DOI: https://doi.org/10.61762/pjstvol7iss2art003

PLATFORM - A Journal of Science & Technology

Received: 15 November 2024, Revised: 10 December 2024, Accepted: 27 December 2024, Published: 31 December 2024, Publisher: UTP Press, Creative Commons: CC BY 4.0

# EMPOWERING HUMAN-LIKE NON-PLAYER CHARACTER INTERACTIONS IN VIRTUAL REALITY THROUGH LARGE LANGUAGE MODELS

**Haarinesh Selvaraju[1]*,Noreen Izza Binti Arshad[1], Ahmad Mukhlason[2], Faizal Mahananto[2]**

[1]Department of Computer and Information Sciences, Universiti Teknologi PETRONAS, Malaysia
[2]Sepuluh Nopember Institute of Technology, Indonesia

*E-mail: haarinesh_20001163@utp.edu.my

## ABSTRACT

*This research paper explores the integration of Large Language Models (LLMs) into Virtual Reality (VR) environments to enhance human-like interactions with non-player Characters (NPCs) in professional training simulations. Current VR systems suffer from repetitive, scripted dialogues that lack naturalness and adaptability, reducing user engagement and learning outcomes. To address this, the study developed a proof-of-concept VR system using state-of-the-art LLMs, including GPT-4, Claude 3.5, and Large Language Model Meta AI (LLAMA), evaluated through the Massive Multitask Language Understanding (MMLU) benchmark. The Agile methodology was employed to iteratively refine the system based on user feedback, optimising NPC interactions for contextual relevance and realism. Results demonstrated significant improvements in naturalness, engagement, and context maintenance, with LLaMA-powered NPCs outperforming others in user acceptance testing. These findings underscore the potential of LLMs to revolutionise VR-based training by delivering lifelike, context-aware dialogues and provide a robust foundation for future research in AI-driven immersive environments.*

***Keywords:*** *Artificial intelligence, business networking, Claude, GPT, metaverse, LLaMA, multitask language understanding (MMLU), Unity*

## INTRODUCTION

Virtual Reality (VR) has rapidly advanced from a gaming and entertainment tool to a transformative platform for professional training and education. By simulating real-world scenarios, VR provides users with an immersive and risk-free environment in which to develop practical skills. However, the full potential of VR remains underutilised in professional training due to the limitations of non-player Characters (NPCs), which play a crucial role in creating interactive and engaging simulations. Current NPCs in VR environments are restricted by pre-scripted dialogues, lacking the ability to deliver dynamic, contextually relevant, and lifelike interactions. These shortcomings diminish user engagement and reduce the effectiveness of VR-based training modules.

Figure 1 presents the Venn diagram illustrating the integration of VR environments, Large Language Models (LLMs), and NPC systems. It highlights their interdependencies in creating a dynamic and immersive professional training ecosystem.

NPCs in existing VR systems struggle to maintain conversational relevance, contextual coherence, and adaptability to real-time user interactions. These challenges stem from their reliance on pre-scripted dialogues, often resulting in repetitive and disengaging user experiences. Moreover, the computational demands of integrating advanced Artificial Intelligence (AI) systems, such as LLMs, into VR environments pose significant obstacles, particularly in ensuring real-time performance and seamless interaction.

Although recent research demonstrates the potential of LLMs in generating dynamic, human-like dialogues, significant gaps remain. Table 1 summarises

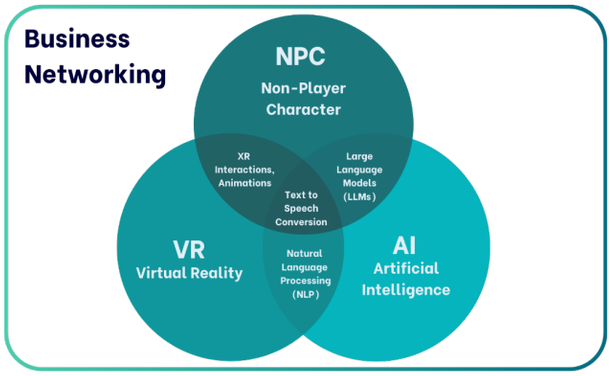


**Figure 1** Venn Diagram of the project's technology relationship

**Table 1** Key research gaps in LLM integration into VR

| Research Gap | Description |
|---|---|
| Real-Time Processing of LLMs | Need for efficient algorithms and computational resources. |
| Contextual Awareness and Coherence | Improving consistency and relevance in NPC dialogues. |

these research gaps, highlighting critical areas for improvement.

These gaps underscore the need for innovative solutions to enhance the realism and engagement of NPC interactions in VR-based training.

The objectives of this research are:

1. To evaluate current limitations in NPC Dialogue within VR Environments.
2. To develop and implement an LLM-Driven NPC system using the top three LLMs selected based on the MMLU benchmark.
3. To assess the effectiveness and user perception of the LLM-Driven NPC system in VR.

## LITERATURE REVIEW

### Historical Context and Evolution of Virtual Reality (VR)

VR has evolved dramatically since its early conceptualisation as a tool for entertainment and experimentation. The pioneering work of Morton Heilig, exemplified by his invention of the Sensorama (Figure 2), demonstrated the potential of multi-sensory immersion in creating engaging and interactive experiences [1]. Sensorama combined stereoscopic visuals, sound, vibration, and even scent to deliver an immersive experience, laying the groundwork for the integration of sensory modalities in VR systems. However, its technological limitations confined its applications primarily to controlled, static environments.

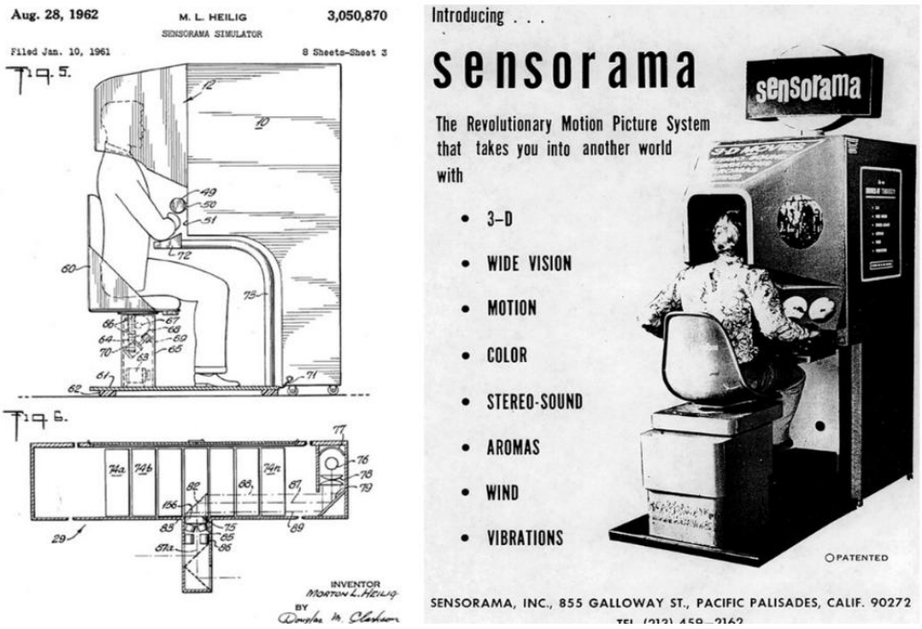


**Figure 2** Sensorama

(*Source*: Brutzman [1])

Subsequent advancements in VR technology focused on overcoming these constraints. Hilfert and König [2] demonstrated how modern VR systems could leverage cost-efficient hardware for practical applications such as engineering and construction, making immersive environments more accessible and functional. Despite these improvements, a persistent challenge remained: creating realistic interactions within VR environments. Traditional NPCs relied heavily on pre-scripted dialogues, which limited their ability to adapt dynamically to user input. This constraint became a significant barrier to using VR for professional training, where realistic, context-sensitive interactions are critical for effective learning.

**The Role of Artificial Intelligence (AI) in VR**

AI has been instrumental in advancing VR's ability to simulate lifelike environments and interactions. AI-driven NPCs have transitioned from being static, predictable entities to dynamic agents capable of adapting to user behaviours. Su [3] demonstrated how AI systems like Voice2Action enable real-time speech recognition and natural language understanding, allowing NPCs to respond contextually to user inputs. Similarly, Roberts et al. [4] highlighted the potential of prompt-based AI models to create fluid, unscripted dialogues that enhance NPC interactions.

While these early applications of AI in VR primarily focused on improving user engagement, they fell short of delivering the depth and contextual relevance required for professional scenarios. This gap paved the way for integrating LLMs, which excel at generating human-like, context-aware text responses.

**Large Language Models (LLMs)**

LLMs such as GPT-4, Claude 3.5, and LLaMA represent a transformative leap in AI's ability to generate natural, coherent, and context-sensitive dialogues. Unlike traditional AI systems, which rely on manually coded rules and datasets, LLMs are pre-trained on vast corpora of text, enabling them to understand complex linguistic patterns and generate relevant and adaptive responses.

Grow et al. [5] illustrated the effectiveness of GPT-based models in enhancing NPC interactions in gaming, where the ability to maintain conversational coherence over multiple exchanges is critical. Similarly, Hong et al. [6] explored the integration of LLMs into three-dimensional virtual environments, showcasing their ability to interpret spatial contexts and adapt their responses accordingly. These advancements make LLMs particularly well-suited for VR applications, where dynamic and context-aware interactions are essential.

**Advantages of Integrating LLMs into VR System**

The integration of LLMs into VR has revolutionised NPC design by enabling dynamic, lifelike dialogues that adapt to the user's inputs and contextual cues. Zhou et al. [7] emphasised that LLM-driven NPCs could transform professional training environments by simulating realistic conversational scenarios like business networking or project management. Unlike traditional pre-scripted NPCs, LLM-powered characters can adapt their tone, content, and behaviour to reflect the complexity of real-world interactions.

Mills [8] and Soliman et al. [9] further demonstrated the potential of LLM-driven NPCs in education and skill-building. LLM-powered NPCs can create personalised learning experiences that foster engagement and retention by tailoring responses based on user profiles. For example, in language learning simulations, NPCs can provide feedback and adjust their conversational style to suit the learner's proficiency level, making the experience both immersive and effective.

**Challenges in integrating LLMs into VR**

Despite their transformative potential, the integration of LLMs into VR systems is not without challenges. Real-time processing remains a significant hurdle, as LLMs require substantial computational resources to generate responses. Zhou et al. [7] and Gupta et al. [10] noted that latency during NPC interactions could disrupt immersion, particularly in professional training scenarios where timely responses are critical.

Maintaining conversational coherence over extended interactions is another challenge. Tan et al. [11] observed that NPCs often lose context during multi-turn dialogues, leading to disjointed or irrelevant responses. Addressing this issue requires the development of memory mechanisms within LLMs to retain and recall conversational context.

Furthermore, the scalability of LLM-powered NPC systems poses practical challenges. Patel [12] emphasised the importance of developing lightweight,

efficient models that can operate seamlessly in diverse VR environments without compromising performance. Additionally, there is a need for multilingual support, as most current LLMs are optimised for English, limiting their applicability in global training contexts.

### Applications of LLM-Driven NPCs in Professional Training

Using LLM-driven NPCs in professional training environments has shown promising results across various industries. In healthcare, NPCs powered by LLMs can simulate patient interactions, enabling medical professionals to practice diagnostic and communication skills in a risk-free environment. Soliman et al. [9] highlighted the role of VR in creating realistic training simulations that enhance both technical and soft skills. In business contexts, LLM-driven NPCs have been used to simulate networking events and client meetings, providing users with opportunities to practice negotiation, persuasion, and decision-making. Mills [8] demonstrated the effectiveness of these systems in teaching business communication, where the adaptability of NPCs allows users to experiment with different strategies and receive real-time feedback.

### Emerging Trends in LLM and VR Integration

Recent trends in LLM and VR integration focus on enhancing the realism and adaptability of NPC systems. Lee et al. [13] explored the integration of spatial audio and facial animations with NPC dialogues, creating a multi-sensory experience that enhances user immersion. Similarly, Zhou and Tan [14] investigated the use of adaptive learning models that allow NPCs to evolve based on user interactions, ensuring that the training experience remains dynamic and personalised.

Additionally, research by Zhang et al. [15] emphasises the importance of developing evaluation frameworks to assess the effectiveness of LLM-driven NPCs. Metrics such as naturalness, engagement, and learning outcomes are critical for refining these systems and ensuring their alignment with training objectives.

## METHODOLOGY

The research methodology for this project is designed to address the integration of LLMs into VR environments for realistic and effective business networking training. The study follows an iterative approach using the Agile methodology, enabling continuous refinement of the VR-AI system based on user feedback, performance metrics, and iterative testing. This section details the technological framework, research process, and evaluation strategy.

### Technological Framework

#### *Integration of LLMs*

The project incorporates advanced LLMs, including OpenAI's GPT-4, Claude 3.5, and Meta's LLaMA. These models were selected based on their performance in the Massive Multitask Language Understanding (MMLU) benchmark, which evaluates AI systems' linguistic and contextual understanding capabilities. Figure 3 highlights the comparative performance of these models in the MMLU benchmark, showcasing why they were chosen as the foundation for NPC development in this VR environment. Each model was assigned to specific NPCs with distinct roles to simulate real-world networking scenarios. Table 2 presents the NPCs and their respective roles.

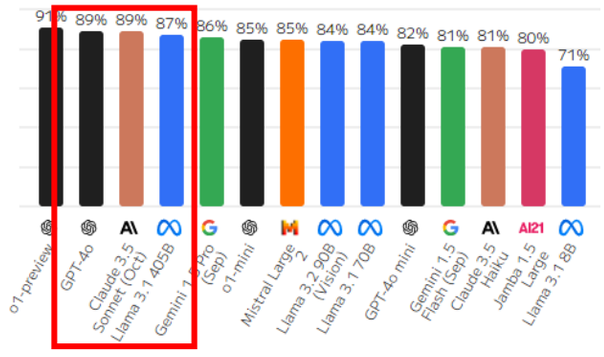


**Figure 3** Top 3 selected LLMs based on MMLU

(*Source*: artificialanalysis.ai [16])

**Table 2** Selected LLMs and their NPC role

| Non-Player Character | Role | Large Language Model |
|---|---|---|
| Intan | Talent Acquisition Specialist | GPT-4o |
| Wong | Digital & Technology Manager | Claude 3.5 Sonnet |
| Raj | Project Manager | LLaMA 3.1 405B |

Each NPC was specifically designed to address distinct professional networking scenarios, ensuring a diverse and enriched user experience. Intan, powered by GPT-4, acts as a Talent Acquisition Specialist, providing career advice on topics such as resume improvement, interview preparation, and professional growth. Wong, supported by Claude 3.5, serves as a Digital and Technology Manager, focusing on discussions about emerging technology trends, their implications for businesses, and practical applications in various industries. Lastly, Raj, driven by LLaMA, takes on the role of a Project Manager, simulating scenarios that emphasise project planning, execution, and problem-solving. Together, these NPCs create a comprehensive and dynamic networking environment that reflects real-world professional interactions and challenges.

***VR Environment Setup***

The VR environment was developed using Unity 3D, a versatile platform renowned for its robust 3D modelling and rendering capabilities. This environment simulates realistic business networking settings, including virtual conference rooms, informal meeting areas, and exhibition halls designed to provide an immersive and engaging experience. The environment incorporates interactive elements such as clickable objects to initiate conversations with NPCs, spatial navigation to allow users to move freely within the virtual space, and context-specific NPC interactions tailored to the unique characteristics of each virtual setting. The development process for this VR environment is illustrated in Figure 4, which outlines the sequence of steps involved in integrating 3D assets, NPC systems, and AI components, ensuring a cohesive and immersive virtual experience.

***Dialogue Management System***

A comprehensive Dialogue Management System was developed to enable seamless communication between users and NPCs, integrating several advanced AI-driven components. The system begins with Speech-to-Text (STT), which uses Google Cloud Speech-to-Text to convert user speech into text in real time, allowing immediate input processing. This is followed by Natural Language Understanding (NLU), which analyses the user's input to identify intent, sentiment, and contextual relevance, ensuring the system accurately interprets user commands or questions. The Dialogue Generation component leverages LLMs to create tailored responses, ensuring dynamic, adaptive, and human-like interactions. Finally, the system employs Text-to-Speech (TTS), powered by Google Cloud TTS, to vocalise NPC responses naturally and clearly, enhancing the conversational flow.

As illustrated in Figure 5, the dialogue management system processes user inputs through a seamless pipeline, from speech recognition to response delivery, creating fluid and responsive interactions. In addition to processing dialogue, the system incorporates synchronised NPC gestures, lip-syncing, and spatial audio to enhance realism. For instance, NPCs dynamically adjust their gaze and facial expressions based on the user's position and the conversation content. This synchronisation improves the immersion and reinforces the lifelike quality of NPCs, making interactions more engaging and contextually relevant.

***Integration of AI and VR Systems***

Integrating the AI-driven dialogue system with the VR environment required meticulous coordination to

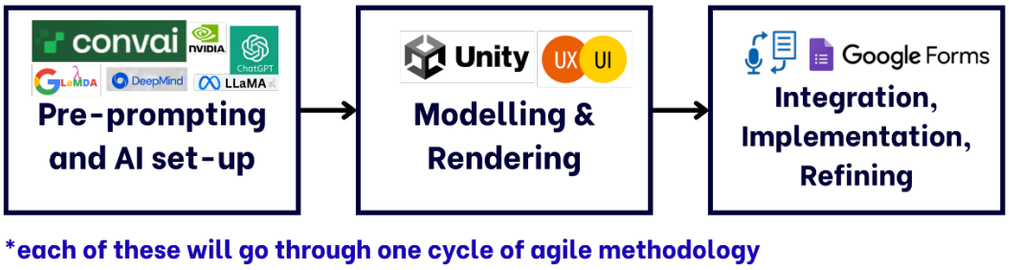


**Figure 4** Sequence of development process in VR environment

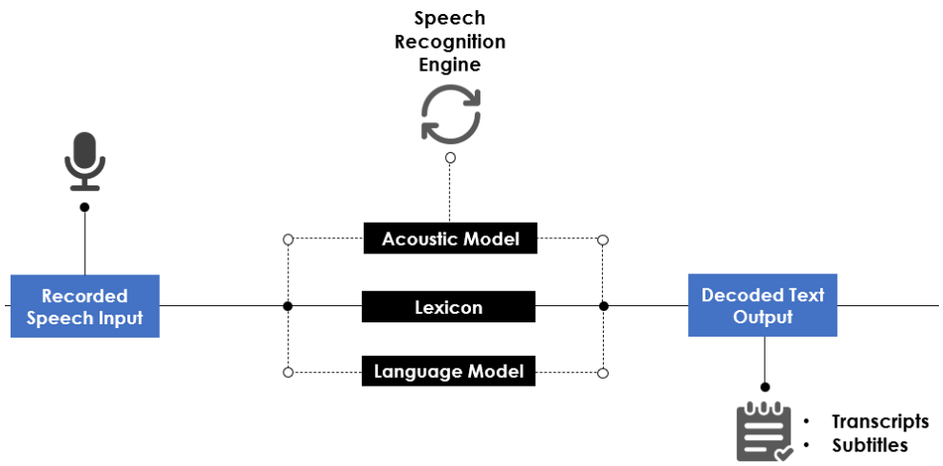


**Figure 5** Dialogue Management System flow

(*Source*: Su [3])

ensure a seamless and cohesive user experience. The integration process began with API Integration, where Convai's AI dialogue models were connected to Unity via APIs, enabling real-time processing of user interactions. This setup allowed the LLMs to dynamically generate contextually relevant responses to user inputs within the VR environment. Speech Recognition technology, specifically Speech-to-Text (STT), was employed to convert user speech into actionable input for the LLMs, ensuring smooth and accurate communication.

To enhance the realism of NPC interactions, Animation Sync was implemented to coordinate NPC animations, such as lip-syncing and facial gestures, with their dialogue. This synchronisation ensured that NPCs' visual and verbal cues aligned naturally, creating a more engaging and lifelike experience. Additionally, Spatial Audio was incorporated to simulate directional sound, allowing NPC voices to be perceived naturally based on their location in the virtual space. These combined efforts ensured users experienced a highly immersive and interactive virtual environment. Table 3 summarises the key components involved in this integration process, highlighting the technologies and their specific roles in bridging AI and VR systems.

## Research and Development Process

### *Requirement Analysis*

Conducted an extensive requirement analysis to identify the necessary tools, technologies, and features essential for developing an effective VR business networking platform. Stakeholder interviews and literature reviews guided the requirement specification phase.

### *Data Collection*

Curated and prepared dialogue datasets relevant to professional networking scenarios. These datasets were used to fine-tune and test the LLMs for accurate and contextually appropriate responses.

**Table 3** Summary of integration of AI into VR

| Component | Description | Purpose |
|---|---|---|
| API Integration | Connected Convai's AI models to Unity via API | Enabled real-time dialogue between users and NPCs |
| Speech Recognition | Used Google Cloud Speech-to-Text for input processing | Converted user speech into text for AI processing |
| Text-to-Speech (TTS) | Google Cloud TTS for NPC speech output | Provided natural-sounding verbal responses |
| Animation Sync | Coordinated lip-syncing and facial expressions | Ensured NPCs appeared realistic and engaging |

### *System Development*

The development process involved building and integrating the NPC behaviour controller, the dialogue management system, and the immersive VR environment. The system was refined iteratively, incorporating user feedback to optimise performance and interaction quality.

### *User Acceptance Testing (UAT) & Evaluation*

A user-centred evaluation was conducted to assess the effectiveness of the VR-AI training tool. Participants engaged with the NPCs in various business networking scenarios, and their feedback was collected using a structured questionnaire focusing on naturalness, relevance, engagement, and context maintenance. Table 4 outlines the participant groups, scenarios, focus area and the initial observations collected.

### Evaluation and Analysis

The project employed both qualitative and quantitative methods to analyse the user feedback. Statistical tests, such as ANOVA, were used to compare the performance of different LLMs in delivering effective dialogues. The findings were used to determine the best-performing model and to propose practical recommendations for further development and implementation in educational and professional settings.

## RESULTS AND DISCUSSIONS

### Development Result

The final prototype represents a significant achievement in integrating AI-driven dialogue systems into an immersive VR environment for professional training. This prototype focuses on meaningful and contextually relevant interactions through three NPCs: Intan, Wong, and Raj. Each NPC is powered by a specific LLM tailored to distinct professional roles, ensuring diverse and dynamic dialogue experiences.

Figure 6 demonstrates the AI-driven NPCs within the VR environment, showcasing their roles in facilitating lifelike conversations. For instance, driven by GPT-4, Intan delivers structured HR guidance, such as resume improvement and interview preparation. Wong, utilising Claude 3.5, provides analytical insights into technology trends, while Raj, powered by LLaMA, engages users in natural, conversational project

**Table 4** Summary of participants group for UAT

| Participant Group | Scenario (NPC) | Focus Area | Initial Observations |
|---|---|---|---|
| Academic Experts | Career Guidance (Intan) | Relevance of career advice and educational value | Positive feedback on realistic and practical advice |
| Industry Professionals | Tech Trends (Wong) | Technical accuracy and clarity | High engagement, with suggestions for deeper technical content |
| Students | Project Planning (Raj) | Ease of use and immersion | Enjoyed the interactive format and clear guidance |


**Figure 6** AI-Driven NPC in VR

management scenarios. These character-specific setups ensure users experience professional-grade, coherent, and immersive dialogue interactions.

The VR environment itself was developed using Unity, incorporating high-quality 3D assets, dynamic lighting, and optimised textures to ensure visual realism and smooth performance. NPCs were created using the Ready Player Me Avatar Creator, with their appearances reflecting their professional roles to further enhance user immersion. The seamless integration of spatial audio and dynamic NPC animations, such as lip-syncing and facial gestures, further elevated the overall user experience. Figure 7 illustrates a sample interaction with one of the LLM-driven NPCs, highlighting the immersive nature of these conversations.

The system was designed for easy navigation and interactivity, with a flow that transitions users from the Loading Scene to the Character Interaction Scene. Key features like spatial audio and a customisable Settings Panel were added to maximise user engagement and realism. Figure 8 shows the complete flow of the system.

**Research Result**

This research focused on evaluating the effectiveness of AI-driven NPCs within a VR environment for professional networking. The evaluation centred on four key attributes: naturalness, relevance, engagement, and context maintenance. A rigorous analysis was conducted using structured user feedback, and the results were validated through statistical methods, including ANOVA, to establish the significance of differences in performance across the three NPCs. Table 5 presents the average scores assigned by participants for each NPC based on the four evaluation attributes.

**Table 5** Average scores for key attributes across NPC characters with different LLM

| Attribute | Raj (LLaMA) | Intan (GPT-4) | Wong (Claude 3.5) |
|---|---|---|---|
| Naturalness | 4.88 | 3.92 | 2.20 |
| Relevance | 3.96 | 3.12 | 3.08 |
| Engagement | 4.84 | 3.15 | 2.20 |
| Context Maintenance | 4.00 | 4.08 | 4.84 |

The results revealed that Raj (LLaMA) outperformed the other NPCs in naturalness and engagement, scoring 4.88 and 4.84, respectively. Participants praised Raj for delivering highly natural, human-like conversations and maintaining captivating dialogue. While Raj performed well in context-maintenance (4.00), there is room for improvement in this area. In comparison, Intan (GPT-4) delivered a balanced performance, excelling in context maintenance (4.08) but showing lower scores in engagement (3.15). Users appreciated her structured responses but suggested incorporating a more conversational tone. Wong (Claude 3.5) achieved the highest score in context-maintenance (4.84) but

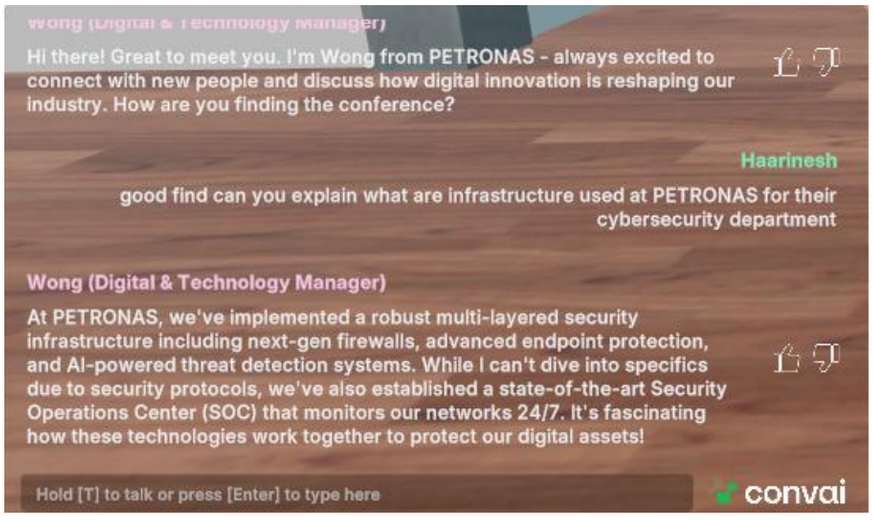


**Figure 7** Sample interaction with LLM-driven NPC

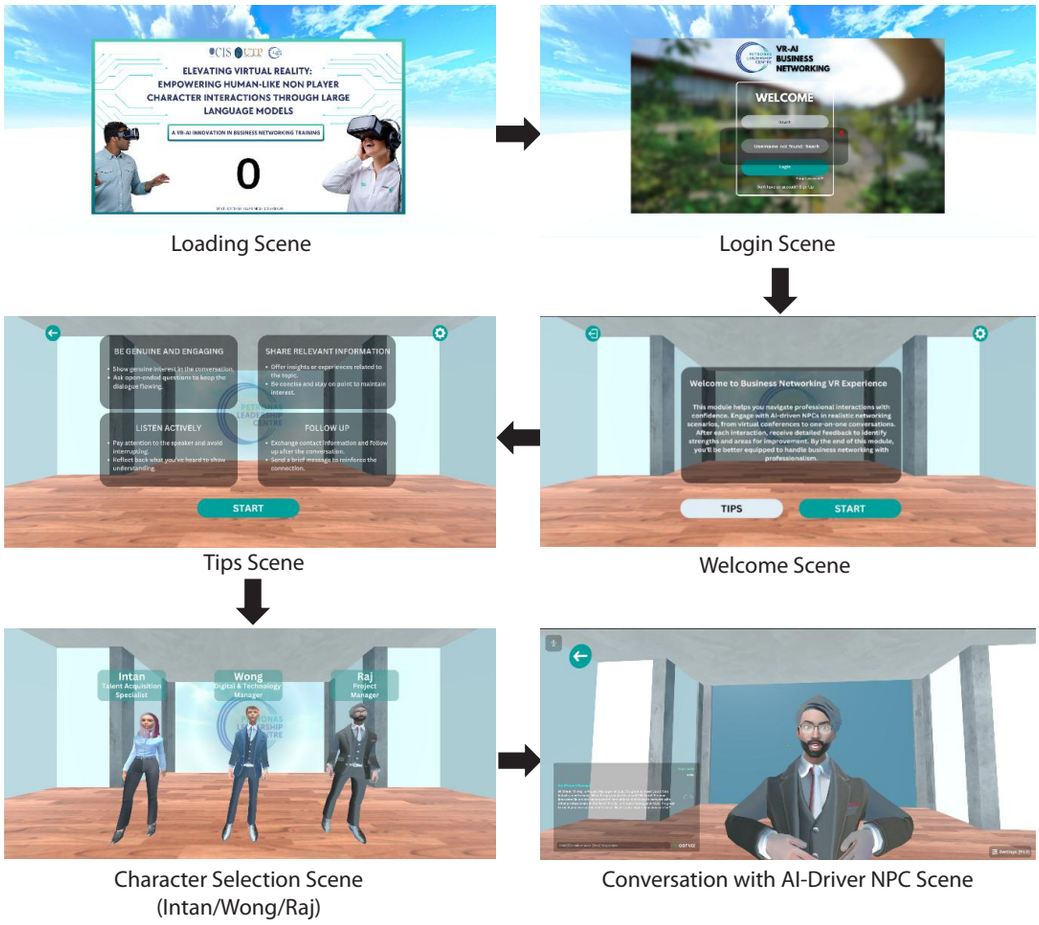


**Figure 8** Complete flow of the system

struggled in naturalness (2.20) and engagement (2.20), with participants describing his interactions as overly formal and less dynamic.

An ANOVA test was conducted to validate these differences, with results summarised in Table 6.

**Table 6** ANOVA Results for Key Attributes

| Evaluation Criterion | F-Value | p-Value | Significance Level |
|---|---|---|---|
| Naturalness | 15.23 | <0.001 | Significant |
| Relevance | 10.67 | 0.002 | Significant |
| Context Maintenance | 20.45 | <0.001 | Significant |
| Human-likeness | 12.89 | 0.001 | Significant |

The ANOVA results confirm that there are statistically significant differences among the three NPC characters across all evaluation criteria. The p-values for each attribute are well below the 0.05 threshold, indicating that the observed differences in performance are not due to chance. Raj's performance was significantly better for naturalness than that of Intan and Wong, aligning with the high user satisfaction reported. The relevance scores also highlight Raj's superiority, although the difference between Intan and Wong was less pronounced. Context maintenance emerged as Wong's strength, with the highest score in this category, reflecting his ability to handle conversation threads effectively, albeit at the cost of a more robotic tone.

In conclusion, the research concludes that as of 2024, LLaMA emerged as the most effective LLM to enhance NPC interaction in VR. This suggests that LLaMA's architecture is particularly suited for generating dynamic and lifelike dialogue in immersive environments.

**Expertise & Industry Validation**

The validation phase was critical to assess the applicability, relevance, and impact of the AI-driven VR business networking training system. Feedback was obtained through structured expert reviews and industry consultations, ensuring that the system met both academic and professional standards.

***Expertise Validation***

The expertise validation was conducted with esteemed professionals from Universiti Teknologi PETRONAS and an international expert in VR. The validation session featured participants such as:

1. Research Officers from the Centre for Excellence in Teaching and Learning (CETaL)
2. Senior Lecturers from the Computer and Information Sciences Department (CISD)
3. Dr. Frederick, Director from Arts et Métiers Institute of Technology, France

The validation session employed structured evaluation criteria, focusing on four key aspects:

1. Naturalness – The ability of NPC dialogues to resemble human-like conversations.
2. Engagement – The extent to which NPC interactions captured and maintained user interest.
3. Context Relevance – The appropriateness of the NPC responses to specific professional scenarios.
4. Applicability to Professional Training – The overall utility of the system in preparing users for real-world networking.

During the session, the panel interacted with the NPCs (Intan, Wong, and Raj) in various networking scenarios and provided comprehensive feedback. Dr. Frederick commended the system's innovative approach to integrating LLMs with VR, emphasising its potential to set a new benchmark in professional training applications. He highlighted that the dynamic and adaptive nature of the NPCs, especially Raj powered by LLaMA, could be transformative in scenarios requiring real-time conversational adaptability.

The CETaL officers noted the importance of refining engagement metrics to ensure that interactions consistently captivate users across all NPCs. They recommended introducing additional measures, such as tracking user dwell time and dialogue progression patterns, to quantify engagement and interaction quality. Meanwhile, the CISD senior lecturers praised the system's adherence to advanced educational technology principles. They suggested incorporating real-world case studies within the NPC dialogue framework to enhance relevance and practicality, particularly for training in project management and technology industries.

***Industry Validation***

Industry validation was conducted through consultations with representatives from the Department of Occupational Safety and Health (DOSH) Perak, focusing on the system's practicality in real-world professional networking and training contexts. The validation session was structured around the following evaluation criteria:

1. Inclusivity – The ability of the system to accommodate users from diverse linguistic and cultural backgrounds.
2. Clarity of Responses – Ensuring NPC dialogues are concise, precise, and easy to comprehend.
3. Real-World Relevance – The alignment of NPC interactions with industry-specific scenarios.

During the session, DOSH officials engaged with the NPCs in simulated business networking scenarios. They provided valuable feedback, particularly emphasising the need for expanded language support to make the system accessible to a broader audience, including users who primarily speak Malay and other regional languages. They suggested that implementing multilingual capabilities would enhance the system's inclusivity and make it more applicable across Malaysia and other multicultural regions.

DOSH officials also highlighted the importance of refining the dialogue structure to ensure responses remain brief yet informative, catering to the fast-paced nature of professional interactions. For example, while Wong (Claude 3.5) excelled in technical accuracy, his responses were perceived as overly formal and lengthy. Streamlining these dialogues while maintaining their analytical depth would improve user satisfaction.

Additionally, the industry representatives praised the system's potential for collaboration with organisations like PETRONAS, where professional networking and safety education are critical. They suggested future research to explore context-specific modules, such as

safety training simulations and leadership development workshops, to broaden the system's application and impact.

## CONCLUSION

The research project successfully demonstrated the transformative potential of integrating Large Language Models (LLMs) into virtual reality environments for professional training, showcasing how advanced AI-driven non-player Characters (NPCs) such as Raj (LLaMA), Intan (GPT-4), and Wong (Claude 3.5) can create immersive, interactive, and contextually relevant networking scenarios. The user testing results highlighted significant improvements in naturalness, engagement, and context maintenance. Raj emerged as the most effective and human-like NPC, a finding supported by rigorous statistical analysis. The visually engaging VR environment, combined with seamless dialogue systems and spatial audio, added to the realism and immersion of the training experience. However, the research also identified areas for improvement, particularly in enhancing conversational spontaneity and reducing interactions' perceived rigidity and formality to better mimic real-world dynamics. These insights lay the foundation for future advancements in AI-VR integration, emphasising the importance of developing more adaptive, multilingual, and culturally inclusive training environments. By addressing these areas, this project contributes to advancing virtual professional training and establishes a robust framework for further research and practical applications, providing valuable lessons for academia, industry, and collaborative stakeholders.

## ACKNOWLEDGEMENT

The authors would like to thank the PETRONAS Leadership Centre for being a collaborator and the department for its continuous support and encouragement. The authors also would like to acknowledge the grant funder, Institute Teknologi Sepuluh Nopember, Indonesia (ITS) 015ME0-397, for funding this research work, paper and publication.

## REFERENCES

[1] D. Brutzman, "The virtual reality modeling language and Java," *Communications of the ACM*, vol. 41, no. 6, pp. 57–64, 1998, doi: https://doi.org/10.1145/276609.276620.

[2] T. Hilfert and M. König, "Low-cost virtual reality environment for engineering and construction," *Visualisation in Engineering*, vol. 4, no. 1, 2016, doi: https://doi.org/10.1186/s40327-015-0031-5.

[3] Y. Su, "Voice2Action: Language Models as Agent for Efficient Real-Time Interaction in Virtual Reality," *arXiv.org*, 2023. [Online]. Available: https://doi.org/10.48550/arxiv.2310.00092.

[4] J. Roberts, A. Banburski-Fahey, and J. Lanier, "Steps towards prompt-based creation of virtual worlds," *arXiv.org*, 2022. [Online]. Available: https://arxiv.org/abs/2211.05875.

[5] A. M. Grow and F. Khosmood, "ChatGPT GameJam: Unleashing the power of Large Language Models for Game Jams," Aug. 2023, doi: https://doi.org/10.1145/3610602.3610605.

[6] Y. Hong et al., "3D-LLM: Injecting the 3D World into Large Language Models," *arXiv.org*, 2023. [Online]. Available: https://arxiv.org/abs/2307.12981.

[7] W. Zhou, X. Peng, and M. Riedl, "Dialogue Shaping: Empowering Agents through NPC Interaction," *arXiv.org*, 2023. [Online]. Available: https://arxiv.org/abs/2307.15833.

[8] K. A. Mills, "Potentials and Challenges of Extended Reality Technologies for Language Learning," *Anglistik*, vol. 33, no. 1, pp. 147–163, 2022, doi: https://doi.org/10.33675/angl/2022/1/13.

[9] M. M. Soliman, E. Ahmed, A. Darwish, and A. E. Hassanien, "Artificial Intelligence powered Metaverse: Analysis, Challenges, and Future Perspectives," *Artificial Intelligence Review*, vol. 57, no. 2, pp. 289–303, 2024, doi: https://doi.org/10.1007/s10462-023-10641-x.

[10] A. Gupta and B. Howard, "Natural Language Processing in NPC Design: A Survey," *Artificial Intelligence Research Review*, vol. 10, no. 6, pp. 511–530, 2023.

[11] R. Tan et al., "Contextual Coherence in AI Dialogue Systems," *IEEE Access*, vol. 9, pp. 45870–45888, 2021.

[12] H. Patel, "LLMs for Adaptive Training in VR Environments," *IEEE Transactions on AI and VR*, vol. 27, no. 9, pp. 1205–1216, 2023.

[13] N. H. Lee et al., "Integrating Spatial Audio with NPC Dialogue in VR," *Journal of VR Research*, vol. 14, no. 5, pp. 556–574, 2022.

[14] B. Zhou and Y. Tan, "Emerging Trends in VR Training Systems," *Virtual Learning Review*, vol. 23, no. 6, pp. 308–319, 2023.

[15] L. Zhang, "LLMs in Education and Professional Training: A Review," *Journal of AI Applications*, vol. 19, no. 7, pp. 441–467, 2023.

[16] "AI Models Leaderboard | Artificial Analysis," *artificialanalysis.ai*. [Online]. Available: https://artificialanalysis.ai/leaderboards/models.