Proceedings of CONF-SPML 2026 Symposium: The 2nd Neural Computing and Applications Workshop 2025
DOI: 10.54254/2755-2721/2026.TJ29592

Evaluating the Effectiveness of Transformer-Based NLP Models in Maintaining Character Consistency (Personality and Tone) in Multi-Turn Game Dialogues

Qiyuan Zheng

Beijing Huiwen Middle School, Beijing, China
zhengqiyuan@me.com

Abstract. Non-Player Character (NPC) dialogue systems in computer games have undergone a significant evolution from fixed scripts to dynamic generation. With the widespread adoption of deep learning models like Transformers, NPC language generation has become increasingly natural and personalized. However, the preservation of character consistency across multi-turn interactions, including personality, tone, and knowledge, persists as a key challenge in current studies and implementations. This paper examines the progress of Transformer-based Natural Language Processing (NLP) models in NPC dialogue generation, with particular attention to studies that compare model selection such as GPT-2, DialoGPT, and GPT-Neo, approaches for sustaining consistency including fine-tuning, knowledge graphs, and rewriting, and evaluation using both automated metrics and human judgment. Through an analysis of existing literature, the results indicate that NPC character consistency is challenged by issues such as persona drift, tonal shifts, and knowledge forgetting, and that existing methods remain limited. Future efforts should integrate memory, multimodal context, and personalization for more immersive and authentic interactions.

Keywords: Non-Player Character (NPC) Dialogue, Character Consistency, Transformer-based NLP Models, Multi-turn Conversations, AI-driven Dialogue Systems

1. Introduction

In computer games, dialogue systems for non-player characters (NPCs) have evolved from static scripts into systems capable of proactive, autonomous AI-generated conversations. Recent studies have shown that Transformer-based NLP models achieve greater coherence and accuracy in dialogue generation, thanks to their self-attention mechanism that captures long-range context and maintains grammatical alignment, as confirmed in low-resource NMT studies [1]. Meanwhile, these models have shown strong capabilities in creative generation tasks and exhibit advantages in language generation beyond gaming contexts [2]. For example, the combination of procedural content generation (PCG) and Transformer architectures can capture complex underlying structures, thus enabling the creation of more diverse and creative content. However, maintaining the consistency of an NPC's character, including personality, tone, and knowledge, remains a significant challenge in long, multi-turn interactions. The proposed system in existing studies combines knowledge fine-

© 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: 10.54254/2755-2721/2026.TJ29592

tuning with dynamic knowledge graphs to ensure that NPCs' language aligns with their cognitive scope, while allowing real-time adjustments based on player interactions [3]. This approach provides a reference for using Transformers to preserve character consistency in games. Therefore, this paper examines the application of Transformer-based models in NPC dialogue generation, analyzing their strengths and limitations in maintaining character consistency. In addition, it compares different studies in terms of model selection, methodological implementation, and evaluation frameworks, and investigates the potential of enhancing NPC immersion and authentic interaction through memory mechanisms, multimodal context, and personalized modeling.

2. Evolution of NPC dialogue systems and transformer applications

2.1. Evolution of NPC dialogue systems

NPC dialogue systems have undergone years of development, beginning with simple scripted systems and evolving into more sophisticated AI models. Initially, these dialogue systems were fully scripted and offered few branching options, which limited player interaction and reduced immersion. The use of more sophisticated AI models by NPCs in games began, keeping pace with game difficulty and plot alterations [4]. However, neural networks did not solve the problem of long-conversation coherence, even as increasingly complex models like Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks emerged. It was then that the Transformer architecture was introduced, which proved to be a game changer.

2.2. Transformer applications in NPC dialogue

By leveraging self-attention, Transformers can track long-range relationships and process extended dialogue effectively, resulting in more sophisticated NPC-player interactions [5]. Besides, these models have been adopted to creative tasks, such as procedural content generation and dynamically generated stories, resulting in NPCs that are more personalized and interactive [4]. The adoption of Transformer-based models reshaped the aims of NPC dialogue systems, enabling them to go beyond basic language fluency and address more complex goals, such as trait consistency and character development. This adjustment is required to ensure that these NPCs display consistent personalities, which helps enhance player engagement in the game [6]. Through the recreation of the static knowledge and the dynamic knowledge graphs, NPCs can now generate dialogues that reflect both their identity and the game environment, enabling more realistic interactions with players. Beisdes, it suggests that NPC behavior is not yet fully optimized, and leveraging evolutionary algorithms alongside knowledge graphs could improve their responsiveness, leading to more natural and flexible NPCs [7]. Thus, the development of NPC dialogue systems marks a move from rule-based to adaptive, self-learning AI. As NPCs advance through Transformers and AI, chances for immersive and personalized gameplay increase.

3. Major challenges in maintaining NPC character consistency in multi-turn dialogues

3.1. Key challenges in multi-turn NPC dialogue

Transformer-based NLP models have markedly improved the fluency and coherence of NPC dialogues in online computer games. However, as NPCs engage in continuous interactions with players, ensuring a stable personality, tone, and knowledge across multi-turn dialogues is crucial. If

DOI: 10.54254/2755-2721/2026.TJ29592

consistency is not maintained, the narrative may be disrupted, leading to lower player engagement and immersion [8].

In multi-turn interactions, a central challenge is character drift, in which NPC-generated dialogue gradually deviates from its intended persona. Although persona-based dialogue generation models can ensure semantic consistency for individual turns, they often cannot uphold coherent personality, tone, and knowledge throughout multi-turn interactions, particularly in dynamic scenarios with unpredictable player behavior [9]. Existing approaches, including persona selection classifiers and dialogue rewriting, incorporate persona-specific memory to align responses with the character's traits. Nevertheless, these mechanisms have limitations in that memory retention is often insufficient, contextual embeddings may not capture long-term links, and slow shifts in tone or knowledge can reduce persona stability.

3.2. Transformer-specific challenges in multi-turn dialogue

Transformer models, such as GPT-3 and DialoGPT, have demonstrated the ability to generate fluent, context-aware, and coherent dialogue across a wide variety of interactive and narrative scenarios [1]. However, sustaining a consistent character persona over extended multi-turn conversations remains challenging. In most Transformer architectures, the inadequate memory retention mechanisms make it difficult to preserve long-range dependencies and cumulative context, frequently resulting in character drift [9,10]. As a result, personality representations can become fragmented, leading NPCs to produce dialogue that gradually diverges from their established traits. Moreover, these models are prone to tonal inconsistencies and knowledge forgetting, which may cause the NPC to contradict earlier statements or display behaviors inconsistent with their persona. Such failures not only disrupt narrative coherence but also diminish the perceived credibility and immersion of the game world, particularly in interactive settings where player input is unpredictable and dialogue paths are highly dynamic.

4. Methodological approaches and evaluation frameworks

The consistency of NPC dialogue in games is crucial for enhancing immersion and realism. As such, various approaches have been proposed to ensure character consistency, mainly including dialogue generation strategies, memory and knowledge integration mechanisms, and evaluation frameworks.

4.1. Dialogue generation methods

To enhance NPC character traits and ensure consistent dialogue, current approaches primarily utilize fine-tuned pre-trained Transformer models along with dialogue rewriting methods.

In terms of fine-tuning, typical pre-trained models include GPT-2 and DialoGPT. These models can be further trained on character-specific datasets to reflect NPC personality traits and generate relatively stable and predictable responses [9]. Due to its open-source nature, GPT-Neo offers greater flexibility for fine-tuning, allowing it to absorb personality and knowledge from character-specific datasets, producing dialogue that better aligns with character design [11,12]. Compared to GPT-2 and DialoGPT, GPT-Neo provides higher flexibility and controllability for personalized dialogue generation. Besides, dialogue rewriting techniques modify the generated dialogue via a secondary model to correct issues in tone or knowledge consistency, making the output more consistent with the character's personality [13]. Moreover, the combination of fine-tuning and

DOI: 10.54254/2755-2721/2026.TJ29592

dialogue rewriting can improve the consistency of multi-turn NPC dialogue and match with character traits, making interactions more natural and believable.

4.2. Memory and knowledge integration

To maintain NPC character consistency in multi-turn dialogues, memory mechanisms and knowledge graphs have been introduced to assist dialogue generation. By storing a character's basic information, background knowledge, and current game state, knowledge graphs allow the model to access pertinent facts and historical interactions during dialogue generation, thereby ensuring that important contextual information is retained [5]. For example, when a player asks multiple questions or makes choices over several turns, the knowledge graph helps the NPC recall prior events, relationships, and decisions, thus preserving logical consistency and character coherence in the dialogue.

Moreover, memory mechanisms enhance both the continuity and adaptability of dialogue generation. While short-term memory retains information from the ongoing dialogue, such as the player's instant feedback and contextual hints, long-term memory collects historical data regarding the character's actions and personality features. These memory systems help maintain consistent personality, tone, and knowledge across extended multi-turn interactions. The memory module dynamically retrieves relevant information for each generated utterance by leveraging optimized attention weights and contextual embeddings, which reduces character drift, tone shifts, and knowledge forgetting. These mechanisms are often combined with fine-tuning and dialogue rewriting. Fine-tuning trains the model on character-specific language and response patterns, knowledge graphs and memory modules ensure information consistency, and dialogue rewriting adjusts any remaining deviations in tone, knowledge, or personality.

4.3. Evaluation methods and comparative analysis

The consistency of NPC dialogue is commonly measured using both automated indicators and human evaluation. Automated metrics, like perplexity, BLEU scores, and embedding similarity, measure the fluency of generated dialogue and its alignment with context [14]. Besides, human evaluation targets subjective perceptions of character traits, tone, and knowledge maintenance, mitigating the inability [15]. In practice, different methods exhibit distinct strengths in multi-turn dialogues. Fine-tuned models can learn character-specific language styles, knowledge graphs and memory mechanisms help maintain long-term information consistency, and dialogue rewriting can correct tone or knowledge deviations. These methods can be combined to enhance dialogue stability and immersion. As shown in Table 1, the main advantages and limitations of each method in maintaining character consistency are provided to guide model selection and integrated application.

DOI: 10.54254/2755-2721/2026.TJ29592

Table 1. Evaluation of techniques for preserving NPC role consistency

Method	Advantages	Limitations
GPT-2 Fine-tuning	Stable training, fluent grammar	Limited personalization, insufficient long-term consistency
DialoGPT Fine-tuning	Better contextual coherence	Character drift still occurs in long dialogues
GPT-Neo Fine-tuning	High flexibility, supports character personalization	High dependency on large-scale data
Knowledge Graphs & Memory Mechanisms	Supports long-term information retention, reduces knowledge forgetting	High construction and maintenance cost
Dialogue Rewriting	Corrects tone and knowledge deviations	Affects generation speed, difficult to apply in real-time

5. Conclusion

This paper presents an overview of the literature available on the use of the Transformer-based models to retain the same character in the NPC dialogues. The field has made a lot of progression whereby it is no longer in the use of scripts but models which provide fluent and contextually reactive conversations. Nevertheless, persistent challenges remain, including character drift, knowledge forgetting, and tonal inconsistencies, which highlight limitations of current architectures in supporting long-term, persona-aware interactions. Existing solutions, like fine-tuning, memory augmentation, and dialogue rewriting, boost consistency but are limited by evaluation complexity and the reliance on both automated metrics and human judgment. Future research should focus on developing more advanced memory structures, techniques to disentangle style and content, and standardized benchmarks to systematically assess long-term character fidelity. The above barriers can only be overcome to achieve the potentials of AI-driven characters in pure text and real worlds of pure text-based games.

References

[1] Javed, A., Zan, H., Orken Mamyrbayev, et al. (2025). Transformer-based re-ranking model for enhancing contextual and syntactic translation in low-resource neural machine translation. Electronics, 14(2), 243

[2] Zhu, S., Ma, T., Rong, H., & Al-Nabhan, N. (2023). A personalized multi-turn generation-based chatbot with various-persona-distribution data. Applied Sciences, 13(5), 3122.

[3] Liu, X., Xie, Z., & Jiang, S. (2025). Personalized non-player characters: A framework for character-consistent dialogue generation. AI, 6(5), 93.

[4] Du, H. (2025). The progress and trend of intelligent NPCs in games. Applied and Computational Engineering, 133(1), 158-164.

[5] Zhang, E. Y., Cheok, A. D., Pan, Z., et al. (2023). From Turing to transformers: A comprehensive review and tutorial on the evolution and applications of generative transformer models. Sci, 5(4), 46.

[6] Liu, X., Xie, Z., & Jiang, S. (2025). Personalized non-player characters: A framework for character-consistent dialogue generation. AI, 6(5), 93.

[7] Armanto, H., Rosyid, H. A., Muladi, & Gunawan. (2025). Improved non-player character (NPC) behavior using evolutionary algorithm - A systematic review. Entertainment Computing, 52, 100875.

[8] Zhao, T., & Fan, Z. (2024). Enhancing procedural game level generation using transformer-based neural architectures. International Symposium on Internet of Things and Smart Cities (ISITSC), 1-6.

[9] Zou, M., Meng, Y., Gómez, S. C., & Sabina, J. (2025). Persistence and evolution within interactive design: An integrated approach to ICT innovations in emergent game narratives. Technologies, 13(5), 179.

DOI: 10.54254/2755-2721/2026.TJ29592

[10] Pu, C., Sun, Z., Li, C., et al. (2024). Dialogue-rewriting model based on transformer pointer extraction. Electronics, 13(12), 2362.
[11] Black, S., Biderman, S., Hallahan, E., Anthony, Q., Gao, L., Golding, L., et al. (2022). GPT-NeoX-20B: An open-source autoregressive language model. arXiv.
[12] Yang, Y., Cao, J., Wen, Y., & Zhang, P. (2022). Multiturn dialogue generation by modeling sentence-level and discourse-level contexts. Scientific Reports, 12(1).
[13] Wang, W., Feng, S., Chen, L., et al. (2021). Learning to improve persona consistency in conversation generation with information augmentation. Knowledge-Based Systems, 228, 107246.
[14] Nguyen-Mau, T., Le, A. C., Pham, D. H., & Huynh, V. N. (2023). An information fusion based approach to context-based fine-tuning of GPT models. Information Fusion, 104, 102202.
[15] Wang, W., Feng, S., Chen, L., Wang, D., & Zhang, Y. (2021). Learning to improve persona consistency in conversation generation with information augmentation. Knowledge-Based Systems, 228, 107246.