Structured Teaching Prompt Articulation for Generative-AI Role Embodiment with Augmented Mirror Video Displays

Llogari Casas
School of Computing
Edinburgh Napier University
Edinburgh, Scotland, United Kingdom
3FINERY LTD
Lanark, Scotland, United Kingdom
llogari92@gmail.com

Kenny Mitchell
School of Computing
Edinburgh Napier University
Edinburgh, Scotland, United Kingdom
3FINERY LTD
Lanark, Scotland, United Kingdom
k.mitchell2@napier.ac.uk

Figure 1: Illustration of various historical dance contexts generated from our structure-guided prompt dialogues for teachers employing *Stable Diffusion* and *ControlNet* in a classroom enhanced with an augmented mirror video display. Given a) a camera image snapshot, and b) an inferred depth map, we generate teaching role embodiments interactively (c-j) to realize the students' appearance as c) *West African dancers, d) Renaissance court dancers, e) Charleston, f) classical ballet, g) Bharatanatyam, h) Kabuki, i) Samba, and j) Flamenco.* The visuals not only capture traditional attire but the environmental settings, offering an immersive representation of each dance form's cultural heritage scenarios enriching the teaching experience.

Abstract

We present a classroom enhanced with augmented reality video display in which students adopt snapshots of their corresponding virtual personas according to their teacher's live articulated spoken educational theme, linearly, such as historical figures, famous scientists, cultural icons, and laterally according to archetypal categories such as world dance styles. We define a structure of generative AI prompt guidance to assist teachers with focused specified visual role embodiment stylization. By leveraging role-based immersive embodiment, our proposed approach enriches pedagogical practices that prioritize experiential learning.

CCS Concepts

• **Computing methodologies → Mixed / Virtual Reality**.

VRCAI '24, December 01–02, 2024, Nanjing, China

ACM ISBN 979-8-4007-1348-4/24/12
https://doi.org/10.1145/3703619.3706049

Keywords

Virtual Reality, Artificial Intelligence, Generative AI, Human-Computer Interaction

ACM Reference Format:
Llogari Casas and Kenny Mitchell. 2024. Structured Teaching Prompt Articulation for Generative-AI Role Embodiment with Augmented Mirror Video Displays. In *The 19th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry (VRCAI '24), December 01–02, 2024, Nanjing, China.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3703619.3706049

1 Introduction

The integration of generative AI into learning environments promises to enable new avenues for enhancing engagement and interaction through immersive teaching experiences. By employing *Stable Diffusion* [Ho et al. 2020], a generative models may create images based on textual prompts, enabling visual forms to adopt various directed appearances and identities and provide accordant environments em-placing their visual settings [Liu et al. 2023; Ramesh et al. 2021]. This yields a powerful capability for teachers as part of their classroom elocution to describe a subject's setting and for that to be realized visually immediately for and with their students on an augmented mirror video display. In this manner, our approach empowers teaching with a visceral form of the concept known as *role embodiment* [Keevallik 2010], where students visually engage as historical figures, scientific professionals, or cultural icons, creating a richer, more immersive learning experience through active role-play [Carnicero Pérez et al. 2023]. To refine and direct image generation with greater precision, this study in particular incorporates *ControlNet*, a development of stable diffusion designed for enhanced control over generated outputs, ensuring that visual transformations align both with the textually prompted educational goals and a provided camera input image snapshot [Zhang et al. 2023]. Upstream of this we integrate speech recognition to provide a natural spoken interface to the controlled directed image generation.

Generative AI models, such as DALL-E or GPT4, can synthesize high-fidelity visual content from textual descriptions. Despite their utility, these models are fundamentally constrained by their reliance on text as the sole conditioning input. This constraint limits their ability to adapt generated outputs to structured spatial inputs, such as depth maps, semantic segmentation masks or pose configurations among others. Consequently, such models are unsuitable for applications that demand precise alignment with live physical contexts, such as interactive environments and real-time personalizations. In contrast, *ControlNet* addresses this gap by enabling the integration of multimodal input modalities, including depth maps, into the generative process. Depth conditioning, in particular, is key for aligning visual outputs with participants' physical profiles, such as body shapes and spatial arrangements, in real-time settings. This capability extends the applicability of generative models into domains requiring contextual and participant-specific outputs. By leveraging depth-based conditioning, *ControlNet* facilitates the generation of visuals that are not only semantically accurate but also spatially coherent, thereby supporting novel applications such as embodied role-play and immersive, context-aware educational experiences.

Role embodiment through generative AI aligns with studies on immersive learning, where students engage more deeply when taking on roles or personas within educational scenarios. Research suggests that embodying historical figures, for instance, develops empathy and enhances memory retention, as students gain a personal connection to the material [Miguel-Revilla et al. 2021]. Similarly, in STEM fields, students can explore roles laterally through archetypes such as scientists, engineers, or astronauts, which translates into a stronger identification with the subject matter and supports sustained engagement [Singer et al. 2020]. In more detail we explore a variety of cultural dance styles as teaching scenarios to connect dance students more directly with an visually embodied teaching context. This paper introduces a framework employing the mechanisms of *Stable Diffusion*[Ho et al. 2020] in combination with *ControlNet*[Zhang et al. 2023] to achieve targeted, role-specific transformations applicable in teaching settings. This integration enables controlled visual customization conforming to the human forms imaged within the classroom, allowing educators to design immersive, contextually accurate experiences that align with educational objectives across a variety of subjects. The primary technical contributions of this paper are:

- Development of structured prompt guidance that directs *ControlNet* [Zhang et al. 2023] to produce specific visual personas and accordant environments, allowing students to adopt roles that align with their learning objectives.
- A framework for speech driven augmented mirror video displays that leverages generative AI for interactive, role-based learning experiences, enhancing engagement and comprehension across various disciplines.

2 A Framework for Interactive Articulated Classroom Augmented Reality Snapshot Generation

The rapid advancement of generative AI, particularly in the area of text-to-image and image-to-image models, has introduced new opportunities for immersive learning. This section describes the background and system elements for our generative AI enhanced augmented classroom.

2.1 Augmented Mirror Video Display with Generative AI

Stable Diffusion[Ho et al. 2020] is a generative AI model built upon the principles of diffusion processes, a method that iteratively refines a noisy image to produce a coherent output and may further steer alignment with a given text prompt [Liu et al. 2023]. This generative process begins with a random noise image and iteratively applies noise reduction techniques, informed by statistical patterns learned from extensive training data. Through these repeated denoising steps, the model is able to produce high-quality images that correspond closely with the content described in the prompt. Stable Diffusion leverages a combination of convolutional neural networks (CNNs) and transformer architectures to achieve high-resolution outputs. *ControlNet*, an extension developed to augment this functionality, introduces a secondary control mechanism that enhances user control over the output by conditioning the generative process on an additional input image [Zhang et al. 2023]. This control mechanism allows *ControlNet* to guide the generation process by imposing specific structural and compositional attributes derived from a reference image, such as pose, depth, or a spatial layout. *ControlNet*'s architecture is based on a branched network design, where control features are processed independently but integrated at multiple stages within the primary diffusion model. This architecture enables *ControlNet* to process control images while maintaining the visual fidelity of the generated output, effectively aligning the generation with the desired structural features from the control image. In practice, *ControlNet* achieves precise alignment by using the input image to condition the output of Stable Diffusion at various network layers (see figure 3). For example, if the goal is to generate an image of a historical figure with a specified posture, *ControlNet* utilizes a reference image containing the desired pose, which guides the generation process to match this posture while retaining other attributes described in the textual prompt. By injecting control features into the diffusion process, *ControlNet* allows the model to adhere closely to both the structural cues from the input image and the semantic content of the prompt.

Specifically by using the source video camera image we use the MiDaS model [Ranftl et al. 2020] to infer a corresponding depth image, which provides the control image to generate a directed stylization of the source full camera color image we, in turn, retain the human forms and their spatial *juxtapositioning* suitable for classroom visual augmentations, such as dance classes (see 4).

Through a PC connected microphone we capture the teacher's voice input as audio frequency signals and processed into a speech-to-text module, which utilizes *Google Cloud, [Cloud 2024]*'s Chirp speech-to-text API.

Thus with both vocalized text prompt inputs and depth control image generation we emcompass a framework for directed stylized snapshot image generation of students in their class for interactive educational role embodiment.

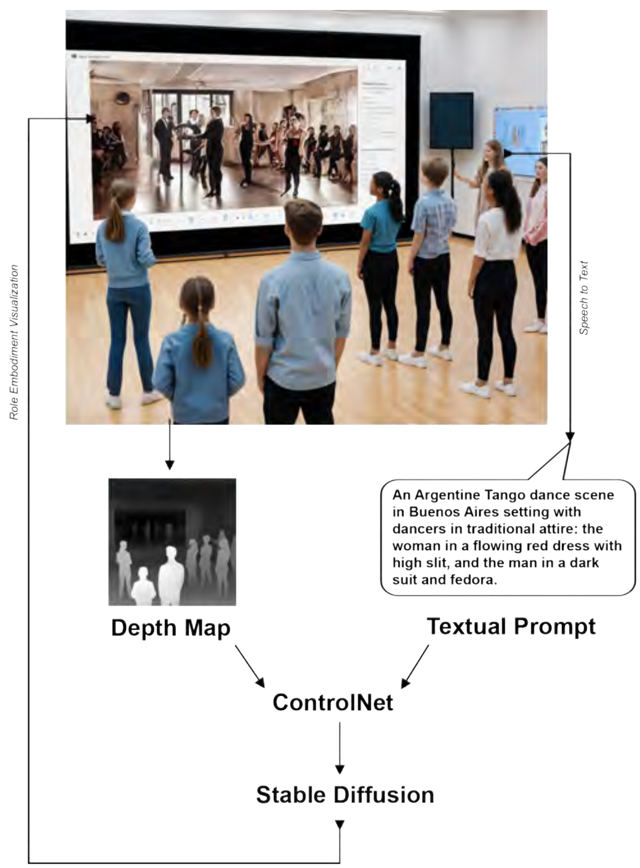

Figure 2: Illustration of the role embodiment process in a mixed reality teaching setting. Students observe a projected visualization where they are embodied as participants in an Argentine Tango dance scene set in Buenos Aires. The process begins with a *Depth Map* generated through MiDaS to direct ControlNet to capture the spatial arrangement and structure of the students. A *Textual Prompt* specifies the details of the desired scene, describing dancers in traditional attire. With this controlled Stable Diffusion, the input from the depth map and textual prompt combine to produce an immersive, contextually accurate image that reflects the role embodiment scenario.

2.2 Generative AI in Teaching Settings

Generative AI, especially models for image synthesis, has peaked significant interest in educational contexts due to its ability to produce engaging and contextually relevant visuals from prompts [Liu et al. 2023; Ramesh et al. 2021]. Models like *DALL-E* and *Stable Diffusion* have demonstrated capability to generate high-quality, prompt-driven images, proving to be new tools for visualization in classrooms [Rizvi et al. 2023]. By enabling the creation of custom visuals, these tools offer educators a way to illustrate abstract concepts, historical contexts and scientific processes that are otherwise challenging to represent in a conventional classroom settings [Cao et al. 2023]. Nonetheless, the applications of generative AI in education extend beyond mere visual aids. These models offer a dynamic platform for immersive learning experiences by creating personalized content that can engage students across various subjects. For instance, in history lessons, generative AI can be used to reconstruct visual representations of historical figures, environments, and events, allowing students to better understand and connect with the material. This approach supports experiential learning, in which students actively participate in the learning process through role-play and contextualized visualization [Swink 1993].

In the sciences, generative AI models enable educators to illustrate complex phenomena such as cellular structures, chemical reactions, or planetary movements with high visual fidelity. This capability helps students to visualize intricate processes, promoting better comprehension and retention [Cao et al. 2023]. Furthermore, the flexibility of generative models allows educators to customize images according to curriculum requirements, making these tools highly adaptable to specific learning objectives [Dehouche and Dehouche 2023]. Generative AI also supports creativity and self-expression in students, offering them opportunities to explore visual storytelling and identity formation. For example, in literature, students can use AI-generated visuals to create avatars or personas aligned with characters and historical figures, developing a deeper understanding of diverse perspectives and enhancing empathy [Kao and Harrell 2021]. This identity exploration, facilitated by generative AI, can be a valuable educational experience, as students learn to connect with viewpoints beyond their own, improving critical thinking and cultural awareness [Singer et al. 2020].

2.2.1 Role-Based Immersive Learning. In this context, immersive learning through role-based pedagogy has been shown to significantly enhance student engagement and understanding [Carnicero Pérez et al. 2023; Miguel-Revilla et al. 2021]. By embodying characters within their studies, students gain a personal connection to the material, fostering empathy and facilitating memory retention [Samsinar and Fitriani 2020]. For instance, embodying historical figures helps students relate to different perspectives, while adopting scientific personas can encourage a deeper identification with STEM fields [Singer et al. 2020]. The application of *ControlNet* for immersive role-based learning supports these findings by providing realistic and adaptable representations aligned with specific educational goals.

2.2.2 Ethical AI-Based Role Embodiment. Concurrently with its potential, the integration of generative AI into educational practices

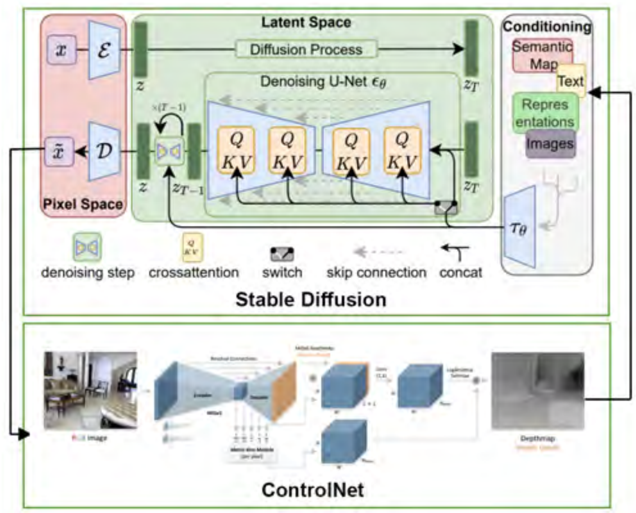

Figure 3: Illustration of the integration between Stable Diffusion and ControlNet for controlled image generation. The upper section depicts the Stable Diffusion model, which utilizes a latent space-based diffusion process with a denoising U-Net architecture. It operates through iterative denoising steps, with cross-attention mechanisms to incorporate conditioning inputs. The lower section shows ControlNet, which enhances Stable Diffusion by introducing structured guidance based on an input image. This process involves extracting structural information via an encoder-decoder network. ControlNet conditions the diffusion model, ensuring that generated images follow specific structural cues while preserving the content specified in the textual prompt.

requires careful consideration of ethical and pedagogical implications. The customization of images must be conducted responsibly to avoid reinforcing stereotypes and creating misrepresentations. Ethical guidelines and appropriate prompts can help mitigate these risks, ensuring that generative AI supports inclusive and accurate portrayals in educational content [Al-kfairy et al. 2024].

Significant ethical concerns regarding identity, representation, and potential misuse must be thoughtfully considered. When using AI to generate visual portrayals of historical and cultural figures, accuracy and respect for its context are paramount [Al-kfairy et al. 2024]. Misrepresentation or oversimplification of a figure's setting can unintentionally reinforce stereotypes and propagate inaccurate historical narratives, which may undermine the educational value of role embodiment.

One primary ethical concern lies in the potential for cultural appropriation and misrepresentation, particularly when generating visuals of figures from historically marginalized communities. The use of generative AI models that lack cultural nuance might lead to visuals that oversimplify and distort complex identities, reducing individuals and communities to stereotypical features. To avoid these pitfalls, educators should employ careful prompt design, use diverse reference materials, and consult relevant historical and cultural resources where possible to ensure that AI-generated content reflects the depth and diversity of the personas being represented [Tafazoli 2024]. Another ethical issue concerns the possibility of reinforcing biases embedded within AI models themselves. Since generative AI models are trained on large datasets that may contain historical and social biases, there is a risk that these biases could manifest in the generated images, affecting accuracy and reinforcing potentially harmful stereotypes [Buolamwini and Gebru 2018]. Educators can mitigate this by critically evaluating generated outputs and choosing prompts carefully to reduce the likelihood of biased representations. Ongoing improvements in AI model training and increased transparency in dataset composition are also needed to minimize inherent biases. Nonetheless, through intentional design, transparency, and adherence to ethical guidelines, educators can leverage generative AI responsibly, ensuring that it serves as a positive, culturally sensitive educational tool.

3 Structured Prompt Design for Role Embodiment

To enable students to visually immerse themselves in various educational themes, we utilize *ControlNet* to craft prompts that reflect distinct historical, scientific, and cultural personas. As depicted in figure 2, this process begins with live spoken prompt inputs in the classroom, allowing educators to dynamically influence the visualization and tailor it to the specific teaching needs. These prompts, coupled with depth map control generated using *MiDas*[Ranftl et al. 2022] within the *ControlNet* pipeline, facilitate a novel form of role embodiment by seamlessly aligning the generated visualizations with the live participants' body shapes. Using the *MiDas* framework, depth maps are generated from RGB input to capture the spatial structure and depth profiles of the live classroom environment. These depth maps serve as input to the *ControlNet* pipeline, which conditions the image generation process of *Stable Diffusion* to align the synthesized content with the physical geometry of the participants. This depth conditioning ensures that the generated augmentations are not only visually coherent but also spatially situated in a way that matches the students' body shapes and movements.

Each of these examples illustrates the role of prompt design in creating immersive educational experiences that align with specific learning objectives. By adopting these visual personas, students are encouraged to develop empathy and deeper engagement with the material, connecting with characters from diverse historical, scientific, and cultural backgrounds. This approach supports experiential learning by transforming passive learning into active, role-based engagement, with a focus on live and contextually accurate enactments of educational themes. Below, we present six general case examples (see figure 4), each demonstrating how prompt-driven transformations can align with educational goals by visually contextualizing each role. For each prompt, the **setting** defines the broader context or environment where the persona is placed. The setting establishes the foundational visual cues, guiding the generation of background elements and contextual details that support immersion in the role. The **persona** represents the historical, cultural, or fictional figure that students will embody. This component is central to role embodiment, ensuring the generated image aligns

Figure 4: Illustration of role embodiment cases generated from custom prompts using *Stable Diffusion* and *ControlNet*. The figure displays the process for each case, starting with the source image (a), followed by a depth map (b), and concluding with target results (c-h) that visually represent different educational personas. These include: (c) Galileo Galilei as a historical figure in the Renaissance, (d) an astronaut in space representing modern STEM, (e) Marie Curie in a 1900s laboratory setting, (f) Cleopatra as a cultural icon in ancient Egypt, (g) Sherlock Holmes as a fictional character in Victorian London, and (h) Leonardo da Vinci as an inventor in a Renaissance workshop. Each generated result emphasizes relevant attire, setting, and details, aligning with educational goals to enhance engagement and understanding through role embodiment.

with the character's visual identity, attire, and recognizable features. the **time** specifies the era associated with the persona, which influences stylistic elements, attire, and environmental accuracy ensuring that the details are visually consistent with the period. The *place* adds specificity to the environment by pinpointing the location where the persona would logically appear, such as a laboratory, an observatory, or a palace. This component influences the spatial composition, guiding the model to include appropriate setting details. **Objects** include distinctive items associated with the persona, such as a telescope for Galileo or a magnifying glass for Sherlock Holmes. Objects serve as focal points that reinforce the persona's role and add layers of authenticity, ensuring students recognize the tools or symbols associated with each figure. And the **sentiment** conveys the emotional tone or atmosphere of the scene, which can range from "exploration" for an astronaut to "mystery" for Sherlock Holmes. Sentiment enhances the engagement by imbuing the scene with an appropriate mood, helping students connect emotionally with the persona and fostering a deeper, empathetic understanding. The generated target results for each case are shown in figure 4. Some samples with a breakdown of the guided prompt stucture behind them are provided in table 1, with corresponding resulting prompt full texts provided in the Appendix A.

- **Historical Figure: Galileo Galilei**. Figure 4c. This prompt captures Galileo's identity as an astronomer during the *Renaissance*, a time of groundbreaking discoveries in science. By including specific details like his attire, telescope, and starry background, the generated image situates Galileo within his historical context, allowing students to engage visually with his contributions to astronomy.
- **Astronaut in Space**. Figure 4d. Designed for STEM education, this prompt captures the iconic image of an astronaut in space. The details of a spacesuit, reflective helmet, and Earth's visibility create an authentic visual, sparking interest in space science and exploration. Such an image can inspire curiosity about astronomy, physics, and technology, allowing students to "step into" the role of an astronaut.
- **Scientist in a Laboratory: Marie Curie**. Figure 4e. This prompt creates a visualization of Marie Curie in her early 20th-century laboratory environment, highlighting her identity as a trailblazer in radioactivity research. The prompt specifies visual cues, such as her attire and laboratory equipment, helping students gain a more personal and immersive understanding of Curie's scientific environment and achievements.
- **Cultural Icon: Cleopatra**. Figure 4f. Cleopatra's portrayal with traditional Egyptian attire and royal surroundings brings the ancient Egyptian world to life. The use of symbolic details—such as the golden headdress, hieroglyphics, and Nile River—enables students to engage with history in a way that feels vivid and personal. This prompt supports the study of ancient cultures by making historical figures visually accessible.
- **Fictional Character: Sherlock Holmes**. Figure 4g. Sherlock Holmes serves as a recognizable example of a fictional character that brings literature and deduction skills to life. By generating an image with iconic elements like the deerstalker hat and magnifying glass, the prompt reinforces his identity as a detective. The foggy Victorian street setting further enhances the sense of mystery, making literature and critical thinking more tangible to students.
- **Inventor in Workshop: Leonardo da Vinci**. Figure 4h. Leonardo da Vinci's portrayal highlights his creative process within a *Renaissance* workshop setting. Details like sketches, tools, and a flying machine emphasize his role as an inventor

Table 1: Breakdown of prompt structure detailing each setting, persona, time period, place, associated objects and the sentiment conveyed, illustrating how each element contributes to the accuracy of the generated visuals in figure 4.

Setting	Persona	Time	Place	Objects	Sentiment
Renaissance	Galileo	16th Century	Observatory	Telescope, charts	Neutral
Space	Astronaut	Contemporary	Earth Orbit	Helmet, stars	Exploration
Science	Marie Curie	1900s	Lab	Equipment, test tubes	Pioneering, focus
Egypt	Cleopatra	Ancient Egypt	River Palace	Jewels, columns	Ancient grandeur
Sleuth	Sherlock Holmes	Victorian Era	London	Magnifying glass, lamps	Mystery, atmospheric
Inventor	Leonardo da Vinci	Renaissance	Workshop	Flying machine	Focus, creativity

and artist, illustrating the interdisciplinary nature of his contributions. This prompt allows students to explore creativity and innovation by visualizing da Vinci's world, making history, science, and art interconnect in a tangible way.

4 Application Notes and Limitations

The integration of AI-driven role embodiment into Mixed Reality offers a wide range of applications across educational and cultural fields, with particular potential in the study and preservation of traditional cultural heritage, such as for instance, traditional dance styles as depicted in figure 1. In cultural and historical studies, this method provides a tool for exploring global dance traditions that might otherwise be difficult to access. For instance, by embodying a Kabuki dancer or a Renaissance court performer, students gain insights into the historical contexts, movements, and societal roles these dances represent. Such applications support experiential learning, allowing students to *"step into"* the roles of dancers from various eras and regions, developing empathy and creating a more personal connection to cultural heritage. Additionally, museums, cultural institutions, and performing arts centers can utilize these tools to enhance visitor engagement. By offering mixed reality experiences that showcase historically accurate dance forms, these organizations can provide audiences with deeper, more interactive encounters with cultural traditions.

In classroom settings, AI-driven role embodiment provides students with opportunities to experience historical events, explore scientific concepts, and understand diverse cultural perspectives firsthand (see figure 4. By embodying figures from history, literature, and science, students can gain deeper insights into subjects that may otherwise feel abstract. For instance, embodying scientists like Marie Curie in a laboratory setting or historical figures in their traditional environments encourages a tangible understanding of their contributions, making learning more relatable and memorable. Mixed reality experiences thus create a space where theoretical knowledge is reinforced through interactive, experiential learning.

Finally, corporate training, team-building exercises, and professional development programs can also benefit from this tool. By allowing participants to embody roles in historical and culturally significant contexts, organizations can improve empathy, cultural awareness, and interdisciplinary skills. In fields like tourism, hospitality, and international business, such experiences may enhance employees' understanding of global cultures, helping them engage more effectively with diverse clients and collaborators.

A limitation of the extent of the embodiment realized by students in this solution lies in the relative performance of image generation using the underlying Stable Diffusion architecture. Here the teacher is limited to relatively infrequency updates of the augmented mirror video display as snapshots with no less than 10 seconds elapsing between newly presented augmentations. Whilst a dance teacher for example can provide formal individual pose guidance through snapshots shown to the students. A future direction of this work would see value to explore live update of augmentations to provide a continuous movement of the generative AI scenarios, for example to illustrate the actual flow of dance motions over time in live embodiment. Achieving real-time generative outputs presents significant technical challenges due to the computational demands of large-scale diffusion models. Rendering continuous, high-fidelity motion requires substantial optimization of inference pipelines and enhanced hardware capabilities, which are areas of active research in the broader field of generative AI. Despite this, the current implementation effectively supports pose guidance through keyframe updates, which are sufficient to convey essential movement details for educational scenarios, allowing students to focus on mastering specific poses and transitions during structured intervals. Future development efforts will aim to overcome these limitations by enabling live augmentations with higher temporal resolution. This would allow for the visualization of continuous movement, such as illustrating the flow of dance motions over time, significantly enhancing the dynamic and immersive aspects of the embodiment experience. Such advancements would extend the utility of the system beyond keyframe-based guidance, enabling broader applications in real-time motion-intensive educational contexts.

5 Conclusion

This article has presented a framework for generative AI role embodiment with augmented mirror displays for the classroom as a tool for enhancing role-based learning in educational contexts. By allowing students to visually embody historical, scientific, and cultural figures through prompt-driven image generation, these models support a dynamic approach to learning that extends beyond traditional methods. The tailored use of prompt design and *ControlNet*'s precision in controlling outputs enables educators to generate visuals that are contextually accurate, culturally relevant, and aligned with specific learning objectives. Our findings show promise in that role embodiment through generative AI has the potential to increase student engagement and assist on promoting deeper understanding of diverse subject matter through visual representations. The nature of this approach may lead to improvements in empathy, memory retention, and active participation, as students connect more personally with the material. Applications across multiple disciplines, from STEM to humanities, illustrate the versatility of these tools in creating customized educational experiences that assist on a diverse curricular needs.

Nonetheless, this study also emphasizes the importance of ethical considerations when using generative AI in education. Issues related to representation, cultural sensitivity, and embedded biases within AI models must be carefully managed to avoid reinforcing stereotypes and providing inaccurate portrayals. Implementing ethical guidelines and rigorous prompt design can mitigate these risks, ensuring that AI-generated content respects diversity and maintains educational value. Future work may explore the long-term effects of AI-driven role embodiment on student outcomes, as well as strategies for ethical AI integration in educational content creation and further accuracy of control mechanisms for generative-AI augmentations both in factual and visual terms.

Acknowledgments

Stable Diffusion was utilized to generate figures and results of this work. This project has received funding from European Union's Horizon 2020 research and innovation programme under grant agreement No. 101017779.

References

Mousa Al-kfairy, Dheya Mustafa, Nir Kshetri, Mazen Insiew, and Omar Alfandi. 2024. Ethical Challenges and Solutions of Generative AI: An Interdisciplinary Perspective. *Informatics* 11, 3 (2024). https://doi.org/10.3390/informatics11030058

Joy Buolamwini and Timnit Gebru. 2018. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* 2018)*. ACM, 77–91. https://proceedings.mlr.press/v81/buolamwini18a.html

Chen Cao, Zijian Ding, Gyeong-Geon Lee, Jiajun Jiao, Jionghao Lin, and Xiaoming Zhai. 2023. Elucidating STEM Concepts through Generative AI: A Multi-modal Exploration of Analogical Reasoning. arXiv:2308.10454 [cs.AI] https://arxiv.org/abs/2308.10454

José David Carnicero Pérez, Ricardo Moreno Rodriguez, and Nerea Felgueras Custodio. 2023. Development of empathy and ethical values through role-playing games as innovation for education in values. *IJERI: International Journal of Educational Research and Innovation* (05 2023), 109–122. https://doi.org/10.46661/ijeri.7273

Google Cloud. 2024. Speech-to-Text AI. https://cloud.google.com/speech-to-text.

Nassim Dehouche and Kullathida Dehouche. 2023. What's in a text-to-image prompt? The potential of stable diffusion in visual arts education. *Heliyon* 9, 6 (2024/11/04 2023). https://doi.org/10.1016/j.heliyon.2023.e16757

Jonathan Ho, Ajay Jain, and Pieter Abbeel. 2020. Denoising Diffusion Probabilistic Models. *Advances in Neural Information Processing Systems* 33 (2020), 6840–6851. https://doi.org/10.48550/arXiv.2006.11239

Dominic Kao and D. Fox Harrell. 2021. Exploring the Use of Role Model Avatars in Educational Games. *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment* 11, 3 (Jun. 2021), 29–35. https://doi.org/10.1609/aiide.v11i3.12814

Leelo Keevallik. 2010. Bodily Quoting in Dance Correction. *Research on Language and Social Interaction* 43, 4 (2010), 401–426. https://doi.org/10.1080/08351813.2010.518065 arXiv:https://doi.org/10.1080/08351813.2010.518065

Chengyi Liu, Wenqi Fan, Yunqing Liu, Jiatong Li, Hang Li, Hui Liu, Jiliang Tang, and Qing Li. 2023. Generative Diffusion Models on Graphs: Methods and Applications. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence (IJCAI-23)*, Edith Elkind (Ed.). International Joint Conferences on Artificial Intelligence Organization, 6702–6711. https://doi.org/10.24963/ijcai.2023/751 Survey Track.

Diego Miguel-Revilla, Mercedes Calle Carracedo, and María Sánchez-Agustí. 2021. Fostering engagement and historical understanding with a digital learning environment in secondary education. *E-Learning* 18 (06 2021), 344–360. https://doi.org/10.1177/2042753020957452

Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. 2021. Zero-Shot Text-to-Image Generation. *OpenAI Research* (2021). https://doi.org/10.48550/arXiv.2102.12092

René Ranftl, Katrin Lasinger, David Hafner, Konrad Schindler, and Vladlen Koltun. 2020. Towards Robust Monocular Depth Estimation: Mixing Datasets for Zero-shot Cross-dataset Transfer. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)* (2020).

René Ranftl, Katrin Lasinger, David Hafner, Konrad Schindler, and Vladlen Koltun. 2022. Towards Robust Monocular Depth Estimation: Mixing Datasets for Zero-Shot Cross-Dataset Transfer. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44, 3 (2022).

Saman Rizvi, Jane Waite, and Sue Sentance. 2023. Artificial Intelligence teaching and learning in K-12 from 2019 to 2022: A systematic literature review. *Computers and Education: Artificial Intelligence* 4 (2023), 100145. https://doi.org/10.1016/j.caeai.2023.100145

Samsinar Samsinar and Fitriani Fitriani. 2020. Character-Based Learning and self-development to improve the students' Character Education. *Lentera Pendidikan : Jurnal Ilmu Tarbiyah dan Keguruan* 23 (06 2020), 108. https://doi.org/10.24252/lp.2020v23n1i10

Alison Singer, Georgina Montgomery, and Shannon Schmoll. 2020. How to foster the formation of STEM identity: studying diversity in an authentic learning environment. *International Journal of STEM Education* 7, 1 (2020), 57. https://doi.org/10.1186/s40594-020-00254-z

David Franklin Swink. 1993. Role-Play Your Way to Learning. *Training & Development* 47 (1993), 91–97. https://api.semanticscholar.org/CorpusID:108246361

Dara Tafazoli. 2024. Exploring the potential of generative AI in democratizing English language education. *Computers and Education: Artificial Intelligence* 7 (2024), 100275. https://doi.org/10.1016/j.caeai.2024.100275

Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. 2023. Adding Conditional Control to Text-to-Image Diffusion Models. arXiv:2302.05543 [cs.CV] https://arxiv.org/abs/2302.05543

A Dance Educator Generative AI Augmentation Prompt Samples

- **Historical Figure: Galileo Galilei**. Figure 4c. *Prompt: "A portrait of Galileo Galilei, Italian Renaissance astronomer, wearing period-accurate 16th-century attire, standing beside a telescope under a starry sky with ancient astronomical charts in the background."*
- **Astronaut in Space**. Figure 4d. *Prompt: "A modern astronaut floating in outer space with Earth visible in the background, wearing a spacesuit with reflective helmet, surrounded by stars, representing space exploration and science."*
- **Scientist in a Laboratory: Marie Curie**. Figure 4e. *Prompt: "Marie Curie in a 1900s laboratory setting, surrounded by lab equipment and test tubes, wearing a long black dress, with focused expression, emphasizing her role as a pioneering scientist in chemistry and physics."*
- **Cultural Icon: Cleopatra**. Figure 4f. *Prompt: "Cleopatra, the ancient Egyptian queen, wearing traditional Egyptian attire with a golden headdress and jewelry, standing in a palace with ornate columns, hieroglyphics on walls, and Nile River visible in the distance."*
- **Fictional Character: Sherlock Holmes**. Figure 4g. *Prompt: "Sherlock Holmes, the iconic detective, in Victorian London, wearing a deerstalker hat and holding a magnifying glass, with a foggy street and gas lamps in the background, emphasizing mystery and investigation."*
- **Inventor in Workshop: Leonardo da Vinci**. Figure 4h. *Prompt: "Leonardo da Vinci in a Renaissance workshop, surrounded by sketches, tools, and a partially built flying machine, wearing a robe with ink stains, focused expression, representing creativity and innovation."*