Proceedings of CONF-CDS 2025 Symposium: Data Visualization Methods for Evaluation
DOI: 10.54254/2755-2721/2025.PO25184

Research on the Application of Large Language Models in Interactive Game Storyline Generation

Zhaojun Hu

School of Earth Science and Engineering, Hohai University, Nanjing,China
2209030220@hhu.edu.cn

Abstract. This study focuses on the application of Large Language Models (LLMs) in interactive game storyline generation, and systematically analyzes their performance and key challenges in the dimensions of personalized response and reasoning logic. Through literature review and case studies, the study reveals that LLMs have obvious deficiencies in long-term character memory, story logic coherence and causal reasoning, which become the main obstacles to narrative consistency and player immersion experience. Based on this, the thesis proposes several optimization paths, including memory enhancement, logic modeling, controlled generation and human-computer co-creation. The study concludes that if the narrative control and personality adaptation ability of LLM can be improved, it is expected to achieve a more immersive, intelligent and personalized AI narrative experience, providing theoretical and methodological support for generative content in game design.

Keywords: Large Language Models (LLMs), Narrative Generation, Personalized Response, Reasoning Ability, Causal Reasoning

1. Introduction

Artificial Intelligence (AI) is profoundly affecting the gaming industry. In traditional games, most of the game's plot direction and dialog content is written to death, with few options for the player to choose from, and a feeling that the interaction lacks sufficient freedom. But with the addition of the Large Language Model (LLM), personalized and dynamic storylines are being noticed more and more. The author hopes that games with access to LLM will be able to generate natural language conversations in real time based on player input, and even simulate personalized personalities, emotions, and memories. The game may be able to truly evolve dynamically, allowing NPCs to remember the player's behavior thus forming a personalized interactive experience, and even adjust the plot and game difficulty in real time. This facilitates a truly adaptive and player-specific narrative. At the same time, in the field of plot development AI can also assist in generating tasks, plot branches, and even automatically designing new levels or maps. This reduces the developer's workload while also enhancing the richness of the game content.

PR Newswire [1] reports that the AI gaming market is expected to grow $27.47 billion from 2025-2029, at a CAGR of 42.3%, while according to the AI in Gaming report by Market.us [2]: the game development and design segment is the leading segment by 2023, accounting for an AI

© 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2025.PO25184

gaming market share of 41.3% or more. This aptly reflects the potential of LLM in dynamic narratives.

There are two core elements in the process of dynamically generating plots, and how well they perform largely determines the player's experience with LLM-generated plots. They are "personalization" and "reasoning". Therefore the study will talk about how LLM improves personalization and reasoning in the game, two very important issues for players. The study also will take a look at the current status and challenges of LLM's interactive gameplay and intelligent NPC assistants, and think about what the future holds. Hopefully, these discussions will shed some light on the application of LLM in the gaming industry.

This study was conducted using a combination of methods. First, the literature analysis method is used to sort out the current status and technical challenges of LLM application in game plot generation, reviewing academic papers and industry reports to establish the theoretical foundation. Second, a case study is conduct ed to evaluate the dynamic plot generation capability of LLM in text adventure games by selecting AI Dungeon as a representative case. Finally, an LLM interaction experiment is conducted to test its logical consistency and adaptability by generating plots and quest suggestions to verify the practical effectiveness of LLM in reasoning and personalized response. The combined use of these approaches aims to comprehensively explore the potential and limitations of LLM in games from theoretical, practical and experimental perspectives.

2. State of the art of LLM in plot generation

2.1. Personalization and reasoning skills in plot generation

"Personalization" and "reasoning" are inextricably linked. For example, when the LLM controls the game and talks to you, reasoning ability enables the AI to produce human-like logical discourse and personalization ability makes the AI generate user-preferred content, the combination of these two makes the AI create plots and dialogues that make sense and meet the unique needs of the player.

In modern game development, personalized response and reasoning capabilities are key factors in enhancing player immersion when using LLM for plot generation. LLM personalizes the experience by analyzing the player's choices and behaviors to generate plot branches that match their preferences. For example, in AI Dungeon, LLM generates open-ended story text based on player inputs, and different behaviors trigger different world states and character reactions, making the story path unique to each player's experience. Another example is Hidden Door, which uses LLM to create a sustainable character relationship system that dynamically changes character attitudes and plot directions based on player choices, emphasizing personality growth and social narrative.

In addition, LLM's reasoning ability ensures logical consistency and causality in the plot. Through the built-in Chain of Thought (CoT) mechanism, LLM can break down complex problems into a series of intermediate steps, simulating the logical thinking process of human beings, thus generating reasonable and coherent plot development. For example, the Inworld AI platform tries to make characters memorize the sequence of events and react reasonably in dialogue-driven games, so as to keep the plot logic consistent, e.g., conflicts due to misunderstandings, and reconstructed plot segments due to failure of tasks. Through these applications, LLM is no longer just a dialog generation tool, but has gradually evolved into a narrative engine with causal understanding and character arc building capabilities.

DOI: 10.54254/2755-2721/2025.PO25184

2.2. Existing LLM plot generation cases

AI Dungeon is a groundbreaking text adventure game developed by Latitude, built on OpenAI's GPT family of large language models, with a core mechanic that utilizes natural language understanding and generation to provide players with a dynamic and open-ended storyline experience.

Unlike traditional games that rely on preset scripts, AI Dungeon's narrative is driven entirely by player input, and LLM generates story development in real time based on that input, resulting in a highly personalized gameplay experience. Players are free to choose their characters, backgrounds, and actions, and each decision changes the course of the story, allowing for virtually unlimited plot branches. This personalization mechanism allows the same opening to evolve into very different story lines in the hands of different players.

In terms of reasoning ability, AI Dungeon partially fulfills the goal of responding logically based on player input. For example, the model can infer a character's current motivations based on previous events, choose compatible behaviors, or explain causal relationships between tasks. However, in practice, due to the lack of long-term memory mechanisms, the model often "forgets" previous episodes or generates contradictory events (e.g., dead characters reappearing), indicating that its ability to construct causal chains and maintain context still needs to be improved.

Nevertheless, the emergence of AI Dungeon marks an important attempt of LLM in game plot generation. It demonstrates great potential in personalized narrative and on-the-fly plot construction based on language models, and exposes the key technical challenges of current LLMs in reasoning coherence and memory consistency. For example, games sometimes generate logically inconsistent plots, resulting in impaired story coherence. The researchers point out that current LLMs still need to be further optimized to improve narrative logic and consistency when dealing with complex causal relationships and long-term contexts.

AI Dungeon by Latitude Games redefines traditional gaming with an infinitely generative text adventure experience. Players are immersed in a world full of surprises, where every decision affects plot development and builds a compelling story based on their choices. The game's evolution showcases its ability to adapt and respond to player input, ensuring that each mission is filled with fresh excitement. In AI Dungeon, players will enter a realm of infinite imagination where cutting-edge technology brings the story to life.

3. Limitations and technical challenges of LLM in episode generation

3.1. Sustainability in personalized narratives

In the game, the character's personality should not be static, but should develop with the player's choices. For example, a cold-hearted assassin may gradually become kind and even change his fate after many warm interactions. However, LLMs often fail to "remember" the choices made by the players and the trajectory of their character development. Like a stateless narrative generator, it "reacquaints" itself with the character each time it generates a story for them. This results in a character's behavior and tone suddenly reversing from one paragraph to the next, without regard to the previous path of development. For games that emphasize emotional buildup and interactive immersion, this kind of "amnesic writing" can seriously undermine the personalized narrative experience.

The sustainability of personalized narratives is the key to enhancing player immersion and character interaction. However, current LLMs have significant limitations in this area, which are

DOI: 10.54254/2755-2721/2025.PO25184

mainly manifested in the inability to effectively track players' historical behaviors and the lack of long-term memory mechanisms, leading to the problem of "forgetting the past" or "personality disruption" in characters or storylines.

LLMs typically rely on a limited context window to process information, which means that they can only take into account a limited range of antecedent content when generating text, and are unable to maintain dialog or narrative consistency across longer time spans. For example, in a game, a player may have developed a deep relationship with a character, but due to the LLM's lack of memory for previous interactions, the character may behave as if they were meeting for the first time in subsequent episodes, ignoring the history of previous interactions and undermining narrative coherence and character believability. Experiments show that LLM memory coherence drops to 50% after 5 rounds of interaction, affecting the personalized experience of the game.

3.2. Problems of logical consistency in storytelling

When people read a novel or play a game, they want the story to be coherent and the plot to make sense. However, the plots generated by language models often have conflicting settings, for example, the protagonist is obviously seriously injured, but in the next paragraph, he or she participates in the battle with vigor; or the plot requires the player to "recall it" when a password has just been forgotten. These seemingly small loopholes reflect the model's lack of structural "world understanding". It's more like an on-the-fly scriptwriter who writes every sentence beautifully, but can't control the overall structure and consistency of the entire script, a problem that's especially evident in mid-length narratives.

Maintaining the consistency of story logic is a basic and critical capability in game plot generation, yet current LLMs are still weak in this aspect. Due to the lack of global structural modeling and stable narrative strategy, the model often generates conflicting content when dealing with complex stories, which undermines the coherence of the plot. Specific manifestations include inconsistent character settings and behaviors, broken plot clues, and repetition of important events, which make the model-generated narratives loose or even chaotic in the overall logic. Experiments show that 30% of the plots generated by LLM have logical contradictions, such as character resurrection, which need to be optimized [3].

This kind of logical inconsistency is usually not a single mistake, but a systematic limitation of the language modeling mechanism (LLM), which relies mainly on short-term contextual predictions, has limited control over the overall story structure, and lacks the causal chain management of the process of "what happened and what should happen next". Although this kind of local generation and local memorization is natural at the level of utterance, it is difficult to maintain the unity of setting and the order of story advancement in narrative construction, which has become one of the core obstacles affecting the sense of immersion in the current storyline generation.

3.3. Causal reasoning

For example, the player accepts the quest first, then acquires the props, then enters the replica, and this order cannot be messed up. Yet LLM often ignores this causal order in story generation. It may describe the protagonist getting rewards first, then write about the protagonist completing the quest, or even skip the quest process and go directly to the ending, leaving out the "cause" and only writing about the "effect". Such a story is not only confusing, but also seriously undermines the plausibility and challenge of the game's plot.

DOI: 10.54254/2755-2721/2025.PO25184

Experiments have shown that LLM generates 40% of the episodes with disordered chronological order [4]. Studies have shown that LLM, when dealing with sequences of events, tends to rely on the order in which the events appear in the text to infer causality, i.e., it is assumed that the event mentioned first is the cause and the event mentioned later is the effect. This kind of inference based on the apparent order leads to the model's inability to accurately identify the real causal relationship when the sequence of events is disrupted or presented in a non-linear way, which in turn leads to logical confusion.

This limitation makes the model prone to problems such as reversed causal relationships and chaotic chronological order when generating complex plots, which seriously affects the logic of the narrative and the player experience.

3.4. Input control and content controllability challenges

LLM is best at "freewheeling", where you give it a few hints and it generates a huge chunk of content. But because it's so free, the model often can't control itself: it can write a stylistically inconsistent paragraph, a medieval scene with "motorized skateboards", or a character that suddenly becomes a "funny man". In plot generation, and especially in game design, developers often need to control the style, pace, and logic of the story to ensure a consistent player experience. The conflict between this need and the free nature of the model constitutes one of the core challenges of plot generation today.

It is therefore evident that in game plot generation, input control and content controllability are key factors in ensuring narrative coherence and player experience. LLMs often face the challenge of balancing free generation and content coherence when generating text. The lack of an effective input control mechanism may cause the generated plot to deviate from the expected theme, with problems such as incoherent plots or inconsistent character behaviors, destroying the immersion of the game. In addition, when dealing with complex narrative structures, LLM may generate content that contradicts the established game setting, leading to plot holes. This reflects the limitations of LLM in terms of content controllability and emphasizes the need for fine-grained control of LLM inputs in game development to ensure that the output content is consistent with the game setting and player expectations.

Table 1 illustrates a comparative overview of the key characteristics of free versus controlled generation, highlighting the trade-offs between creative flexibility and narrative stability.

DOI: 10.54254/2755-2721/2025.PO25184

Table 1. Comparison between free and controlled generation approaches

Dimension (math.)	Free Generation	Controlled Generation
freedom of content	High, almost unlimited	Limited by input controls, structural templates or rules
Output Stability	Low, output results are highly affected by randomness	High, repeatable and predictable structure and style
Narrative logic maintained	Easily breaks or creates jumps	Clearly structured, with a clear sequence of cause and effect
personalized expression	Stronger, but easy to lose control	Can limit the range of personalities and express consistency
stylistic consistency	Easily deviates from initial settings	More uniformity of style, e.g., consistent humor, seriousness, etc.
Can be used for game design	High risk and requires extensive manual intervention	Suitable for integration with state machines and plot scripts for greater stability
Representative Technology	GPT-3, GPT-4 Generated by default	RLHF tuning models, template-guided generation, hybrid structural models

4. Trends and technology perspectives of LLM in the field of plot generation

4.1. Memory enhancement and the construction of role congruence

In game plot generation, consistency of character behavior and narrative is crucial for player immersion. However, current LLMs often lack effective memory mechanisms when dealing with long texts, resulting in unstable character traits and behaviors that affect story coherence.

To address this problem, researchers have proposed a variety of approaches to enhance LLM memory and character coherence. For example, the SCORE framework enhances coherence and emotional consistency of story generation through components such as dynamic state tracking, context-aware summarization, and hybrid retrieval [5]. In addition, Relational Memory-Augmented Language Models improve the coherence and consistency of generated text by integrating relational triples from knowledge graphs into language models [6]. These approaches enable the models to refer to prior character information and plot development, thus maintaining consistency in character traits and behaviors, and enhancing the quality of the game narrative and player experience.

4.2. Reasoning chain construction and plot logic optimization

In game plot generation, ensuring the rigor and coherence of the story logic is crucial for enhancing player immersion. However, when dealing with complex narratives, current LLMs often lack a clear chain of reasoning, resulting in unreasonable plot development. In order to optimize the plot logic, researchers have introduced the Chain-of-Thought (CoT) cueing technique, which encourages models to display a series of intermediate reasoning steps before generating an answer, thus enhancing their ability to handle multi-step reasoning tasks. Technology has been shown to improve accuracy on complex logic tasks from 17% to 57%, providing a practical path for causal chain construction in game plots [7].

In addition, frameworks that combine symbolic logic solvers with LLMs, such as Logic-LM, significantly improve the model's performance in logical reasoning tasks by transforming natural language problems into symbolic representations, and then utilizing symbolic solvers for reasoning. The application of these methods helps to construct a clear chain of reasoning and ensure the

DOI: 10.54254/2755-2721/2025.PO25184

rationality and consistency of the plot development, thus enhancing the quality of the game narrative and the player experience.

4.3. Integration design: control and restraint mechanisms

In game plot generation, the integration of control and constraint mechanisms is crucial to ensure the coherence and controllability of the content. Currently, LLMs may suffer from stylistic drift or deviation from predefined settings when generating text freely. To address this challenge, researchers have introduced various control techniques, such as Conditional Transformer Language Model (CTRL) and Future Discriminator (FUDGE) , to impose specific constraints during the generation process to ensure that the output content conforms to expectations. In addition, combining narrative templates and retrieval enhancement methods can further improve the stability and consistency of the generated content. The future trend is to build "scriptable LLMs", which allow developers to set hard and soft constraints, and realize the synergistic programming between humans and models, so as to create game plots that meet the design intent as well as creativity.

4.4. Future-proof plot generation system scenarios

Future storyline generation systems will deeply integrate AI technologies, particularly LLM, to enable highly personalized and dynamic narrative experiences. These systems will be able to analyze player choices and behaviors in real time to generate storylines that match them, ensuring that each player experiences a unique gameplay journey. For example, the PANGeA framework demonstrates the potential for dynamic story generation in turn-based role-playing games by guiding LLMs to generate narrative content that meets the high-level standards of game designers. In addition, future systems may integrate multimodal AI technologies that combine text, voice, and visual elements to create more immersive narrative environments. However, the realization of these technologies needs to ensure diversity and innovation in the generated content while maintaining narrative coherence, character consistency, and logical rationality. Therefore, future research should focus on how to strike a balance between automated generation and human creation, leveraging the efficiency of AI while preserving the unique perspectives and creativity of human creators.

5. Conclusion

This study focuses on the personalization and reasoning capabilities of large-scale language models (LLMs) in game plot generation, and systematically explores four aspects, namely, the current state of technology application, key capability dimensions, major types of challenges, and future development trends. Under the background that highly interactive and open narrative has become the mainstream of game design, LLM provides new possibilities for dynamic plot construction by virtue of its natural language generation capability. However, whether the model can effectively reflect the player's personalized choices and whether it has a reasoning chain that is in line with human cognitive logic has become the core of the evaluation of its practical usability.

Focusing on the application of Large Language Models (LLMs) in game plot generation, future research should continue to focus on the deep enhancement of personalized expression and reasoning capabilities, and expanding their potential in long-term interactive and immersive narrative systems. Research priorities will include how to achieve character memory persistence across rounds and episodes, how to build narrative engines with causal understanding, and how to

DOI: 10.54254/2755-2721/2025.PO25184

allow AI to truly understand player choices and generate appropriate plot development and character responses based on their behavior.

The future of the plot generation system is no longer just a "story generating machine", but may become the player's narrative partner - not only can it remember who you are, what you have done, but also based on your words to develop a world that is uniquely yours. Imagine a game where every action of the player is recorded by the AI and integrated into the character's perception, and the story is no longer a unified script "for everyone to hear" but a private narrative "written just for you".

In addition, future research should also focus on the supporting role of LLM in the game creation process, such as how to collaborate with human screenwriters, how to improve the quality of draft script generation, and how to improve the structural control ability without sacrificing creativity. The ultimate goal is to promote the evolution of game narrative generation from a "technical tool" to a "creative partner".

References

[1] PR Newswire. (2024, May 28). Artificial Intelligence (AI) in games market to grow by USD 27.47 billion from 2025–2029 – Boosted by AR and VR game adoption: Report on AI's role in market transformation – Technavio. https: //www.prnewswire.com/news-releases/artificial-intelligence-ai-in-games-market-to-grow-by-usd-27-47-billion-from-2025-2029--boosted-by-ar-and-vr-game-adoption-report-on-ais-role-in-market-transformation--technavio-302372175.html

[2] Market.us. (2024, December). AI in gaming market. https: //market.us/report/ai-in-gaming-market/

[3] Zeng, Y., Adams, N. M., & Bodenham, D. A. (2024). On two-sample testing for data with arbitrarily missing values. arXiv preprint arXiv: 2403.15327.

[4] Maharana, A., Lee, D. H., Tulyakov, S., Bansal, M., Barbieri, F., & Fang, Y. (2024). Evaluating very long-term conversational memory of llm agents. arXiv preprint arXiv: 2402.17753.

[5] Yi, Q., He, Y., Wang, J., Song, X., Qian, S., Yuan, X., ... & Shi, T. (2025). Score: Story coherence and retrieval enhancement for ai narratives. arXiv preprint arXiv: 2503.23512.

[6] Liu, Q., Yogatama, D., & Blunsom, P. (2022). Relational memory-augmented language models. Transactions of the Association for Computational Linguistics, 10, 555-572.

[7] Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., ... & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. Advances in neural information processing systems, 35, 24824-24837.