncanny Valley effect arises from conflicting visual and behavioral cues in a character's appearance [9], leading to a perceptual mismatch. In contrast, cartoon characters, which are consistently unrealistic, are easily accepted by users. By comparison, generated avatars with inconsistent realism, such as highly detailed models paired with stiff expressions, are more likely to trigger the Uncanny Valley effect.

This theory introduces the "buckets effect," which suggests that the perceived realism of a character is dictated by its least well-executed component, whether it be the 3D model, animations, voice, or conversational manner. Consequently, it becomes crucial to address each aspect comprehensively when developing a virtual character system, significantly increasing both the complexity and cost of implementation. Additionally, research indicates that the Uncanny Valley effect is often amplified during interactions via VR headsets [10], a technology increasingly utilized in educational settings.

Valentin et al. underscored the critical role of the eyes in virtual characters, observed that users primarily rely on eye contact to assess a character's realism and demonstrated that users fixate on the eyes before evaluating other facial features [9]. This aligns with interactions in real life, where individuals focus on the eyes as a key element of nonverbal communication and social interaction [11]. The phenomenon of "dead eyes" in virtual characters can evoke discomfort, as the eyes are integral to expressing intentions, emotions, and overall well-being, all of which are essential for shaping user perception and establishing a sense of connection.

Focusing on the recent development of language model-powered virtual characters, numerous researchers have proposed systems that leverage this technology. James introduced a system incorporating multisensory feedback, including haptics and olfactory VR, to engage users in a dialogical relationship, while also addressing recent safety concerns and limitations of characters powered by large language models [12]. Sepehr et al. developed an AI-driven system that was integrated into a gamified heritage education platform [13], specifically showcasing the story of the Silk Road. Similarly, Alexander et al. implemented an AI-driven chat system featuring multiple historical figures and conducted extensive evaluations of the system's effectiveness [14].

The growing number of similar systems provide adaptive and dynamic learning experiences for users. As AI continues to advance, such systems show great promise in preserving and promoting cultural heritage while offering an engaging educational experience. However, Tajik et al. also highlighted potential risks and concerns associated with relying on large language models for educational purposes [15], primarily focusing on issues related to accuracy, trustworthiness, and ethical considerations.

3 Design and Implementation

3.1 Unreal Engine and MetaHuman

As mentioned earlier, this project builds upon a previous conference paper that produced highly photo-realistic historical characters. Cross-version compatibility enables seamless integration with previous projects, allowing the reuse of existing assets. Moreover, Unreal Engine 5 (UE5) offers a significant advantage through the integration of the MetaHuman framework, which provides a comprehensive set of tools for creating highly realistic and detailed human characters that are fully rigged for animation. While achieving photorealism does not guarantee positive user acceptance during interactions, this robust toolset significantly reduces the effort required to ensure that the models appear photorealistic when static.

3.2 Integration of Convai

Convai is a modular platform that integrates a range of AI services essential for creating virtual characters [16], including language models, speech generation, animation actions, and perception components. It provides tools for refining character behavior based on their backstory, knowledge base, and personality, allowing for a more tailored and dynamic user experience.

For each character, a unique backstory is developed from historical sources and integrated into the Convai interface. Additionally, each character's tone of speech is adjusted to reflect their most appropriate communication style.

3.3 Virtual Character and Environment

The project aims to simulate a virtual museum environment set within a medieval-style building, providing users with a time-travel-like experience. The environment is constructed using modular assets from the Unreal Marketplace for ease of development. Four historical characters are placed within the museum, each accompanied by textual information and a historical portrait displayed alongside them.

The project includes four interactable characters based on Convai: Queen Marie de Guise, King James V, David Beaton, and John Knox. These figures represent significant historical personalities from their era and are placed within the virtual environment to interact with users as the primary witnesses of history. Their presence in the museum is designed to enhance user engagement, providing opportunities for dialogue and exploration of the historical context in which they lived.

3.4 Pedagogy and Limitations

From a cognitive load perspective, immersive and interactive environments can structure historical information in ways that support gradual knowledge acquisition. Rather than presenting large amounts of abstract historical content all at once, virtual characters deliver segmented, contextually relevant information based on user interaction. This aligns with Vygotsky's Zone of Proximal Development [17], where learning is most effective when scaffolded within a learner's current level of understanding. Additionally, social constructivist perspectives emphasize the role of dialogue in learning. While AI-driven interactions cannot replace human discussion, they can serve as an engaging supplement, allowing learners to practice historical reasoning and contextual analysis through guided conversations.

However, this approach is not a replacement for traditional methods, as current generative AI models have inherent limitations. AI-generated responses, particularly when dealing with incomplete or ambiguous historical records, may sometimes be inaccurate or speculative, as history often contains gaps and uncertainties. Unlike textbooks, which provide rigorously validated information, AI-generated content does not always offer absolute accuracy. To address this, the system includes explicit disclaimers reminding users that it should not be considered a definitive source of historical knowledge. This encourages learners to approach virtual experience with critical thinking, using it as a tool for exploration and engagement rather than a substitute for verified historical scholarship.

3.5 Future Steps

Lively expressions of eyes are an important aspect of virtual avatars. For the future development of the project, additional facial animations, particularly for the eyes, are planned to further enhance the realism of the characters.

Future development will focus on refining facial animations, especially for eye movements, to enhance the realism of the characters and improve the overall interactive experience.

In addition to technical improvements, further research is planned to evaluate the system effectiveness in historical education. The target group for this study will include students, educators, and museum visitors, as they represent key audiences for immersive learning applications. The research will adopt a mixed-methods approach, combining both qualitative and quantitative analyses to assess user engagement, learning outcomes, and perceived authenticity of historical interactions.

Data collection will involve structured user testing sessions, where participants will interact with the virtual characters and provide feedback through surveys and interviews. Surveys will measure engagement, perceived realism, and educational value using Likert-scale questions, while interviews will gather in-depth insights into user experiences and expectations.

For data analysis, quantitative responses will be statistically analyzed to identify trends in engagement and learning impact, while qualitative data from interviews will be thematically coded to extract key insights into user perceptions and areas for improvement. By integrating these findings, the study aims to refine the system iteratively, ensuring that both its technological and pedagogical aspects align with the needs of learners and educators.



Fig. 1. An exemplary scene with interactable characters powered by MetaHuman and Convai.

4 Conclusion

The creation of highly realistic, interactive, and emotionally engaging virtual spaces has the potential to revolutionize how we engage with historical and cultural resources. By bridging the gap between static digital records and immersive experiences, new paradigms for exploring history and culture can emerge, allowing for deeper engagement with historical events, cultural practices, and personal narratives.

Beyond historical education, the integration of immersive technologies with AI-driven virtual characters presents practical applications across various domains. Museums and cultural institutions can use such systems to create interactive exhibits that engage visitors with personalized storytelling and real-time historical dialogues. In educational settings, this approach can complement traditional learning by offering students an interactive way to explore historical events, figures, and cultural contexts. Additionally, similar frameworks could be adapted for language learning, soft skills training, or historical reenactments, where users can practice conversations and gain deeper insights. As immersive technologies continue to evolve, future iterations of such systems could incorporate multi-user interactions, adaptive learning pathways, and more advanced emotion recognition to further enhance engagement and accessibility, making historical and cultural education more dynamic and widely accessible.

In conclusion, as technology continues to advance, the integration of AI-driven virtual characters and immersive environments into educational and cultural contexts presents transformative opportunities. Such systems could serve as valuable platforms for preserving and promoting cultural heritage. By overcoming current challenges and refining these tools, future research and development can further enhance immersive learning experiences, deepening our connection to history.

References

1. Fabola A, Kennedy S, Miller A, Oliver I, McCaffery J, Cassidy C, Clements J, Vermehren A: A virtual museum installation for time travel. *Communications in Computer and Information Science*, 255–270 (2017)
2. Gamage V, Ennis C: Examining the effects of a virtual character on learning and engagement in serious games. In: *Proc. 11th ACM SIGGRAPH Conf. Motion, Interact. Games*, pp 1–9 (2018)
3. Bissinger B, Märtin C, Fellmann M: Support of virtual human interactions based on facial emotion recognition software. *Lecture Notes in Computer Science*, 329–339 (2022)
4. Seyama J, Nagayama RS: The uncanny valley: effect of realism on the impression of artificial human faces. *Presence: Teleoperators and Virtual Environments* **16**, 337–351 (2007)
5. Cao Q, Yu H, Charisse P, Qiao S, Stevens B: Is high-fidelity important for human-like virtual avatars in human computer interactions? *International Journal of Network Dynamics and Intelligence*, 15–23 (2023)
6. Llobera J, Charbonnier C: Interactive characters for virtual reality stories. In: *ACM International Conference on Interactive Media Experiences*, pp 322–325 (2021)
7. Yuan V, Miller A, Nyman PJ, Oliver I: Designing metahuman-based historical characters in virtual exhibitions and scenes: a case study on st andrews. *Communications in Computer and Information Science*, 232–246 (2025)
8. Mori M, MacDorman K, Kageki N: The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine* **19**, 98–100 (2012)
9. Schwind V, Wolf K, Henze N: Avoiding the uncanny valley in virtual character design. *Interactions* **25**, 45–49 (2018)
10. Hepperle D, Odell H, Wolfel M: Differences in the uncanny valley between head-mounted displays and monitors. In: *2020 International Conference on Cyberworlds (CW)* (2020)
11. Jongerius C, Hessels RS, Romijn JA, Smets EM, Hillen MA: The measurement of eye contact in human interactions: a scoping review. *Journal of Nonverbal Behavior* **44**, 363–389 (2020)
12. Hutson J: Combining large language models and immersive technologies to represent cultural heritage in the metaverse context. *Springer Series on Cultural Computing*, 265–281 (2024)
13. VaezAfshar S, Eshaghi S, Hadighi M, Varinlioglu G: From past to present: a study of AI-driven gamification in heritage education. In: *eCAADe proceedings*, vol 2, pp 249–258 (2024)
14. Steinmaurer A, Dengel A, Comanici M, Buchner J, Memminger J, Gütl C: Immersive learning in history education: exploring the capabilities of virtual avatars and large language models. *Communications in Computer and Information Science*, 363–374 (2025)
15. Tajik E, Alijani F: The ethical concerns surrounding GPT in education (2024)
16. Documentation, Convai API Docs. <https://docs.convai.com/api-docs>. Accessed 6 Mar 2025
17. Cowling M, Vanderburg R: An interactive virtual reality physics instructional environment based on Vygotskian educational theory. In: *ASCILITE's First Virtual Conference. Proceedings ASCILITE 2020 in Armidale*, pp 168–173 (2020)