

CharacterMind: An AI-Powered Fictional Character Simulation System for Immersive Conversational Experiences

Samanyu Srivastava ¹, Vikram Yadav ¹, Chaitanya Kumar ¹, Somya Tiwari ²

¹ School of Engineering and Technology, Sushant University, Gurgaon, Haryana, India

² Associate Professor, School of Engineering and Technology, Sushant University, Gurgaon, Haryana, India

O 0000-0003-1106-9700

Abstract

The rapid evolution of conversational artificial intelligence has been largely driven by advances in large-scale language models capable of generating coherent, context-aware, and linguistically rich responses. These systems demonstrate strong performance across a wide range of applications, including virtual assistance, customer support, storytelling, and knowledge retrieval. Despite these advancements, one persistent limitation continues to affect their usability in immersive and long-term interactions: the inability to maintain a consistent persona.

While modern conversational models can imitate a specific tone or character in short interactions, they often fail to sustain that identity across extended dialogue. Over time, responses tend to shift toward a neutral or generic style, resulting in a loss of character authenticity and reduced user engagement. This phenomenon, commonly referred to as persona drift, highlights the absence of explicit mechanisms for enforcing personality constraints during response generation.

This research presents a structured framework for developing persona-consistent conversational AI systems by integrating transformer-based language models with explicit persona encoding and retrieval-augmented memory. Unlike traditional approaches that rely primarily on prompt-based conditioning, the proposed system introduces a multi-layer architecture that reinforces persona alignment throughout the generation pipeline. A persona consistency filter is employed to evaluate generated responses and ensure alignment with predefined character attributes.

The system is evaluated using both automated metrics and human-centered assessments, including persona consistency scores, contextual coherence measures, and user engagement ratings. Experimental findings indicate that the integration of persona embeddings and memory retrieval significantly improves identity retention and conversational continuity. The proposed approach demonstrates measurable improvements over baseline systems in maintaining tone, behavior, and contextual alignment.

Introduction

Conversational artificial intelligence has undergone a remarkable transformation over the past decade. Early systems relied on rule-based frameworks that followed predefined scripts and lacked flexibility, resulting in rigid and predictable interactions. The introduction of machine learning enabled adaptive responses, but the real breakthrough came with transformer-based architectures, which allow models to capture contextual relationships across long sequences using attention mechanisms.

Despite these advances, conversational AI systems still struggle to replicate one of the most essential aspects of human interaction: personality. Human conversations are shaped by consistent behavioral patterns, emotional tendencies, and communication styles. Whether interacting with a fictional character or a real individual, users expect a coherent identity.

In contrast, most AI systems lack a stable persona. While they can generate accurate and grammatically correct responses, they fail to maintain consistency over time. This issue becomes especially critical in applications such as storytelling, gaming, and virtual assistants, where personality plays a central role in user engagement.

One of the primary causes of this limitation is the probabilistic nature of language generation. Transformer models generate responses based on statistical likelihood rather than deterministic rules, making it difficult to enforce strict persona constraints. Additionally, limited context windows result in loss of earlier conversation history, further contributing to inconsistency.

Emotional inconsistency is another significant issue. A character defined as calm and wise may suddenly produce aggressive or contradictory responses, breaking immersion. These challenges highlight the need for a system that can maintain both personality and context over extended interactions.

This research proposes a persona-consistent conversational AI system that integrates persona encoding, memory retrieval, and response validation mechanisms. The goal is to create a system that not only produces accurate responses but also maintains a coherent identity throughout the conversation.

Objectives of the Study

The primary objective of this research is to develop a conversational AI system capable of maintaining consistent character personas across extended interactions.

The specific objectives include:

- Designing a framework that embeds persona traits directly into the response generation process
 - Enabling users to interact with AI characters through a conversational interface
 - Generating responses that reflect consistent tone, vocabulary, and behavior
 - Implementing a memory system to preserve contextual continuity
 - Introducing a validation mechanism to ensure persona alignment
 - Evaluating system performance using both qualitative and quantitative methods
 - Comparing persona-based AI with traditional chatbot systems
-

Scope of the Study

This research focuses on developing a prototype system that demonstrates persona-consistent conversational capabilities.

The system includes:

- A text-based interaction interface
- Structured persona modeling using attribute-based representations
- NLP-based input processing
- A memory retrieval mechanism for context preservation

The study does not include voice interaction, large-scale deployment, or advanced emotional intelligence modeling. The focus is on demonstrating feasibility and establishing a foundation for future work.

Literature Review

The development of conversational AI has been significantly influenced by transformer-based models, which enable contextual understanding through self-attention mechanisms. These models have demonstrated strong performance in language generation tasks but lack explicit mechanisms for maintaining consistent personas.

Early persona-based systems relied on datasets such as Persona-Chat, which introduced structured persona descriptions into dialogue generation. While these systems improved short-term consistency, they struggled with long-term persona retention.

Retrieval-Augmented Generation (RAG) introduced memory mechanisms that allow models to retrieve relevant information during response generation. This approach improves contextual awareness but does not enforce personality consistency.

More advanced systems such as BlenderBot incorporate dialogue history and long-term memory to improve coherence. Fine-tuning techniques such as LoRA and QLoRA enable efficient adaptation of models for specific tasks, including character simulation.

Despite these advancements, maintaining a consistent persona remains a challenge. Existing systems either lack memory, fail to enforce persona constraints, or struggle with long-term consistency. This research builds upon these approaches by integrating persona modeling, memory retrieval, and validation mechanisms.

Methodology

The proposed system follows a modular architecture consisting of five key components: input processing, persona encoding, memory retrieval, response generation, and persona validation.

Input Processing

User input is processed using transformer-based encoders, which convert text into contextual embeddings. These embeddings capture semantic meaning and contextual relationships.

Persona Encoding

Each character is represented as a structured persona profile consisting of attributes such as personality traits, tone, and behavioral tendencies. These attributes are encoded into vector representations, allowing them to influence the response generation process.

Memory Retrieval

The system stores past interactions as embedding vectors. When a new input is received, relevant memories are retrieved using similarity measures such as cosine similarity. This enables the system to maintain context across multiple conversational turns.

Response Generation

The response generation module uses a transformer-based decoder conditioned on input, persona, and memory. This ensures that responses are both contextually relevant and aligned with the character's identity.

Persona Validation

A persona consistency filter evaluates generated responses using a scoring function that considers both persona alignment and contextual coherence. Responses that do not meet the required threshold are refined or regenerated.

Results

The system was evaluated across multiple scenarios to assess persona consistency, contextual awareness, and user engagement.

Results indicate that the system successfully maintains persona consistency in most interactions. Responses reflect appropriate tone and behavior, demonstrating alignment with predefined character traits.

The memory module significantly improves contextual awareness, allowing the system to maintain continuity and avoid repetition. Users reported higher engagement levels compared to traditional chatbots, highlighting the effectiveness of persona-based interaction.

However, minor deviations were observed in longer conversations, indicating the presence of residual persona drift.

Discussion

The integration of persona modeling and memory mechanisms significantly enhances conversational AI performance. The system successfully maintains character identity across multiple interactions, improving realism and engagement.

The memory module plays a crucial role in maintaining context, while the persona validation mechanism ensures consistency. Together, these components create a more immersive conversational experience.

However, challenges remain. There is a trade-off between creativity and consistency, as strict persona enforcement may limit response diversity. Additionally, performance may decline in very long or complex conversations.

Future improvements may include advanced memory systems, reinforcement learning, and improved emotional modeling.

Expected Outcomes

The system is expected to:

- Maintain consistent character personalities
 - Provide engaging and immersive interactions
 - Improve conversational quality compared to traditional chatbots
 - Demonstrate applications in storytelling, gaming, education, and entertainment
-

Limitations

The system has several limitations:

- Occasional deviation from persona in long conversations
 - Limited data availability for certain characters
 - Difficulty handling highly complex queries
 - Computational constraints affecting performance
-

Conclusion

This research demonstrates that persona-consistent conversational AI systems can significantly improve the quality of human-computer interaction. By integrating persona encoding, memory retrieval, and validation mechanisms, the system maintains character identity, enhances contextual awareness, and increases user engagement.

The findings highlight the importance of personality in conversational AI and provide a foundation for future research. As AI continues to evolve, persona-consistent systems have the potential to transform digital interaction by enabling more natural and immersive conversations.

References

- [1] S. Zhang et al., "Personalizing dialogue agents," ACL, 2018.
- [2] A. Vaswani et al., "Attention is all you need," NeurIPS, 2017.
- [3] P. Lewis et al., "Retrieval-augmented generation," NeurIPS, 2020.
- [4] K. Shuster et al., "BlenderBot 3," 2022.
- [5] E. Hu et al., "LoRA," 2022.
- [6] T. Dettmers et al., "QLoRA," 2023.
- [7] X. Wang et al., "CharacterChat," 2023.
- [8] D. Jurafsky & J. Martin, *Speech and Language Processing*, 2023.
- [9] I. Goodfellow et al., *Deep Learning*, 2016.