Computers and Artificial Intelligence；Vol.3, No.3; 2026
E-ISSN: 3006 -3728 P-ISSN：3006-371X
DOI: https://doi.org/10.70267/cai.26v3n3.223231
Published by: Zeus Press

Research on the Application of Generative AI in Immersive Tasks for RPG Games

Weixuan Zhou*

College of Computer Science and Technology, Dalian Maritime University, Dalian 116026, China

**Corresponding author: Weixuan Zhou.*

Abstract

In recent years, generative artificial intelligence has developed rapidly, supported by improvements in computing resources, data availability, and model architecture design. The introduction of the Transformer architecture in 2017 provided a key technical basis for later large-scale language models. The release of ChatGPT in 2022 further increased the visibility and adoption of generative AI systems, contributing to their wider application beyond research settings. At present, research in this area has increasingly involved multimodal models and more efficient deployment strategies. In the field of role-playing games (RPGs), many existing task systems are still based on pre-designed scripts, with relatively fixed task structures, limited narrative variation, and constrained interaction patterns between non-player characters (NPCs). These characteristics limit the flexibility of gameplay experience to some extent and make it difficult to fully support diverse player interactions. This paper reviews studies on generative AI applications in RPGs since 2020, focusing on three aspects: narrative generation, dynamic task design, and NPC dialogue systems. It further summarizes current research developments, discusses issues such as content controllability, experience consistency, and computational cost, and outlines several research directions, including multimodal task modeling and personalization mechanisms.

Keywords

generative AI, RPG games, immersive tasks

1. Introduction

The core experience of role-playing games (RPGs) is built around players taking on fictional roles and engaging with narratives within a virtual environment. Within this structure, quest systems function as a central mechanism that connects gameplay actions with narrative progression and objective setting.

Quest systems in RPGs have developed through several stages. Early titles such as *Ultima* relied on relatively simple command-based tasks, while later works like *Final Fantasy* introduced more structured linear storytelling. In *The Witcher 3*, main narratives were combined with a large number of branching side quests, and more recent titles such as *Black Myth: Wukong* have further expanded quest design through branching structures and multiple endings. Across these examples, quest design has gradually shifted toward more complex and flexible forms.

Despite this evolution, a persistent issue in RPG quest design lies in the tension between pre-designed content and players' demand for freedom. Most traditional quest systems are still based on manually scripted structures. As a result, story progression and gameplay paths are often fixed in advance, which limits variation across playthroughs and reduces flexibility in accommodating different player preferences.

Conventional procedural content generation techniques can partially address content scalability, but they often struggle to maintain narrative coherence. In addition, generated content may still exhibit a degree of repetition, which affects the overall player experience.

Recent progress in large language models (LLMs), with systems such as ChatGPT as representative examples, has introduced new possibilities for RPG quest design. These models can generate narrative text based on prompts and context, and have been used to produce quests with coherent structure and character-consistent dialogue without task-specific training. Their practical advantage lies mainly in reducing the dependency on manually written scripts rather than replacing them entirely.

Värtinen et al. proposed Quest-GPT, one of the earlier attempts to apply pretrained language models such as GPT-2 and GPT-3 to automated RPG quest generation, and evaluated its feasibility through user studies [1].

However, research in this area remains relatively fragmented. There is still limited agreement on how AI-generated quests should be evaluated, particularly in terms of balancing narrative coherence and player autonomy.

Against this background, this study reviews existing work on generative AI in RPG quest systems, focusing on technical approaches, design frameworks, key findings, and current limitations in the literature.

2. Theoretical Foundation

2.1 Core Dimensions of Immersion

Research on immersion is commonly discussed from several psychological perspectives, among which flow theory, narrative transportation, and presence are the most frequently referenced.

Flow theory describes a mental state in which individuals become highly focused on an activity, often accompanied by a reduced awareness of effort and a sense of control and enjoyment [2]. In RPG quest contexts, this experience is typically associated with task difficulty that is neither too easy nor too demanding, along with clear objectives and immediate feedback.

Another frequently used framework is narrative transportation theory. This concept refers to the psychological process in which individuals become absorbed in a narrative, emotionally responding to events and identifying with characters. It involves a combination of attentional focus and mental imagery [3]. In RPGs, quests function as interactive narratives, and the extent to which players become emotionally engaged often influences how strongly this "transportation" effect is experienced.

Presence is usually defined as the subjective feeling of "being in" a mediated environment [4]. It is often discussed in terms of spatial presence and social presence. The former relates to the sense of physically locating oneself within a virtual world, while the latter concerns whether non-player characters are perceived as socially responsive entities. Together, these two aspects contribute to how "real" the game world feels to players.

2.2 Immersion Mechanisms and Break Points in Traditional RPG Quests

In traditional RPG quest systems, immersion is typically formed through a combination of role identification, structured narrative progression, and interaction cycles based on goal setting, player action, and feedback loops [5]. Within this process, players gradually build an understanding of their in-game identity while engaging with predefined narrative structures.

However, many conventional quest designs still rely on relatively fixed templates. A common pattern involves repetitive structures such as traveling to a designated location, obtaining a specific item, and delivering it to a non-player character. Once players become familiar with this pattern, the sense of engagement may gradually weaken, particularly when outcomes become predictable.

At the same time, strict constraints on player behavior can reduce the flexibility of narrative progression, which in turn affects the continuity of narrative engagement. Limited variation in NPC responses also makes interactions feel repetitive over time, reducing the sense of social presence within the game world.

Dialogue systems in many traditional RPGs are still based on predefined branching structures. Although these systems provide a degree of interaction, the repetition of response patterns often makes NPC behavior relatively easy to anticipate. As a result, maintaining a perception of dynamic or responsive interaction becomes more difficult.

Another limitation lies in content uniformity. Since most quests are designed in advance, players with different play styles—such as combat-oriented or narrative-oriented preferences—often experience largely similar progression paths. Procedural generation can increase content volume, but it does not always ensure narrative depth, and in some cases may produce tasks that feel loosely connected to the overall story progression.

The characteristics of generative AI align in several ways with the limitations described above. Real-time task generation, for example, allows quests to adapt to player behavior and game state, introducing variation that is not fixed in advance. Similarly, interactive NPC systems capable of responding dynamically to player input can improve the perceived responsiveness of character interactions, provided that narrative consistency is maintained.

Memory-based mechanisms offer another possible direction. By storing and reusing information about past player actions, such systems can support continuity across quests and enable more persistent narrative consequences [6]. This allows earlier decisions to influence later events without requiring fully hand-authored branching structures.

In practice, many current implementations follow a hybrid design pattern in which key narrative structures are defined by designers, while AI systems are used to generate localized dialogue, side interactions, or adaptive responses. This approach attempts to balance narrative control with flexibility in content generation.

3. Applications of Generative AI in Enhancing Immersive RPG Tasks

3.1 Dynamic Task Generation: Addressing Predictability and Enhancing Autonomy

Dynamic task generation is generally proposed as a way to reduce the rigidity of pre-scripted quest systems and introduce more variation into gameplay. Existing studies typically explore this direction through different levels of control over large language model (LLM) outputs.

One common approach is based on prompt design. By constructing structured prompts that include narrative context, role definitions, and task constraints, LLMs can be guided to generate quest descriptions and objectives that align with specific design requirements [7]. This method is relatively easy to implement and does not require additional training. However, because the model output relies heavily on prompt quality, the generated content may occasionally deviate from the intended game logic or world consistency, especially in longer interactive sequences.

Another line of work focuses on adapting model behavior through fine-tuning or retrieval-augmented generation (RAG). Fine-tuning allows models to better align with domain-specific narrative styles, but it depends on relatively large amounts of curated training data. RAG methods, on the other hand, introduce external information sources such as prior quest logs or world-state data during generation, which helps improve factual consistency and contextual relevance [8]. These approaches are often used in practice together, with prompt-based methods providing flexibility and retrieval or fine-tuning mechanisms improving stability.

3.2 Intelligent NPC Dialogue Systems: Enhancing Social Presence

In traditional RPGs, NPC dialogue is usually organized through branching trees, where players select from a limited set of predefined options. Once these options are familiar, interaction tends to become predictable, and it is difficult to accommodate inputs that fall outside the designed pathways [9].

By contrast, dialogue systems based on large language models allow players to input free-form text. Responses are generated dynamically according to character settings and prior dialogue context, which expands interaction beyond fixed-choice structures. This shift changes the interaction model from selection-based communication to more open-ended conversational exchange.

One issue in traditional systems is that NPCs often lack continuity in memory. To address this, recent studies distinguish between short-term and long-term memory mechanisms. Short-term memory helps maintain coherence within a single interaction session, while long-term memory stores key player-related information across multiple encounters [10]. In later conversations, relevant past events can be retrieved and incorporated into responses, making NPC behavior appear more consistent over time. This can create the impression that characters retain awareness of previous interactions.

Studies have reported that NPC systems incorporating memory mechanisms are associated with higher perceived social presence, particularly in terms of continuity and relational engagement.

3.3 Hybrid Task and Narrative Architectures: Balancing Freedom and Control for Immersion

Pure generative approaches offer a high degree of flexibility in content creation, but they may also introduce inconsistencies in narrative logic or produce outputs that deviate from established world rules. On the other hand, fully pre-scripted systems provide stronger narrative control but are limited in scalability and variation.

To balance these constraints, many current designs adopt hybrid structures. A common approach is the "predefined narrative skeleton + AI-based content filling" pattern, where major plot points are authored in advance while dialogue and side interactions are generated dynamically by AI systems [11]. Another frequently used configuration combines AI generation with human review, where candidate content is produced by the model and then refined or filtered by designers.

In practice, this type of division allows core narrative structure to remain stable while still introducing variation at the level of moment-to-moment interaction.

4. Research Methods and Evaluation

4.1 Common Research Designs

Research in this area is mainly conducted through two relatively common approaches.

One approach is prototype development combined with user experience evaluation. In this type of study, researchers build experimental systems based on GPT-series models or open-source large language models such as LLaMA. These systems are often implemented in simplified environments, including text-based interactive platforms, lightweight 2D game scenes, or modifications of existing game frameworks. Participants are then invited to complete predefined tasks while interacting with the system, and their experiences are recorded through questionnaires, interviews, or behavioral logs. This type of design is often used to compare AI-generated content with traditional scripted systems or to examine differences between model configurations. While it allows relatively direct observation of user experience, the controlled experimental setting may not fully reflect long-term gameplay conditions in real RPG environments [12].

Another commonly used approach is case study analysis. This method focuses on existing systems or applications, such as AI Dungeon or AI-driven modules integrated into commercial games. Instead of controlled experiments, researchers rely on gameplay observation, developer interviews, and user feedback analysis to understand design patterns and practical issues [13]. This approach provides access to more complex real-world usage contexts, although the findings are often closely tied to specific cases, which limits the extent to which conclusions can be generalized.

4.2 Evaluation Tools

In studies examining immersion and related player experiences, a combination of subjective scales and task-oriented metrics is commonly used.

One widely adopted approach is the use of immersion-related questionnaires. These instruments are often based on established tools such as the Game Engagement Questionnaire (GEQ) and the Immersive Experience Questionnaire (IEQ), sometimes with modifications to fit specific experimental contexts [14]. The items typically cover aspects such as perceived realism, attentional focus, time perception, and emotional engagement. In comparative studies, AI-generated quests are often reported to perform better in terms of novelty and perceived autonomy, while differences in narrative coherence tend to be less consistent across experiments.

Another category of evaluation focuses on the characteristics of the generated tasks themselves. Rather than measuring player perception directly, these metrics examine properties such as logical consistency across quest steps, degree of novelty compared to template-based content, feasibility of task completion, and alignment with the game world setting. In practice, these aspects are often assessed through human annotation, although some studies also explore automated proxies such as text similarity or perplexity-based measures.

In addition to general immersion scales, the Flow Experience Questionnaire is frequently used to capture task engagement from a flow-theory perspective. This includes dimensions such as challenge–skill balance, clarity of goals, immediacy of feedback, and action awareness. When applied to AI-generated content, findings are relatively mixed: some studies report improved alignment in challenge–skill balance, while others note that unclear task instructions may reduce perceived goal clarity, occasionally leading to confusion during gameplay [15].

5. Main Findings and Controversies/Challenges

5.1 Main Findings

Across existing empirical work, a relatively consistent pattern has emerged regarding the role of generative AI in RPG task design. The most stable finding is that AI-generated tasks tend to increase perceived novelty and unpredictability. Compared with traditional pre-scripted quest structures, AI-based generation introduces greater variation in task framing and progression, which reduces players' ability to anticipate outcomes and sustains attention over longer gameplay periods. In many experimental studies, this effect appears robust, with novelty-related scores for AI-generated content significantly exceeding those of control conditions ($p < 0.05$), often with medium to large effect sizes.

Several experimental comparisons further support this tendency. In controlled user studies involving different task-generation methods—including human-designed content, stochastic baselines, and language model outputs—neural language models such as GPT-2 have been shown to produce tasks that approach or, in some cases, match human-authored content in perceived originality [16]. Similar patterns are observed in comparisons with commercial RPG quest datasets, where AI-generated content is often evaluated as comparable to human-written tasks in dimensions such as interest and novelty, and occasionally exceeds baseline handcrafted content on specific metrics [17]. Larger-scale user evaluations also suggest that a substantial portion of AI-generated quests is perceived as sufficiently coherent and engaging for practical game use, although acceptance thresholds remain uneven across player groups [1].

5.2 Controversies and Challenges

Despite growing agreement on the benefits of generative diversity, several unresolved issues continue to shape the field.

One central tension lies in the balance between pre-scripted narrative structure and real-time generation. Fully generative approaches are often associated with higher flexibility and emergent storytelling, yet they risk weakening narrative coherence and emotional pacing. By contrast, strongly pre-authored systems preserve story structure but limit variability in player experience. Although hybrid architectures combining fixed narrative skeletons with AI-generated micro-content are widely adopted, the boundary between structured and generated components remains loosely defined in current practice, and design principles for integrating the two layers are still underdeveloped.

Another persistent constraint is related to contextual limitation in large language models. Even with recent improvements in context window size, maintaining long-term coherence across extended gameplay sessions

remains difficult. External memory modules have been introduced to partially address this issue, but efficient retrieval and prioritization of relevant past events without increasing system latency remains an open technical problem.

A further practical concern involves system robustness under player interaction. Once players become aware of the generative nature of task systems, some attempt to deliberately manipulate outputs, producing incoherent, exploitative, or inappropriate content. In open environments, this raises additional risks, particularly when generated content deviates from intended narrative boundaries or violates content safety constraints. Such issues highlight the need for stronger alignment mechanisms and more controlled generation frameworks in interactive game settings [18].

6. Future Directions

Drawing on the current state of research and its remaining limitations, future work in generative AI for RPG task systems is likely to evolve along two closely related trajectories, both centered on extending temporal consistency and interaction complexity.

A first direction concerns the development of long-context modeling for interactive narrative systems. While recent large language models have significantly expanded their context windows, sustaining coherent storytelling across extended gameplay sessions remains difficult in practice. Much of the current workaround still depends on external memory modules and retrieval mechanisms, which introduce additional architectural complexity and may not fully capture long-range dependencies in player behavior.

Long-context models offer an alternative path by allowing more continuous incorporation of interaction history directly into the generation process. In principle, this would enable narrative systems to reference earlier player actions without explicit retrieval, improving the continuity of character memory and quest logic. Over time, such mechanisms may support more stable narrative trajectories, where non-player characters respond consistently to events occurring across extended in-game timelines. At the same time, richer contextual input could support more differentiated player modeling, where behavioral tendencies and interaction styles are gradually reflected in task design, leading to more adaptive and individualized gameplay experiences [19].

A second direction shifts the focus from single-agent generation to multi-agent narrative systems. Most existing implementations still rely on a single model responsible for generating tasks or controlling non-player characters, which limits the structural complexity of emergent narratives. By contrast, RPG worlds naturally involve distributed agency, where multiple characters and systems interact simultaneously.

Multi-agent frameworks introduce the possibility of modeling these interactions more explicitly. Instead of isolated character generation, different non-player characters can be embedded within a shared environment where their actions and decisions influence one another. This makes it possible to generate narrative situations that resemble coordinated or conflicting intentions among agents, resulting in more dynamic story development beyond designer-authored scripts.

Within this broader shift, player participation can also be reconsidered. Rather than being positioned solely as the recipient of generated content, the player may function as an additional agent within the system, whose actions feed back into the state of other entities in the environment. This creates a more continuous interaction loop in which narrative evolution is shaped jointly by system-driven and player-driven decisions, gradually blurring the boundary between authored content and emergent storytelling [20].

7. Conclusion

This paper reviews existing research on the application of generative AI in immersive RPG task systems, with a particular focus on how it reshapes task design, narrative delivery, and player experience. Rather than producing a uniform shift in game experience, current evidence suggests a more uneven transformation.

On the one hand, generative AI is relatively effective in expanding the variability of quest content. Across different experimental settings, improvements are most visible in task novelty and the perceived responsiveness of non-player character interactions. Players are generally more likely to describe AI-generated

content as less predictable and more conversational in tone. These aspects align with short-term engagement indicators, particularly those related to attention capture and interaction freshness.

At the same time, the limitations are equally consistent across studies. Narrative coherence does not improve at the same pace as surface-level diversity, and in some cases becomes more fragile as generation freedom increases. Issues such as unclear task direction or loosely connected quest steps are repeatedly reported, especially in systems without strong structural constraints. This also contributes to a recurring experience in which players understand what to do locally, but lose clarity about why they are doing it in the broader narrative context.

In response to these tensions, most existing systems converge on a hybrid design logic rather than fully generative or fully scripted approaches. The combination of fixed narrative scaffolding with AI-generated micro-level content is not simply a technical compromise, but a practical response to the need to preserve narrative structure while introducing variability. Pure generation methods remain constrained by consistency issues, while fully pre-authored systems struggle to scale content diversity in meaningful ways.

More broadly, the field is still characterized by early-stage experimentation. Many findings are based on limited sample sizes or short-term user studies, and evaluation standards remain far from unified. Technical constraints such as memory length and context management, together with behavioral risks introduced by open interaction, continue to shape system design in practice.

If there is a central takeaway from existing work, it is that generative AI does not automatically make RPG experiences more immersive in a uniform way. Its contribution is more selective, strengthening certain dimensions of experience while exposing weaknesses in others. Whether these systems ultimately enhance or disrupt immersion depends less on generation capability itself, and more on how narrative structure, constraint mechanisms, and interaction design are combined in practice.

References

[1] Värtinen, S., Hämäläinen, P., & Guckelsberger, C. (2024). Generating role-playing game quests with GPT language models. *IEEE Transactions on Games*, 16(1), 127-139. https://doi.org/10.1109/TG.2022.3228480.

[2] Wang, Y., & Qian, S. F. (2026). Research on the design of children's popular science games based on flow theory. *Industrial Design*, (04), 61-65. https://doi.org/10.26920/j.cnki.231516cn.2026.04.039.

[3] Zhang, M. (2023). From game to tourism: A study on the mechanism of influence on tourism intention from the perspective of narrative transportation theory (Doctoral dissertation, Sichuan Agricultural University). https://doi.org/10.27345/d.cnki.gsnyu.2023.001236.

[4] Wu, L., Wang, Y. H., & Wu, B. (2026). The formation mechanism of tourists' presence in digital immersive scenarios: A case study of Yangzhou Grand Canal Museum of China. *Journal of Central South University of Forestry & Technology (Social Sciences)*, 20(02), 72-83. https://doi.org/10.14067/j.cnki.1673-9272.2026.02.006.

[5] Wu, R. N. (2025). Wandering in the "gaps": An exploration of immersion in RPG games. *Contemporary Animation*, (04), 109-114.

[6] Liu, C. X., Wang, M., Yan, S. M., et al. (2026). Unpaired image dehazing algorithm based on memory mechanism and frequency domain learning. *Journal of Computer-Aided Design & Computer Graphics*, 1-12. https://link.cnki.net/urlid/11.2925.TP.20260227.1406.002.

[7] Li, J. S., Li, R., & Deng, L. (2026). An automatic extension method of fault tree based on large language model prompt engineering. *Computer Applications*, 1-7. https://link.cnki.net/urlid/51.1307.TP.20260324.1359.005.

[8] Liu, H. R., Miao, R. S., & Ye, C. C. (2025). Emergency management question-answering method driven by RAG and fine-tuning of large language models. *Smart Computer and Applications*, 15(12), 164-170. https://doi.org/10.20169/j.issn.2095-2163.25092801.

[9] Schlünder, B., & Klabunde, R. (2013). Greetings generation in video role playing games. In *Proceedings of the 14th European Workshop on Natural Language Generation* (pp. 167–171). Association for Computational Linguistics.

[10] Pal, P., et al. (2024). Scalable GaN-HEMT model for X-band RF applications. In *2024 8th IEEE Electron Devices Technology & Manufacturing Conference* (EDTM) (pp. 1-3). IEEE. https://doi.org/10.1109/EDTM58488.2024.10512345.

[11] Colinsky, C. (2026). Lunar Command: The GRAG architecture - A white paper on hybrid generative narrative systems. Lunar Command.

[12] Ashby, T., Webb, B. K., Knapp, G., Searle, J., & Fulda, N. (2023). Personalized quest and dialogue generation in role-playing games: A knowledge graph- and language model-based approach. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (CHI '23) (Article 290, pp. 1–20). Association for Computing Machinery. https://doi.org/10.1145/3544548.3581441.

[13] Al-Nassar, S., Schaap, A., Van Der Zwart, M., Preuss, M., & Gómez-Maureira, M. A. (2023). QuestVille: Procedural quest generation using NLP models. In *Proceedings of the 18th International Conference on the Foundations of Digital Games* (FDG '23) (Article 50, pp. 1–4). Association for Computing Machinery. https://doi.org/10.1145/3582437.3587188.

[14] Chen, D., Zhu, C. K., Sun, Y. Y., et al. (2023). Translation and reliability and validity test of the Clinical Simulation Learning Immersion Scale. *Journal of Nursing Science*, 38(04), 12-15.

[15] He, Y. C., Fan, Y. T., & Gao, X. D. (2026). Exploration of the path of college students' digital sports flow experience to alleviate social anxiety: Based on the perspective of technology immersion and anonymity regulation scenarios. *Digital Sports Research*, (01), 188-199.

[16] Ammanabrolu, P., Broniec, W., Mueller, A., Paul, J., & Riedl, M. (2019). Toward automated quest generation in text-adventure games. In *Proceedings of the 4th Workshop on Computational Creativity in Language Generation* (pp. 1–12). Association for Computational Linguistics.

[17] van Stegeren, J., & Myśliwiec, J. (2021). Fine-tuning GPT-2 on annotated RPG quests for NPC dialogue generation. In *Proceedings of the 16th International Conference on the Foundations of Digital Games* (FDG '21) (Article 2, pp. 1–8). Association for Computing Machinery. https://doi.org/10.1145/3472538.3472595.

[18] Li, L. (2025). Research on safety detection of network audiovisual content based on AI technology. *Audio Engineering*, 49(06), 94-96+115. https://doi.org/10.16311/j.audioe.2025.06.030.

[19] Sun, H. Y., Liu, Y. Jun., Yu, L. N., et al. (2025). Unknown intent recognition method based on large model in-context learning. *Cybersecurity and Data Governance*, 44(11), 38-44. https://doi.org/10.19358/j.issn.2097-1788.2025.11.007.

[20] Xia, B., Hu, M. L., Yuan, T., et al. (2026). A review of multi-agent collaborative workflows. *Software Guide*, 25(03), 1-8.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.