The Roleplay Mirror: Leveraging Digital Avatars to Address Discomfort in Sensitive Roleplays

JOEL HARMAN, Queensland University of Technology, Australia
NELLI HOLOPAINEN, Queensland University of Technology, Australia
ALESSANDRO SORO, Queensland University of Technology, Australia

Roleplay is a powerful tool for exploring social dynamics, particularly in education and training on sensitive topics like gender and power dynamics in STEM. However, traditional methods often cause discomfort when participants must embody roles that differ from their own identities. The Roleplay Mirror, an augmented reality (AR) system, addresses this by allowing participants to project their actions and voices onto digital avatars, reducing discomfort and enabling more authentic engagement. The system also facilitates the recording and review of role-plays in three-dimensional spaces, enhancing their educational value and fostering empathy through perspective-taking. Initial feedback highlights both the potential of the system and the challenges of using digital avatars for representation. This paper details the design, implementation, and early insights from the Roleplay Mirror, contributing to the development of virtual roleplay systems that are both effective and ethically sensitive.

CCS Concepts: • **Human-centered computing** → *Interactive systems and tools*; **Collaborative interaction**; *Collaborative and social computing systems and tools*; • **Software and its engineering** → *Designing software.*

Additional Key Words and Phrases: Roleplay, Augmented Reality, Motion Capture, Avatar Embodiment, Interactive Storytelling

ACM Reference Format:
Joel Harman, Nelli Holopainen, and Alessandro Soro. 2024. The Roleplay Mirror: Leveraging Digital Avatars to Address Discomfort in Sensitive Roleplays. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (OzCHI '24).* ACM, New York, NY, USA, 13 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Roleplay has long been recognized as a powerful method for exploring complex social dynamics, allowing participants to step into different roles and engage with scenarios that challenge their perspectives [2]. In educational and training contexts, roleplay is particularly valuable for simulating real-world situations that involve nuanced interactions, such as navigating gender and power dynamics in STEM higher education [1]. However, traditional roleplay methods often encounter significant challenges. Participants may feel uncomfortable or reluctant to embody characters that do not reflect their own identities, leading to a diminished effectiveness of the roleplay exercise [15].

To address these challenges, this initial, in-progress research aims to develop and evaluate a virtual roleplay system that can provide a more effective, immersive, and psychologically safe environment for participants. We introduce the "Roleplay Mirror," an augmented reality (AR) tool designed to facilitate roleplay experiences by allowing participants

Authors' Contact Information: Joel Harman, ja.harman@qut.edu.au, Queensland University of Technology, Brisbane, Queensland, Australia; Nelli Holopainen, nelli.holopainen@qut.edu.au, Queensland University of Technology, Brisbane, Queensland, Australia; Alessandro Soro, alessandro.soro@qut.edu.au, Queensland University of Technology, Brisbane, Queensland, Australia.

Manuscript submitted to ACM

to project their actions and voices onto digital avatars. This system offers a crucial degree of separation between the participants and the roles they assume, reducing discomfort and enabling more authentic engagement with the scenarios.

A key focus of this system is its ability to construct role-plays that are designed to be viewed by third parties, rather than just the individuals who create them. By enabling others to observe these role-plays, the Roleplay Mirror seeks to foster empathy and a deeper understanding of complex social issues [8]. This exacerbates the need to separate the role-player from the scenario, as participants may be particularly sensitive about being personally associated with the roles they portray, especially when these scenarios are intended for external audiences.

The need for virtual roleplay systems is underscored by the increasing complexity of social dynamics in various fields, such as STEM education, where power imbalances and gender issues can significantly impact learning and professional environments [10]. Traditional roleplay, while valuable, often fails to fully engage participants due to the psychological barriers of directly associating with the roles they must play. The Roleplay Mirror addresses this by leveraging advanced motion capture and AR technologies to create an environment where participants can interact with digital avatars, preserving their anonymity and allowing for a more flexible and impactful roleplay experience.

However, our early findings also reveal that some participants found it unnerving to have their words and actions represented by a digital avatar. This reaction highlights the complexity of balancing immersion and psychological comfort in virtual roleplay systems. As this research progresses, we will continue to explore these nuances to better understand how to create a system that is both effective and comfortable for all users.

Overall, this work aims to provide three main contributes to this research area:

(1) **Prototype Design**: We present the design and implementation of the Roleplay Mirror, a novel AR roleplay tool that integrates ZED cameras for real-time motion capture and audio processing. This system captures participants' movements and voices, applying them to digital avatars in real-time, creating an immersive and interactive storytelling environment.
(2) **User Feedback**: We present preliminary feedback from users who have interacted with the Roleplay Mirror, highlighting both the strengths and areas for improvement in the current system. While many users reported a greater sense of comfort and immersion, some found the use of digital avatars to represent their words and actions to be disconcerting, particularly given the system's focus on creating scenarios intended for third-party viewing.
(3) **Design Insights**: Through the development of the Roleplay Mirror, we gained valuable insights into the design challenges and considerations necessary for creating effective AR roleplay tools. These insights include the importance of seamless motion capture, the need for intuitive user interfaces, and the role of avatar customization in enhancing user engagement.

By addressing the challenges associated with traditional roleplay, the Roleplay Mirror represents a significant step forward in the development of virtual roleplay systems that support immersive, yet psychologically safe, roleplay experiences. This paper outlines the technological underpinnings of the system, shares the insights gained from its early development, and provides initial user feedback, laying the groundwork for future iterations and broader applications in educational and professional training contexts.

2 Related Work

The development of virtual roleplay systems and the integration of advanced technologies such as real-time body capture and text-to-speech (TTS) have significantly impacted various fields, including education, therapy, and social dynamics research. These systems offer innovative ways to simulate and analyze complex scenarios, enhancing both user experience and research outcomes. This section explores existing research in several key areas relevant to our work: virtual roleplay systems, multi-user virtual roleplays, the use of roleplay for perspective-taking, and advancements in real-time body capture and TTS technologies. By examining these areas, we aim to contextualize our contributions and highlight how our approach builds upon and extends current knowledge.

2.1 Challenges of Roleplaying Mismatched Perspectives

Roleplaying perspectives that differ significantly from one's own identity presents unique psychological challenges, particularly in sensitive scenarios such as gender and power dynamics. This discomfort, often referred to as identity dissonance, can hinder participants' engagement and effectiveness in the roleplay [15]. Participants may feel uneasy when embodying roles that conflict with their self-concept, leading to forced or insincere performances that diminish the impact of the exercise [2].

To mitigate these challenges, the Roleplay Mirror uses digital avatars, allowing participants to project their actions onto characters that provide a layer of anonymity. This separation reduces the discomfort of roleplaying unfamiliar perspectives and supports more authentic engagement [8]. However, the ethical considerations of asking individuals to roleplay potentially distressing roles remain significant. Future research should continue to explore ways to enhance participant comfort and engagement, such as through customizable avatars or guided post-roleplay reflections.

2.2 Virtual Roleplay Systems

Virtual roleplay systems have been developed to facilitate immersive simulations in a wide range of domains, including education, training, and therapy. These systems leverage virtual environments to create scenarios where users can engage in roleplay without the constraints of physical space or the need for elaborate setups. Early work by Gee [3] emphasized the educational potential of video games as roleplay environments, where players could explore different identities and scenarios in a controlled yet engaging manner. More recent systems, such as VR roleplay platforms, have enhanced this experience by adding elements of immersion and interactivity. For example, the "Virtually Better" platform explored by Rizzo et al. [11] utilized VR for therapeutic roleplay, helping users confront and manage anxiety in simulated environments. These systems, however, often focus on individual experiences rather than the social dynamics explored in the Roleplay Mirror.

2.3 Multi-user Virtual Roleplay

Multi-user virtual roleplay systems extend the concept of single-user environments by enabling multiple participants to interact within the same virtual space. These systems are often used in collaborative training scenarios where participants must work together to solve problems or navigate complex social interactions. For example, Molka-Danielsen et al. [6] examined the use of Second Life for collaborative learning, demonstrating how virtual worlds can facilitate social interaction and collaborative problem-solving.

2.4 Roleplay for Perspective Taking

Roleplay has been recognized as a powerful tool for perspective-taking, allowing individuals to step into the shoes of others and experience situations from different viewpoints. This technique is particularly useful in educational and therapeutic settings, where it can be used to foster empathy and understanding. Research has also demonstrated that perspective-taking through roleplay could reduce biases and promote more inclusive attitudes [2]. In virtual environments, roleplay for perspective-taking has been further explored through the use of avatars that differ from the participant's own identity. For instance, researchers found that embodying an avatar of a different race or gender could influence participants' attitudes and behaviors, demonstrating the potential of virtual roleplay to challenge and expand users' perspectives [15].

2.5 Real-Time Body Capture

Real-time body capture technology, such as the use of ZED cameras, has seen significant advancements in recent years. These systems are capable of capturing and processing a user's body movements in real-time, allowing for accurate representation within virtual environments. The ZED camera, in particular, offers depth-sensing capabilities that enable detailed motion capture without the need for extensive setup or calibration.

Researchers have explored the integration of ZED cameras in VR systems, highlighting their ability to enhance user immersion by providing realistic and responsive avatar movements and interactions [5]. ZED cameras have also been effectively used for tracking physical objects, rather than user bodies, using virtual markers [12].

2.6 Realtime High-Quality TTS

Real-time Text-to-speech (TTS) technology has undergone significant advancements, evolving from basic robotic voices to highly realistic speech synthesis. Early TTS systems were limited by their mechanical sound and lack of natural intonation, which often detracted from the user experience. However, modern TTS methods, such as those employed by ElevenLabs, have revolutionized the field by using deep learning models to generate speech that closely mimics human intonation, rhythm, and emotion. Tacotron 2 by Shen et al. [13] is one such model that has significantly improved the quality and naturalness of synthetic speech. Another notable development is the work by Ping et al. [9] on the Deep Voice system, which introduced end-to-end models for TTS that can generate high-fidelity, emotionally expressive speech.

The quality and easily availability of real-time high-quality TTS has led to many researchers considering the potential impacts within several domains. For example, it has been considered as potentially valuable in education [14], interactive language learning [4], and video games [7].

3 Research Questions

Developing a multi-user AR roleplay system that integrates real-time body capture and text-to-speech (TTS) technologies requires overcoming significant technical challenges. The complexity of ensuring low-latency processing, synchronization across multiple users, and seamless integration between motion capture and speech synthesis is crucial to maintaining an immersive experience.

RQ1: What are the technical challenges and design considerations in developing a multi-user AR roleplay system that incorporates real-time body capture and text-to-speech technologies?

The realism and responsiveness of virtual environments are key to user immersion and engagement. Real-time body capture and advanced TTS technologies promise to enhance the authenticity of roleplay experiences, but their actual impact on user perceptions needs to be understood.

RQ2: How does the incorporation of real-time body capture and text-to-speech affect the overall user experience and perceived authenticity of roleplay in virtual environments?

Roleplaying sensitive scenarios, such as those involving gender and power dynamics in STEM education, can lead to discomfort for participants. Digital avatars provide a layer of separation that may help reduce this discomfort, enabling participants to engage more fully in these challenging discussions.

RQ3: To what extent does the separation provided by digital avatars in AR roleplay reduce participant discomfort and increase engagement in scenarios involving sensitive topics, such as gender and power dynamics in STEM education?

Exploring these questions is essential to advancing the design and effectiveness of AR roleplay systems, ensuring they provide immersive, authentic, and psychologically safe environments for participants to engage with complex social scenarios.

4 System Design

In this section, we provide a summary of the overall design process for the Roleplay Mirror and Roleplay Planner systems. The development followed an iterative approach, with each stage informed by user feedback and ongoing testing. We began with a foundational prototype that prioritized core functionalities, such as real-time avatar tracking and motion capture. Through multiple iterations, we refined the user interface, enhanced the system's accuracy, and introduced additional features like avatar customization and the Roleplay Planner tool. This process was driven by the goal of creating an intuitive, immersive, and flexible system that could effectively support complex role-play scenarios, particularly in educational and training contexts. As many versions of this tool are discussed, an overview of the features in each stage are outlined in Figure 1 below.

4.1 Design Discussions

During the development of the Roleplay Mirror system, we engaged in extensive discussions with potential user groups, particularly HCI researchers interested in high-impact role-play scenarios. These conversations were crucial in shaping the design and functionality of the system, ensuring it would meet the specific needs and sensitivities of its intended users.

One of the primary groups we consulted with was researchers focused on navigating gender and power dynamics within STEM higher education. These scenarios are often complex and sensitive, requiring participants to engage with topics that can be uncomfortable or challenging. Through our discussions, it became clear that many participants are reluctant to role-play such scenarios, particularly when the roles they must assume do not align with their personal identities. The prospect of being recorded while taking on a role that contrasts with their self-image added another layer of discomfort, which could potentially hinder their engagement and the effectiveness of the role-play.

To address these concerns, we emphasized the use of virtual avatars within the Roleplay Mirror system, which allows participants to engage in role-plays without directly associating themselves with the characters they portray. This approach was designed to provide a psychological buffer, enabling users to explore challenging scenarios more freely without the fear of being personally identified with the roles they were assigned.

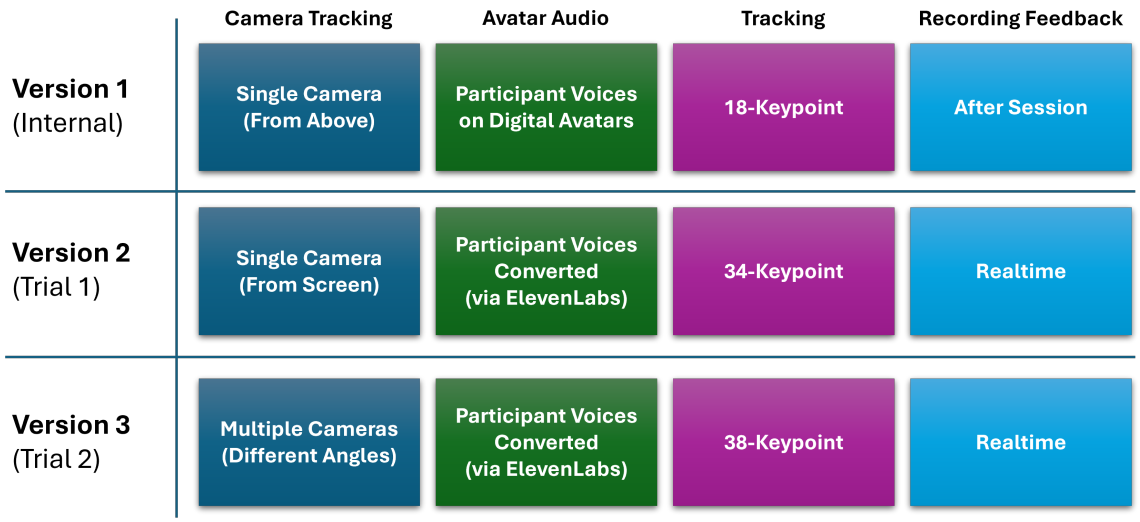

Fig. 1. An overview of the key features of the Roleplay Mirror at each of the three major versions. Version One was tested by the research team and other HCI researchers known by the research team. Version Two and Version Three were both tested with users outside of the research team. The four key dimensions that changed during iteration of the tool were: (1) the camera position and location for tracking, (2) the method for overlaying audio onto the virtual avatars, (3) the number of keypoints involved in the tracking, and (4) when the participants were able to see their actions on the virtual avatars.

Another key point that emerged from these discussions was the desire for a system that would allow users to revisit and explore the recorded scenarios after their creation. Researchers were particularly interested in how the perception of these role-plays might change when viewed through augmented reality (AR). They hypothesized that watching these scenarios unfold in a real-world setting, augmented with digital avatars, could potentially enhance the impact and realism of the experience. This led to the inclusion of features that not only allow scenarios to be recorded and reviewed in 3D but also enable others to engage with these recordings in various settings, offering a deeper exploration of the role-played dynamics.

These insights from our discussions were instrumental in guiding the development of the Roleplay Mirror, ensuring that it would be both a powerful tool for immersive role-play and sensitive to the needs of its users.

4.2 Version One

In the initial implementation of the Roleplay Mirror, body tracking was conducted using a single ZED 2i camera. This system employed an 18-keypoint tracking model that effectively captured the movement of major bones but lacked the precision to track finer details, such as individual finger movements. The system tracked users using a body-tracking machine learning model, rather than marker-based methods which can have non-trivial setup requirements. This ensured that the system would be easy for people to use, and would better allow a group of people to roleplay scenarios without needing to put on and take off markers for accurate tracking. As with most motion capture, a post-processing step was conducted after the recording to clean up the captured motion capture data.

The camera was mounted on a tripod and positioned to monitor all participants within the designated roleplay area, approximately a 3x3 meter space. To ensure comprehensive tracking, including the lower extremities, an angled overhead placement was selected, optimizing visibility of the entire body.

Fig. 2. An example visualisation simulating a roleplay created by the system. The current speaker is highlighted in yellow to improve clarity for the viewer. The viewer can move around the environment to view the situation, and the system uses spatial audio to improve the ability of the viewer to track the conversation.

Following the roleplay sessions, participants could review their performance using a Varjo XR-3 augmented reality headset. The replay environment offered three options: the real world, a digital twin background replicating the physical setting (if one existed for the location), or a generic background for the simulation (e.g. an office or a classroom). The body movements of the roleplay participants were mapped onto virtual avatars chosen by the participants before the roleplay. An example of this roleplay being visualised through the AR headset can be found in Figure 2 below.

After trialing the initial version, several key issues emerged. Test users expressed dissatisfaction with the delay between performing the roleplay and viewing the final results, as they wanted to confirm that the system was functioning correctly before starting the main session.

Additionally, participants were uncomfortable with having their voices "mapped onto" the virtual characters. They felt that the voices did not always align well with the avatars and that this approach compromised the level of anonymity they expected.

In response to this feedback, we quickly implemented real-time avatar movement simulation during the session. However, this solution was suboptimal, as the overhead camera angle used for tracking did not align with the participants' natural viewing perspectives, leading to a disconnect in their experience.

4.3 Version Two

In the second iteration of the Roleplay Mirror, we implemented significant modifications based on the feedback received from earlier trials. Instead of mounting the camera on a tripod above the scene, we introduced a large monitor into the roleplay space with the camera positioned directly above the screen. We also updated the body tracking model to instead use 34-keypoint tracking, improving the hand tracking capabilities of the system.

This setup also inspired the system's name, "Roleplay Mirror," as the monitor now provided a real-time reflection of participants' actions, with avatars dynamically anchored to their movements. This adjustment also enabled participants to select their preferred avatar prior to the recording, a feature they found particularly helpful during preparation. Figure 3 is an example showing the system before unique avatars have been applied to each person.

Fig. 3. This is an example of the Roleplay Mirror as viewed from the perspective of the users. Due to the position of the camera on the screen, it would act as a digital mirror, performing the actions on the assigned avatars in realtime. Note that during the roleplay, a clean frame is used to "hide" the participants, and the UI is also hidden to remove distractions.

We initially considered using voice actors to record the dialogue from the roleplays to explore how participants would respond to the removal of their own voices. However, this approach proved cumbersome to implement and lacked scalability.

As an alternative, we opted to transcribe the dialogue spoken during the sessions and convert it back to speech using ElevenLabs' text-to-speech software. This solution allowed participants to choose which voices to assign to each role-play participant and to refine the dialogue to their preference, enhancing both the customization and anonymity of the experience.

User testing for this version proved to be significantly more successful than the initial iteration. Participants reported a high level of enjoyment when interacting with the digital characters, and they particularly appreciated the immediacy with which their actions were reflected in the avatars, making the experience feel more immersive and convincing.

However, while the real-time feedback was well-received, the tracking accuracy still fell short of our desired quality. To address this, we planned to implement a post-processing step to refine the recorded animation data, ensuring a higher level of precision until the tracking system could be fully optimized.

4.4 User Trial One

After trialing Version 2 with other HCI researchers, we determined that the system had reached a level of maturity suitable for supporting role-play scenarios. Specifically, the tool was deployed by researchers to facilitate role-playing exercises focused on exploring gender and power dynamics within STEM higher education. This application was chosen as an initial use case, given that these researchers had been instrumental in guiding the system's initial design, making it a promising fit for their needs.

This user trial was conducted as part of a larger workshop. Participants were tasked with designing scenarios that examined challenging power dynamics within STEM higher education and then role-playing these scenarios using the system. Each scenario involved 2-4 participants in the "scene," each with speaking roles. Two groups participated in the

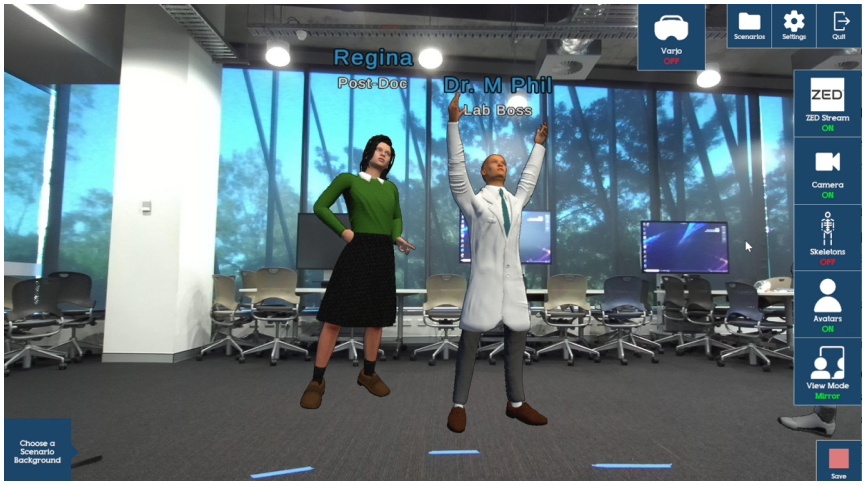

Fig. 4. An example frame captured during the roleplay. The participants conducting the roleplay are hidden using a clean frame, leaving only the virtual avatars which match the real-world movements of the users. They each have unique avatars associated with them which mirror the persona they have assumed for the roleplay.

trial, with each group creating a unique scenario. Participants were instructed to face the camera during the role-play and to avoid occluding each other, ensuring that the camera could accurately capture all individuals. Figure 4 shows can example of the roleplay in progress, with the people doing the roleplay hidden and their chosen virtual identities shown instead.

However, during this trial, we identified issues that had not surfaced during earlier testing. During the testing phase, participants primarily interacted with the system, whereas during the role-play, participants frequently engaged with each other. This shift in interaction led to situations where participants were often side-on or even fully turned away from the camera, creating significant challenges for accurate tracking and recording. A particularly common issue was participants facing side-on to the camera. For example, in a two-person role-play, participants would initially stand side-by-side but then turn to face each other once the dialogue began. This created difficulties for the system, as adjusting the camera's position would still result in occlusions.

Following the role-play sessions, participants were able to view their recordings using the Varjo XR-3 headset. Due to the tracking challenges encountered, a post-processing step was necessary to enhance the accuracy and authenticity of the recorded animations. After this step, participants could don the headset and experience the scenario they had role-played, complete with the chosen 3D models and voices.

Overall, participants found the process engaging and insightful. However, they noted that having more guidance in designing the scenarios—such as knowing in advance which avatars and voices were available—might have influenced their creative decisions and enhanced the overall experience.

4.5 Version Three

Based on the feedback from the user trials, we made further enhancements to improve the system's functionality. To address the accuracy issues, we integrated multiple ZED 2i cameras into the setup, linked together to enhance body tracking. While one camera remained positioned on the main display to maintain the mirror simulation, two additional

Fig. 5. The Roleplay Planner provides tools to help users construct their roleplay and make changes as necessary. Users can select the avatars they want to use, the voices they want the avatars to have, and the dialogue they want the avatars to speak. They can then watch a simulation of the scenario and make adjustments as needed. When they are happy with the scenario, then can then use the Roleplay Mirror to act out the scenario and add their own animations to the recording.

cameras were strategically placed around the room at different angles to significantly improve overall tracking precision. We also again updated the body tracking model to instead use 38-keypoint tracking, improving the hand tracking capabilities of the system once more.

To better support users in designing their scenarios, we developed an additional tool called the Roleplay Planner. This tool allowed participants to plan their role-plays by arranging virtual characters within a scene and specifying the dialogue for each participant. Once the scenario was set, users could press play to simulate the role-play, with each character delivering the pre-scripted lines, providing a comprehensive preview of the session before recording. Figure 5 shows can example of the Roleplay Planner being used to design a scenario for later roleplay and visualisation.

4.6 User Trial Two

Following these updates, we conducted another trial where participants were tasked with designing and simulating role-plays that addressed problematic power dynamics. With the updated system, participants could begin by creating a rough outline of the role-play, selecting their desired avatars and voices using the Roleplay Planner. They then physically enacted the scenario using the Roleplay Mirror, ultimately producing a finished role-play that could be viewed in augmented or virtual reality.

The Roleplay Planner proved particularly valuable in allowing users to develop and refine their scenarios. Unlike in the initial trial, where participants performed their first draft, the ease of iteration in this version enabled them to make multiple adjustments, such as refining dialogue that sounded stilted when spoken aloud. Participants also identified potential applications beyond the initial use case. One participant suggested that the tool could be useful for ensuring presentations sound natural, while others noted its potential for facilitating the development of realistic conversations between characters.

The enhanced accuracy of the Roleplay Mirror also led participants to envision a wider range of applications. Several participants noted its potential for recording dances or other performances. Additionally, one participant with expertise in gaming and digital effects highlighted the possibility of linking real-world actions to visual effects on the avatars, such as triggering a flash of light with a clap of hands.

4.7 Exhibit Trial

Building on the suggestions from previous participants, we decided to trial the system during a public event. For this event, the Roleplay Mirror was continuously running at a booth. As attendees passed by, they could see a body-tracked skeleton mesh of themselves displayed on the screen, along with instructions on how to interact with the system.

Users could initiate avatar selection by holding their hands together in front of them, and release their hands to stop. They could then record a brief snippet of their actions, which would be displayed to subsequent users. The system was met with high levels of engagement, likely due to its simplicity and ease of use, which required minimal investment and no complex controls. Additionally, the design's anonymity encouraged users to comfortably record their actions.

Given the success of this trial, the Roleplay Mirror shows promise for applications in playful systems and interactive exhibits.

5 Discussion

The ongoing development and iterative refinement of the Roleplay Mirror and Roleplay Planner have provided valuable preliminary insights into enhancing role-play experiences, particularly in scenarios involving complex social dynamics. While our findings are still emerging, this section discusses how the system potentially addresses the research questions posed earlier in the paper, with a focus on the system's ability to reduce discomfort in role-playing sensitive scenarios, the impact of real-time body capture and text-to-speech technologies on user experience, and the broader implications for educational and training contexts.

5.1 Benefits of Recording with Virtual Avatars

A key potential benefit of using virtual avatars in the Roleplay Mirror is the capacity to record role-play scenarios anonymously. This feature could allow participants to engage more fully without the self-consciousness that might arise from directly associating themselves with challenging roles, a concept directly linked to **RQ3**. The digital avatars offer a layer of separation that may help reduce psychological barriers, possibly fostering more open and honest engagement in complex scenarios. However, it's important to note that we have not yet conclusively shown that this separation effectively reduces discomfort, and further research is needed to substantiate these initial observations.

The system's capability to record role-plays in 3D, viewable through augmented reality, shows promise for enhancing the review and analysis process, as it allows scenarios to be experienced from multiple perspectives. This aligns with **RQ2**, as it may provide a richer understanding of the interactions that took place. The flexibility to update or modify recorded scenarios also suggests a valuable tool for iterative learning and development. Nevertheless, these are early insights, and more extensive testing is required to fully validate the system's effectiveness in these areas.

5.2 System Usefulness

The Roleplay Mirror was designed to improve the role-playing experience by allowing participants to project their actions onto digital avatars in real-time, potentially reducing the psychological barriers often associated with traditional role-play—a key concern of **RQ3**. The addition of the Roleplay Planner further expanded the system's utility by enabling

scenario pre-design, avatar selection, and dialogue scripting. These features suggest that the system can facilitate more immersive and convincing role-plays, as indicated by initial user feedback. However, it's important to emphasize that these conclusions are still preliminary, and further research will be necessary to confirm the system's effectiveness and utility.

5.3 Design Insights

Preliminary trials and user interactions have provided early insights that suggest potential answers to our research questions:

(1) Real-Time Feedback and Immersion: The Roleplay Mirror's capacity for immediate feedback through digital avatars appears to enhance user immersion and engagement, which may support the hypotheses of RQ2. However, these findings are not yet conclusive.
(2) Customization and Anonymity: The ability to customize avatars and voices may contribute to reducing discomfort and increasing engagement in sensitive scenarios, particularly those involving gender and power dynamics. This observation relates to RQ3, though further study is required to confirm its validity.
(3) Technical Challenges: Achieving precise tracking accuracy remains a significant challenge, as noted in RQ1. While the system has made progress, further refinement is needed, and these challenges underscore the importance of ongoing technical development.
(4) Broader Applications: Early feedback suggests that the Roleplay Mirror has potential applications beyond its original scope, including in playful systems and interactive exhibits. This aligns with our exploration of RQ2 and RQ3, though these applications will need to be explored in more depth in future work.

6 Conclusion and Future Work

The Roleplay Mirror and Roleplay Planner show considerable promise in enhancing role-play experiences, particularly in educational and training settings. Although our work is still in progress, the tools appear to address some of the challenges outlined in our research questions, offering potentially valuable solutions for role-playing sensitive scenarios. However, further research and extensive testing are needed to confirm these early findings and to fully understand the system's impact.

Future work will focus on addressing the technical challenges identified, exploring additional features such as sensory feedback (e.g., haptic feedback), and broadening the system's applications across various domains. As we continue to refine and develop the Roleplay Mirror, we aim to create a robust, versatile tool that can support a wide range of immersive and interactive experiences, ultimately contributing to the advancement of virtual roleplay systems.

Ethical Approval

This research was approved by the Queensland University of Technology Human Research Ethics Committee (Approval Number: 3254).

References

[1] Sara De Freitas and Tim Neumann. 2009. The use of 'exploratory learning'for supporting immersive learning in virtual environments. *Computers & Education* 52, 2 (2009), 343–352.
[2] Adam D Galinsky and Gordon B Moskowitz. 2000. Perspective-taking: decreasing stereotype expression, stereotype accessibility, and in-group favoritism. *Journal of personality and social psychology* 78, 4 (2000), 708.
[3] James Paul Gee. 2003. What video games have to teach us about learning and literacy. *Computers in entertainment (CIE)* 1, 1 (2003), 20–20.

[4] Jordi Linares Carrasquer. 2024. Interactive language learning application using artificial intelligence. (2024).
[5] Antoine Maiorca, Seyed Abolfazl Ghasemzadeh, Thierry Ravet, François Cresson, Thierry Dutoit, and Christophe De Vleeschouwer. 2024. Self-Avatar Animation in Virtual Reality: Impact of Motion Signals Artifacts on the Full-Body Pose Reconstruction. *arXiv preprint arXiv:2404.18628* (2024).
[6] Judith Molka-Danielsen and Mats Deutschmann. 2009. *Learning and teaching in the virtual world of Second Life.* Tapir Academic Press.
[7] Valeriu-Liviu Nițu. 2024. *Enhancing Player Immersion: Automatic AI Localisation of Romanian Dialogue in Video Games.* B.S. thesis. University of Twente.
[8] Tabitha C Peck, Sofia Seinfeld, Salvatore M Aglioti, and Mel Slater. 2013. Putting yourself in the skin of a black avatar reduces implicit racial bias. *Consciousness and cognition* 22, 3 (2013), 779–787.
[9] Wei Ping, Kainan Peng, Andrew Gibiansky, Sercan O. Arik, Ajay Kannan, Sharan Narang, Jonathan Raiman, and John Miller. 2018. Deep Voice 3: Scaling Text-to-Speech with Convolutional Sequence Learning. arXiv:1710.07654 [cs.SD] https://arxiv.org/abs/1710.07654
[10] E Ashby Plant, Amy L Baylor, Celeste E Doerr, and Rinat B Rosenberg-Kima. 2009. Changing middle-school students' attitudes and performance regarding engineering with computer-based social models. *Computers & Education* 53, 2 (2009), 209–215.
[11] Albert Rizzo, Greg Reger, Greg Gahm, JoAnn Difede, and Barbara O Rothbaum. 2009. Virtual reality exposure therapy for combat-related PTSD. *Post-traumatic stress disorder: Basic science and clinical practice* (2009), 375–399.
[12] Sergio Serra, Redouane Kachach, Ester Gonzalez-Sosa, and Alvaro Villegas. 2020. Natural user interfaces for mixed reality: Controlling virtual objects with your real hands. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, 712–713.
[13] Jonathan Shen, Ruoming Pang, Ron J Weiss, Mike Schuster, Navdeep Jaitly, Zongheng Yang, Zhifeng Chen, Yu Zhang, Yuxuan Wang, Rj Skerrv-Ryan, et al. 2018. Natural tts synthesis by conditioning wavenet on mel spectrogram predictions. In *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 4779–4783.
[14] Chaonan Xu et al. 2024. Integrating AI Tools into Teaching Practice: Unleash the Potential of Your AI Co-pilot. In *Conference Proceedings. The Future of Education 2024*.
[15] Nick Yee and Jeremy Bailenson. 2007. The Proteus effect: The effect of transformed self-representation on behavior. *Human communication research* 33, 3 (2007), 271–290.

Received 2 September 2024