Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 5, **Number** 2, November 2025
https://doi.org/10.47709/brilliance.v5i2.6779

Non-Playable Characters Based On Large Language Models For Role Playing Games (RPG)

Ade Mulyana[1], Yudi Wibisono[2*], Ani Anisyah[3]
[1,2,3]Universitas Pendidikan Indonesia, Indonesia
[1]adem01@upi.edu, [2]yudi@upi.edu, [3]anianisyah@upi.edu

open access

***Corresponding Author**

Article History:
Submitted: 05-08-2025
Accepted: 12-08-2025
Published: 19-08-2025

Keywords:
Artificial Intelligence; Large Language Model; Non-Playable Character; Retrieval-Augmented Generation; Role-Playing Game.

ABSTRACT

Interactive dialogue is a central element in role-playing games (RPG), particularly those that emphasize storytelling and immersion. This study explores the development of a dynamic Non-Playable Character (NPC) system using a Large Language Model (LLM) to simulate responsive conversations in a fictional world. The objective of this research is to design an NPC dialogue system that can maintain contextual consistency with the game's lore while adapting to player input dynamically. The method used is engineering-based development, involving prompt engineering and a Retrieval-Augmented Generation (RAG) approach to embed narrative context into the LLM prompts. The system is implemented in a 2D RPG titled Kage no Meiyaku: Shinobi no Michi, where players interact with multiple NPCs whose responses evolve based on both pre-defined lore and game progression. Evaluation is conducted using a Likert scale across four dialogue quality dimensions: coherence, emotional engagement, narrative relevance, and persona consistency. The results show that the system generates engaging and contextually accurate responses, with average scores ranging from 4.0 to 4.5. Some limitations are identified, such as occasional misspellings and generic responses in ambiguous inputs. However, the approach demonstrates strong potential for AI-assisted storytelling in games. This research contributes to expanding LLM applications in interactive fiction and opens future work toward feature-rich RPG elements such as transactional systems, branching narratives, and real-time battle interactions.

INTRODUCTION

The Role-Playing Game (RPG) genre stands as one of the most popular and lucrative segments of the global games industry, renowned for its deep narratives, imaginative worlds, and complex character interactions (Newzoo, 2021). In 2020, RPGs dominated the mobile gaming market, accounting for $18.5 billion, or 21.3% of total global revenue, largely driven by markets in East Asia (Newzoo, 2021). A critical component of the RPG experience is the Non-Player Character (NPC), who populates the game world and enriches the narrative by providing players with information, quests, and dialogue (Warpefelt & Verhagen, 2016). The realism and quality of these interactions are paramount, as they significantly enhance player immersion and emotional engagement with the story (Jennett et al., 2008).

Traditionally, NPC dialogue has been built on branching dialogue trees, a system where developers pre-script all possible responses. However, this approach has notable limitations: dialogues are often rigid, unable to adapt to unexpected player inputs, and struggle to maintain narrative consistency in dynamic game worlds (Collins et al., 2016; Radež & Bohak, 2024). A well-known example is the generic and repetitive dialogue from guards in The Elder Scrolls V: Skyrim, such as, "I used to be an adventurer like you, then I took an arrow in the knee." This line became infamous for its lack of context relative to the player's actions or story progression. Furthermore, manually scripting these extensive dialogue trees is a time-consuming and resource-intensive process for developers (Gao et al., 2023).

The advancement of Artificial Intelligence (AI), particularly the emergence of Large Language Models (LLMs), presents a transformative solution to these challenges (Shahsavar & Choudhury, 2023). LLMs can process and generate human-like text, offering a promising alternative for creating dynamic, context-aware NPC conversations that enhance narrative consistency and player immersion (Gallotta et al., 2024; Ou et al., 2024). However, the direct integration of LLMs into NPCs is not without obstacles. Key issues include maintaining role consistency, where an LLM might "hallucinate" and generate responses that are out of character or contradict the game's lore (Xu et al., 2024). Another significant challenge is the inherent memory limitation of LLMs, which makes it difficult to recall past interactions, leading to less relevant and immersion-breaking responses (Gallotta et al., 2024).

To address these limitations, this research leverages a combination of Prompt Engineering and Retrieval-Augmented Generation (RAG). Prompt engineering involves strategically structuring inputs to guide the LLM toward generating responses that align with a specific context, style, or persona (Zhang et al., 2025). The RAG framework further enhances this by augmenting the LLM with an external knowledge retrieval mechanism. This allows the model

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 5, **Number** 2, November 2025
https://doi.org/10.47709/brilliance.v5i2.6779

to dynamically pull relevant information such as conversation history, world lore, and character backgrounds from a database, grounding its responses in factual context and overcoming token limitations (Gao et al., 2023). By combining these methods with persona-aware prompting, it is possible to significantly improve the behavioral consistency of LLMs in role-playing scenarios (Ji et al., 2025).

This study aims to bridge that gap by designing and evaluating a prototype open-world RPG where NPCs generate dynamic, context-aware dialogue using LLMs enhanced by RAG and prompt engineering.

The research objectives are fourfold:

1. To develop an LLM-based NPC system capable of generating adaptive, coherent, and lore-consistent dialogue.
2. To implement RAG mechanisms for external knowledge access.
3. To utilize prompt engineering for reducing hallucinations and ensuring persona alignment.
4. To evaluate the system's performance in terms of immersion, consistency, and narrative quality.

By focusing on this integration within a game prototype developed using Python and Pygame, this study contributes to the advancement of intelligent virtual agents in game design. It provides both a technical framework and a proof-of-concept for future implementations of conversational AI in interactive entertainment. Ultimately, this research underscores the potential of combining AI language technologies with narrative game mechanics to redefine how players experience stories through meaningful dialogue.

LITERATURE REVIEW

Large Language Models

LLMs are neural network-based models trained on a vast corpora of text data to predict the next token in a sequence, enabling them to generate coherent and context-aware language. Architectures such as Transformer (Vaswani et al., 2017) form the basis of modern LLMs, including GPT, LLaMA, and PaLM. These models can adapt to a wide range of tasks, including text generation, summarization, and dialogue. However, they often face challenges such as hallucination, memory limits, and maintaining consistent behavior in character-driven applications (Ren, 2024).

Retrieval-Augmented Generation (RAG)

RAG combines LLMs with an external knowledge retriever, allowing the model to pull contextually relevant information from a vector database before generating output. This helps address the memory limitation of LLMs and reduces hallucination by grounding responses in real, up-to-date knowledge (Lewis et al., 2020; Gao et al., 2023). In games, RAG can dynamically incorporate evolving game lore into NPC dialogues.

Prompt Engineering

Prompt engineering refers to the method of crafting effective inputs for LLMs to guide their responses. In the context of game NPCs, persona-aware prompting ensures that the character maintains a consistent voice, attitude, and background knowledge. Techniques like zero-shot, few-shot, and chain-of-thought prompting can also be employed to optimize response quality and maintain narrative coherence (Mao et al., 2024).

Lore and Storytelling in Games

Lore refers to the background story, world history, character relationships, and mythologies that form the narrative foundation of a game world. It plays a central role in RPGs by enriching the player's immersion, providing context for quests and interactions, and offering players a sense of belonging in the game universe (Jenkins, 2004). Well-crafted lore influences gameplay mechanics, dialogue outcomes, and even character development. Environmental storytelling delivering narrative through in-game objects, architecture, and ambient cues further enhances the player experience. Embedding this lore into dialogue systems using LLMs and RAG enables more personalized and believable NPC interactions.

Game Development with Pygame

Pygame is a cross-platform set of Python modules designed for writing 2D games. It enables rapid prototyping of game mechanics, user interface elements, and event handling. Its simplicity and extensibility make it ideal for academic and independent game development. In this study, Pygame was used to develop a functional RPG prototype integrating LLM-based NPCs, allowing researchers to simulate interaction scenarios within a controlled environment.

METHOD

This study employed the ADDIE development model, which stands for Analysis, Design, Development, Implementation, and Evaluation. The ADDIE model provides a systematic and iterative framework for designing and evaluating interactive systems. The purpose of this research is to build an RPG game prototype titled *"Kage no Meiyaku: Shinobi no Michi"*, featuring dynamic Non-Playable Character (NPC) dialogue powered by Large Language Models (LLMs). The system allows players to engage in context-sensitive conversation with NPCs in a narrative-driven

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 5, **Number** 2, November 2025
https://doi.org/10.47709/brilliance.v5i2.6779

game world.

Analysis

The researcher identified the primary problem in traditional RPG design: NPC dialogue is often static, pre-defined, and disconnected from narrative progress. The game aims to solve this by embedding a Retrieval-Augmented Generation (RAG) system that references background lore and previous interactions to generate adaptive, immersive dialogue. The game was designed for players who prefer immersive storytelling and personalized character interaction. Narrative content (lore, characters, world-building) was written by the researcher, while game assets were adapted from open resources.

Design

During this phase, system architecture was planned in three major components: user interface (built using Pygame), prompt engineering (to construct input messages for the LLM), and a RAG module to retrieve relevant lore from the story database. Each NPC was designed with specific personalities, memories, and narrative roles, allowing their dialogue to reflect their emotional state, backstory, and current situation. Prompt templates were structured to include player input, retrieved lore, and NPC persona, ensuring the generated dialogue remains coherent and consistent. The overall system architecture is illustrated in Figure 1, which shows how these components interact to support dynamic NPC conversations within the game environment.

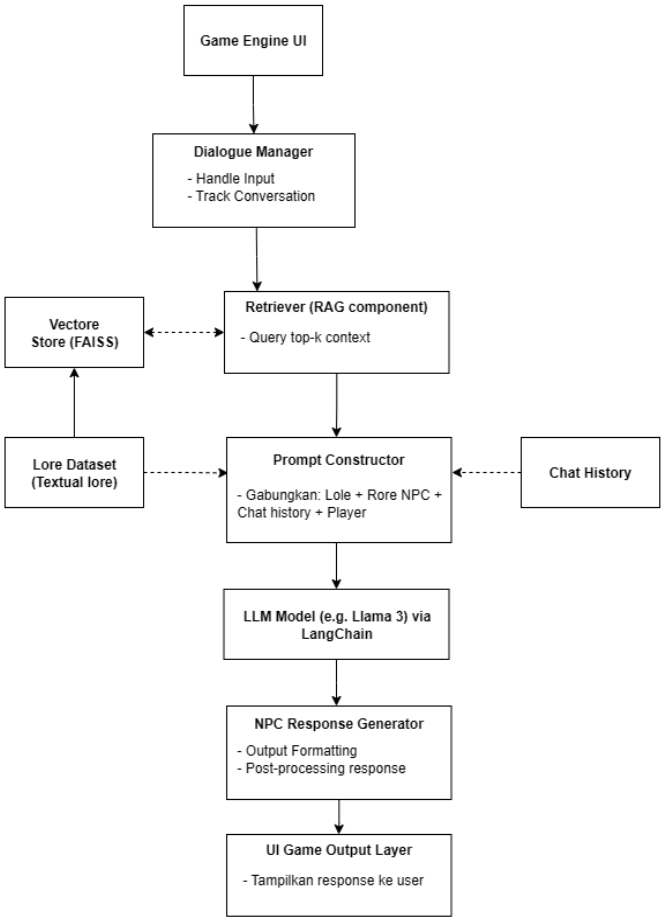


Fig. 1 System Architecture Diagram

Development

In the Development phase, the system was implemented using Python 3.9 and Pygame for game mechanics and interface, LangChain for LLM integration, FAISS for vector-based document retrieval, and LLaMA3 8B as the core language model via API. Game features include main and side missions, a dynamic world state, and changes in NPC personality based on story progression.

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 5, **Number** 2, November 2025
https://doi.org/10.47709/brilliance.v5i2.6779

Implementation

The implementation phase involved integrating the full system and conducting a real-time interaction test. Players could explore the fictional world of Kurokami, engage with NPCs, and trigger branching scenarios. Each NPC response was generated in real-time based on context, prior interactions, and relevant narrative lore, demonstrating how the system adapts to player choices without using predefined dialogue trees. An example of the player-NPC interaction interface is shown in Figure 2, illustrating how dialogue appears in the game environment and how user input is handled dynamically.

Fig. 2 Interaction Dialog Player-NPC UI

Evaluation

Finally, the evaluation stage involved user-based testing through a Likert-scale questionnaire. The evaluation focused on four dimensions: coherence (logical flow of responses), relevance (alignment with lore), character consistency (faithfulness to NPC persona), and emotional engagement (the degree of immersion). Respondents were asked to rate statements after playing the game. The instrument was inspired by the evaluation criteria used in the LLM-as-Judge framework, particularly the work by Shao et al. (2023), although this research relied solely on direct human assessment. The full list of evaluation items is presented in Table 1, which outlines the questionnaire used to assess the perceived quality of NPC dialogue across the four dimensions.

Table 1. NPC Dialogue Quality Evaluation Questionnaire

Code	Evaluation Statement	Strongly Disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Strongly Agree (5)
P1	The NPC's response aligns with the background story (lore) I understood.	☐	☐	☐	☐	☐
P2	The NPC's dialogue sounds natural and not robotic.	☐	☐	☐	☐	☐
P3	The NPC maintains a consistent speaking style in line with their character traits.	☐	☐	☐	☐	☐
P4	Interacting with the NPC enhanced my perception of the game world.	☐	☐	☐	☐	☐

RESULT

The prototype game titled “Kage no Meiyaku: Shinobi no Michi” was successfully developed using Python and integrated with a Large Language Model (LLaMA3 8B) via an API, alongside a Retrieval-Augmented Generation (RAG) mechanism. The system allows players to engage in free-form dialogue with NPCs, where the model generates responses based on contextual lore and character personality.

After implementation, the system was tested by a group of nine users. Each participant interacted with multiple NPCs across several narrative scenarios, including main quests and side quests that influenced character behavior. The interactions were designed to reflect character-specific tone, memory of past interactions, and current narrative progression.

Result

The results focus on the quality of the NPC responses as perceived by users. Evaluation was conducted using a Likert-scale questionnaire comprising four statements related to coherence, relevance to the lore, consistency of character, and emotional engagement. Users rated each statement on a scale of 1 to 5. The results are summarized in

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 5, **Number** 2, November 2025
https://doi.org/10.47709/brilliance.v5i2.6779

Table 2.

Table 2. NPC Dialogue Evaluation Scores

Statement Code	Evaluation Statement	Average Score
P1	The NPC responses align with the story background (lore).	4,4
P2	The dialogue generated by NPCs sounds natural and not robotic.	4,0
P3	Each NPC maintains a consistent speech style based on their persona.	4,4
P4	Interacting with NPCs enhances my emotional connection to the game world.	4,5

The response generation mechanism was able to adapt to different player inputs and narrative changes, with each NPC responding in a way that reflected their designed personality and contextual knowledge. The dynamic nature of the dialogue system allowed players to experience unique conversations that varied based on the progress of the main storyline or side missions.

Additionally, the system demonstrated the ability to recall and refer to recent player interactions by leveraging lore fragments through RAG, resulting in more immersive and context-aware exchanges.

Study case NPC Dialogue

To better illustrate the dynamic response capability of the system, the following test case samples show actual player inputs and the generated NPC responses.

Table 3. Study case NPC Dialogue awal

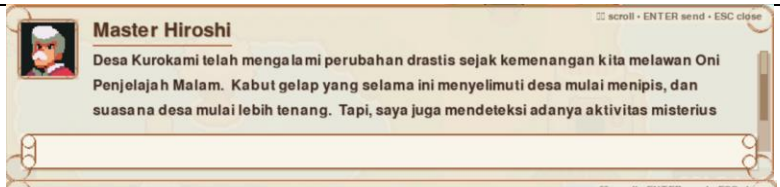
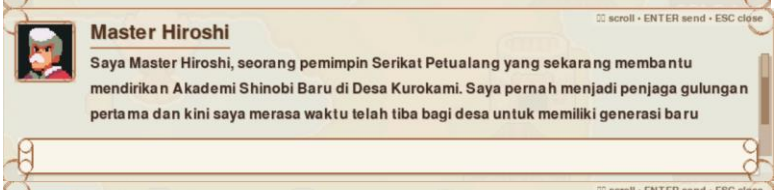
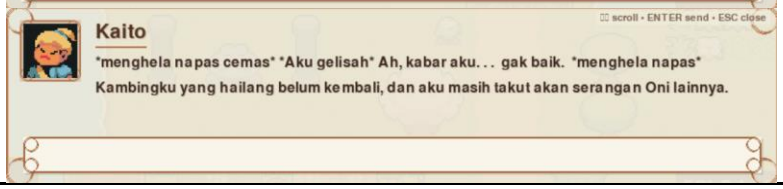
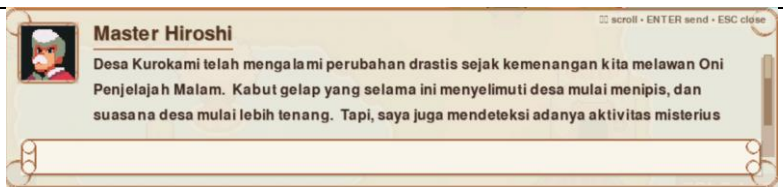
Player Input	Output NPC
"Apa yang terjadi pada desa ini?"	
"Siapa Kamu?"	
"Bagaimana kabarmu, kaito?	

Table 3 presents early-stage interactions between the player and selected NPCs before narrative progression or major player decisions have occurred. At this point, the system has minimal prior context about the player's actions or story development. The NPC responses are based on static persona information and foundational lore only. This table serves to demonstrate how the dialogue system operates before dynamic context (such as past interactions or quests) has influenced the conversation flow. It also serves as a baseline for evaluating improvement over time.

Table 4. Study case NPC Dialogue akhir

Player Input	Output NPC
"Apa yang terjadi pada desa ini?"	

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 5, **Number** 2, November 2025
https://doi.org/10.47709/brilliance.v5i2.6779

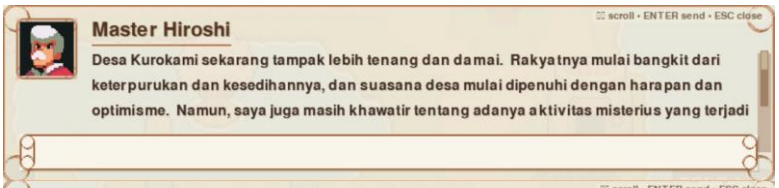
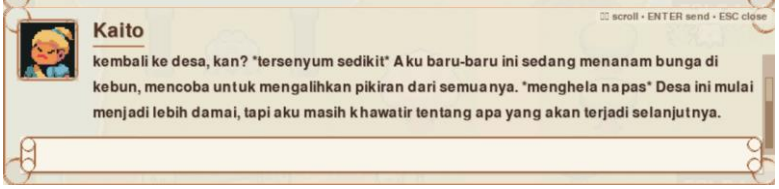
"Bagaimana keadaan warga saat ini"	
"Bagaimana kabarmu sekarang, Kaito?"	

Table 4 shows follow-up interactions with the same NPCs after story events have occurred (e.g., quest completion, emotional turning points). This table highlights how the system's dialogue generation evolves dynamically as the player progresses, proving the effectiveness of RAG-enhanced LLM integration for context-aware storytelling.

Table 5. Study case NPC Dialogue Side Mission

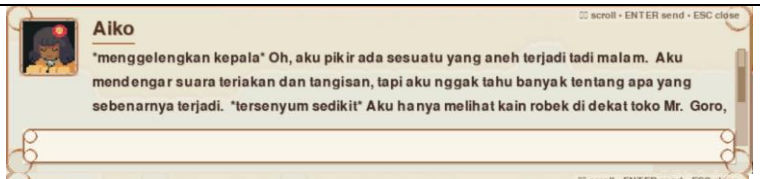
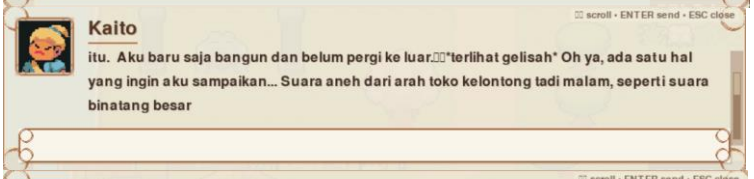
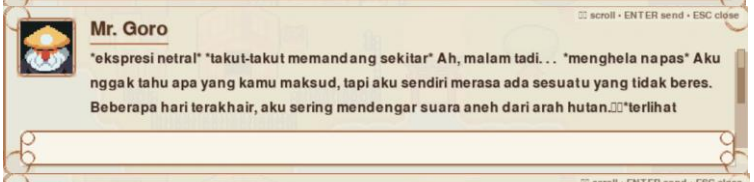
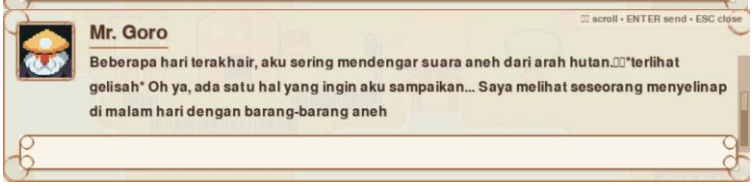
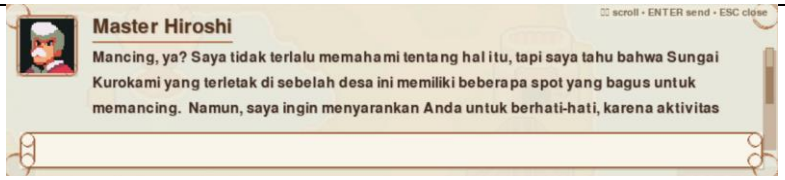
Player Input	Output NPC
"Apa yang aneh tadi malam?"	
"Ada hal mencurigakan atau aneh tadi malam?	
"Ada hal mencurigakan atau aneh tadi malam?"	
	

Table 5 illustrates examples of player interactions during a side mission scenario where the player is investigating a mysterious event that occurred the previous night. In this case, the dialogue system dynamically tailors NPC responses based on their respective roles, personalities, and knowledge scope within the game world. As the player continues the investigation and gathers more clues, it is eventually revealed through in-game narrative triggers and further questioning that Mr. Goro is the murderer. This progression showcases how the system supports branching narratives, emergent storytelling, and player-led discovery, aligning perfectly with the goals of a narrative-rich RPG enhanced by LLMs.

Table 6. Robustness Test: Handling Irrelevant Input

Player Input	Output NPC
"Apakah kamu tahu tempat terbaik untuk memancing hari ini?"	

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 5, **Number** 2, November 2025
https://doi.org/10.47709/brilliance.v5i2.6779

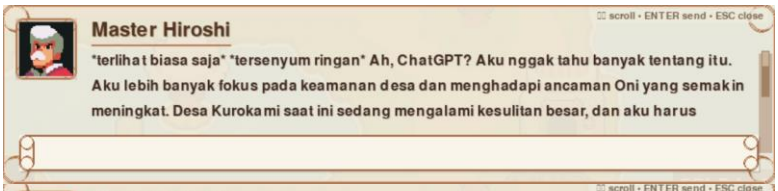
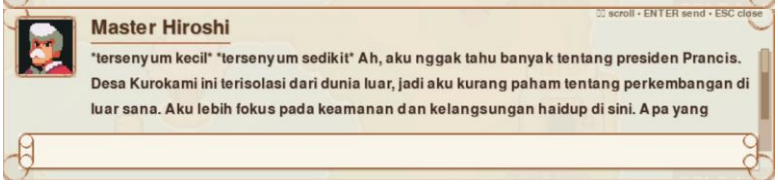
"Apa pendapatmu tentang ChatGPT?"	
"Siapa presiden Prancis?"	

Table 6 demonstrates how the system responds to player inputs that are irrelevant or out-of-universe. Rather than breaking immersion or producing nonsensical answers, the NPCs maintain their persona coherence by responding with either lore-adapted improvisation or evasive phrases. This behavior is aligned with the design goal of ensuring that dialogue remains immersive even when user input is unexpected or inappropriate to the game context.

DISCUSSION

The implementation of dynamic NPC dialogue using a Large Language Model (LLaMA3 8B) combined with Retrieval-Augmented Generation (RAG) has demonstrated the feasibility of integrating context-aware conversation in narrative-driven RPG games. Through the prototype "Kage no Meiyaku: Shinobi no Michi", players were able to interact freely with NPCs in a way that reflected both their personalities and the evolving story context. This marks a significant departure from traditional pre-scripted dialogue trees commonly used in commercial RPGs.

Based on the evaluation results, users perceived the system to be effective across four key dimensions: alignment with lore, naturalness of response, character consistency, and emotional engagement. The average scores ranged from 4.0 to 4.5, suggesting that most interactions were coherent and immersive. This indicates that prompt engineering combined with story retrieval can generate meaningful and believable interactions, even without manually scripted responses.

Furthermore, the study cases including both main quest and side mission interactions show that NPCs responded differently to the same input depending on their persona and narrative position. For instance, the character Mr. Goro, who was revealed to be the murderer in a side mission, responded evasively when questioned about suspicious activity, while other NPCs gave fragmented clues. This illustrates the system's ability to simulate character motivation and selective knowledge within a fictional world.

In addition, a robustness test using irrelevant input showed that the system handled out-of-context queries gracefully. Rather than producing incoherent or out-of-character replies, NPCs either adapted the input to the game world (e.g., "Siapa presiden Prancis?" => "Aku, nggak tahu tentang presiden Prancis…") or subtly rejected it in-character, preserving immersion. This further validates the potential of LLM-based systems to sustain consistent and believable roleplay under diverse interaction conditions.

Despite these promising outcomes, certain limitations were also observed. The system's responsiveness heavily depends on prompt structure and retrieval quality. In some instances, the retrieval component failed to provide highly relevant lore, leading to more generic or vague responses. Also, due to latency in LLM API calls, the real-time experience may require optimization for practical deployment in commercial games.

Overall, the results reinforce the potential of LLM + RAG-based NPC systems in enhancing narrative interaction in games. The study offers a foundation for future work in this domain, including the integration of emotion tracking, long-term memory, and dynamic lore updates over time.

CONCLUSION

This study successfully developed an RPG game prototype titled Kage no Meiyaku: Shinobi no Michi, which features dynamic Non-Playable Character (NPC) interactions powered by a Large Language Model (LLM) integrated with Retrieval-Augmented Generation (RAG). The system enables NPCs to respond to player inputs in a natural and contextually relevant manner by retrieving fragments of game lore and shaping responses according to each NPC's unique persona.

The dialogue system was evaluated based on four key dimensions: coherence with the story (lore), naturalness of generated dialogue, character consistency, and emotional engagement. Results from Likert-scale questionnaires indicated a high level of user satisfaction across all dimensions, confirming the effectiveness of combining LLMs with contextual retrieval mechanisms in enhancing narrative immersion.

Furthermore, the use of FAISS as a vector search database within the RAG pipeline allowed for relevant story elements to be dynamically retrieved based on player queries. This mechanism reduced reliance on static, scripted dialogue and provided more personalized and adaptive interactions throughout the gameplay.

E-ISSN : 2807-9035
Volume 5, **Number** 2, November 2025
https://doi.org/10.47709/brilliance.v5i2.6779

While the findings are promising, several improvements are recommended for future work. These include incorporating rule-based systems for inventory and trading mechanics, expanding tactical or text-based combat scenarios, and implementing more complex branching dialogues and questlines. Additionally, integrating automatic spell-checking modules and conducting broader user testing would help improve system polish and generalizability.

Overall, this research demonstrates that the integration of LLMs with narrative-driven design principles offers significant potential for creating adaptive and immersive dialogue systems in RPG games. The approach paves the way for more personalized and emotionally resonant player experiences in future game development.

REFERENCES

Collins, J., Hirst, W., Tang, W., Luu, C., Smith, P., Watson, A., & Sahandi, R. (2016). *EDTree: Emotional dialogue trees for game based training.*

Gallotta, R., Todd, G., Zammit, M., Earle, S., Liapis, A., Togelius, J., & Yannakakis, G. N. (2024). *Large Language Models and Games: A Survey and Roadmap*. 1–19. https://doi.org/10.1109/TG.2024.3461510

Gao, Q., Isaac, S., Way, B., Catharines, S., Emami, A., & Catharines, S. (2023). *The Turing Quest : Can Transformers Make Good NPCs ?* 93–103.

Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., Wang, M., & Wang, H. (2023). *Retrieval-Augmented Generation for Large Language Models: A Survey*. 1–21. http://arxiv.org/abs/2312.10997

Jenkins, H. (2004). Game Design as Narrative Architecture. In N. Wardrip-Fruin & P. Harrigan (Eds.), *First Person: New Media as Story, Performance, and Game* (pp. 118-130). MIT Press.

Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., & Walton, A. (2008). Measuring and defining the experience of immersion in games. *International Journal of Human-Computer Studies, 66*(9), 641–661. https://doi.org/10.1016/j.ijhcs.2008.04.004

Ji, K., Lian, Y., Li, L., Gao, J., Li, W., & Dai, B. (2025). Enhancing Persona Consistency for LLMs' Role-Playing using Persona-Aware Contrastive Learning. http://arxiv.org/abs/2503.17662

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... & Kiela, D. (2021). Retrieval-augmented generation for knowledge-intensive NLP tasks. Advances in Neural Information Processing Systems, 33, 9459–9474. https://papers.nips.cc/paper/2020/file/6b493230205f780e1bc26945df7481e5-Paper.pdf

Mao, Y., Ge, Y., Fan, Y., Xu, W., Mi, Y., Hu, Z., & Gao, Y. (2024). *A Survey on LoRA of Large Language Models*. 1–144. https://doi.org/10.1007/s11704-024-40663-9

Newzoo. (2021). RPGs are mobile's biggest genre by revenues: How do gamers across East and West engage with the genre?https://newzoo.com/insights/articles/rpgs-are-mobiles-biggest-genre-by-revenues-how-do-gamers-across-east-and-west-engage-with-the-genre

Ou, J., Lu, J., Liu, C., Tang, Y., Zhang, F., Zhang, D., & Gai, K. (2024). *DialogBench: Evaluating LLMs as Human-like Dialogue Systems.* https://arxiv.org/abs/2311.01677

Radež, G., & Bohak, C. (2024). *Integrating Environmental Awareness Into NPCs : Contextual Conversational Interaction in Games.*

Ren, M. (2024). *Advancements and Applications of Large Language Models in Natural Language Processing : A Comprehensive Review. 0,* 55–63. https://doi.org/10.54254/2755-2721/97/20241406

Shahsavar, Y., & Choudhury, A. (2023). User Intentions to Use ChatGPT for Self-Diagnosis and Health-Related Purposes: Cross-sectional Survey Study. *JMIR Hum Factors, 10,* e47564. https://doi.org/10.2196/47564

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems, 30.*

Warpefelt, H., & Verhagen, H. (2016). *A model of non-player character believability Keywords.*

Xu, Z., Jain, S., & Kankanhalli, M. (2024). Hallucination is inevitable: An innate limitation of large language models. *arXiv preprint arXiv:2401.11817*

Zhang, Q., Wen, R., Hendra, L. B., Ding, Z., & LC, R. (2025). Can AI Prompt Humans? Multimodal Agents Prompt Players' Game Actions and Show Consequences to Raise Sustainability Awareness. https://arxiv.org/abs/2409.08486