

Automated Dialogue Generation for Video Games

Sahil. Anant. Belurkar¹, Prof. Sandarsh Gowda M M²

¹ Student, Department of MCA, Bangalore Institute of Technology, Karnataka, India

² Professor, Department of MCA, Bangalore Institute of Technology, Karnataka, India

Abstract

"Automated Dialogue Generation for Video Games (ADGVG)," addresses the need for dynamic and context-aware dialogue in modern video games. Traditional dialogue systems, which are manually scripted and rule-based, are often repetitive, costly to produce, and lack the flexibility to adapt to player actions. This report presents a framework that integrates large, pre-trained language models (LLMs) with a retrieval-augmented generation (RAG) pipeline and prompt engineering to create coherent, engaging, and character-specific dialogues for non-player characters (NPCs). By grounding generative models in curated knowledge bases and using structured dialogue turn-taking, this approach overcomes key challenges such as maintaining personality consistency, long-term narrative coherence, and avoiding "hallucinated" outputs. Evaluation, which combines automatic metrics with human-centered testing, demonstrates that this hybrid system produces more consistent and engaging dialogues compared to traditional methods. The ADGVG framework provides a scalable and versatile solution with broad applications in interactive storytelling, training simulations, and conversational AI beyond the gaming industry.

Keywords—Automated Dialogue Generation, Video Games, NLG, Large Language Models, Retrieval-Augmented Generation, Prompt Engineering, NPC Dialogue, Context-Aware Dialogue, Interactive Storytelling, Conversational AI.

I. INTRODUCTION

Dialogue is a fundamental component for crafting immersive experiences in video games, especially within the narrative-heavy genres of role-playing games (RPGs) and open-world titles. Non-player characters (NPCs) serve as crucial narrative agents, guiding players through storylines, providing information, and enriching the game world. Historically, the creation of NPC dialogue has been a labour-intensive process, relying on manually authored scripts and fixed dialogue trees. While this approach offers predictability and full narrative control, it has significant drawbacks: it is resource-intensive, leads to repetitive interactions, and struggles to scale to the sprawling environments of modern games.

The rapid evolution of Natural Language Generation (NLG) offers a promising alternative. Modern transformer-based architectures, such as GPT-2 and GPT-3, can produce fluent, human-like responses, opening new possibilities for dynamic and adaptive interactions. However, deploying these powerful generative systems in a domain-specific context like video games introduces a unique set of challenges. A model's output must not only be grammatically correct but also maintain a

character's consistent personality, adhere to established game lore, and avoid generating information that contradicts the game world—a phenomenon often referred to as "hallucination".

The ADGVG project aims to bridge this critical gap by designing a hybrid system that combines the creative flexibility of large language models with the structured control of retrieval mechanisms and prompt engineering. This report details a framework that leverages curated datasets and a RAG pipeline to generate dialogues that are both engaging and contextually grounded. The following sections will provide a comprehensive overview of existing systems, the proposed ADGVG framework, its technical implementation, and a discussion of its potential impact and future directions.

II. LITERATURE SURVEY

Early natural language generation was dominated by rule-based systems and Context-Free Grammars (CFGs). These methods, exemplified by work like the Talk of the Town dialogue generation system, provided authors with a high degree of structural control and predictability over the generated output. The system used an authoring tool called Expressionist to specify CFGs with explicit metadata, allowing for a hierarchy of symbolic markup to guide text production. However, this control came at the cost of linguistic diversity and scalability, as these models lacked the flexibility to produce novel and varied text beyond their predefined rules.

The rise of neural networks introduced a new paradigm. Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTMs) became popular for sequence modelling, as they could generate text iteratively by predicting the next token in a sequence. This approach offered a greater degree of creativity than rule-based systems but struggled with long-term dependencies, meaning they often lost context in longer texts. A key limitation of these early neural methods was that they were monolithic, end-to-end systems, which stood in stark contrast to earlier, more modular approaches that divided the generation process into distinct phases, such as Text Planning, Sentence Planning, and Realization. This modularity in older systems allowed for specific control over different aspects of generation, a feature that modern models often forgo in favour of a single, powerful, end-to-end network.

The introduction of the Transformer architecture and its attention mechanism in 2017 fundamentally changed the landscape of NLG. By allowing a model to weigh the importance of different words in a sequence, the Transformer was able to process long-range dependencies with unprecedented efficiency and accuracy, effectively overcoming the limitations of RNNs. This innovation led

directly to the development of powerful pre-trained language models (PLMs) such as GPT-2.

GPT-2 and its successors, fine-tuned on massive, unlabelled corpora like the Web Text dataset, learned a deep, robust latent representation of the English language. This pre-training paradigm enabled these models to be highly effective at a wide range of downstream tasks, from general-purpose text generation to more creative applications like generating dialogue for interactive text adventures such as AI Dungeon. The ability of these models to produce text that is often indistinguishable from human-written content has been a major milestone, pushing the boundaries of what is possible in NLG. The shift from older, manual methods to modern, data-driven LLMs represents a move from systems that are "controlled but non-creative" to those that are "creative but uncontrolled". This is the core problem that research is now attempting to solve.

The challenges inherent in story generation, specifically the need to produce a coherent, multi-sentence narrative, have led researchers to explore innovative training paradigms. One such paradigm is framing a task as a "pretext task" for learning broader, more transferable representations.

A similar logic applies to our proposed approach. By training a model to generate text that adheres to a set of commonsense inference rules, the task of "coherence enforcement" becomes a powerful pretext task. The model is forced to learn the intricate causal relationships and emotional connections that drive a narrative forward, rather than simply predicting the most probable next word in a vacuum. This approach transforms story generation from a purely generative problem into a robust training paradigm for representation learning, with the potential to yield more logically and semantically grounded outputs.

A recurring theme in the literature is the need to impose constraints on the open-ended nature of modern generative models to prevent common failures like factual inconsistencies and logical drift. Various methods have been proposed to inject external knowledge and control signals into the generation process. These can be broadly categorized into "soft-constrained" and "hard-constrained" approaches.

Controlled generation in video game dialogue involves multiple techniques to shape character output and improve coherence. Style can be guided explicitly, using brief trait descriptions, or implicitly, using a character's prior lines to allow adaptation over time. Template-Guided Generation (T2G2) employs simple predefined templates representing system actions, which a language model rewrites into fluent, natural text, mitigating representation bias. Meta-information and tagging further enhance structured generation, such as RPG quests, by providing contextual details like titles, locations, and quest givers, ensuring outputs remain coherent and grounded. Additionally, Retrieval-Augmented Generation (RAG) leverages external curated databases to anchor outputs in factual knowledge, helping maintain consistency with a game's lore while generating dialogue or narrative content. These approaches together allow precise control over tone, style, and factual accuracy in AI-generated game content.

A significant theme that emerges from this review is the continuous search for a way to impose authority on a model's generative power. The journey has progressed from brittle, handcrafted grammars to the integration of external, verifiable knowledge. The central challenge is that generative models

lack a fundamental, a priori understanding of cause and effect, leading to a disconnect between sentences. The use of a commonsense knowledge base, as seen in this paper, is a specific, targeted example of this broader movement toward hybrid, grounded systems. This trend indicates that the future of generative AI research will focus less on scaling model size and more on creating sophisticated, modular architectures that can seamlessly integrate external, verifiable knowledge to produce outputs that are not just plausible but also factually and contextually sound.

III. EXISTING SYSTEM

The field of Natural Language Generation (NLG) has progressed from early rule-based and template-driven systems to modern neural network architectures. Initial approaches, such as those using Context-Free Grammars (CFGs), provided precise control and predictability over text generation, exemplified by systems like "Talk of the Town" which used the Expressionist authoring tool to define CFGs with explicit metadata to guide output. However, these methods were rigid, limiting linguistic diversity and scalability, as they could not produce text beyond predefined rules. The emergence of neural networks, particularly Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) models, introduced iterative text generation with greater creativity, though they struggled with long-term dependencies and often lost context in extended narratives. The advent of Transformer architectures and large pre-trained language models (LLMs) such as GPT-2 further advanced NLG by producing fluent, human-like text. Yet, these models are "creative but uncontrolled", lacking structured understanding of cause and effect, which poses challenges in maintaining coherence, logical consistency, and contextual integrity over longer narratives. Moreover, unconstrained generative models are prone to hallucinations or factual inconsistencies due to the absence of structured grounding to validate their outputs.

Disadvantages

- Repetitiveness, players encounter the same dialogue options repeatedly, reducing engagement.
- High development cost, since creating and managing extensive dialogue trees by hand demands considerable effort and resources.
- Lack of context awareness, responses do not adapt to changing player actions or in-game events.
- Limited personality representation, NPCs lack distinct traits or emotions, making them feel generic.

IV. PROPOSED SYSTEM

The proposed system, NPC Dialogue Generator, is an advanced web-based application designed to automate the creation of dynamic and contextually relevant dialogue for non-player characters (NPCs) in video games. Implemented with FastAPI and powered by advanced large language models (LLMs) through the Hugging Face Inference API, the

system allows game developers and writers to create rich, multi-character dialogues tailored to specific narrative contexts and character profiles. Users can submit dialogue requests either through a web interface or by uploading structured JSON files containing detailed character attributes such as name, personality, occupation, and relationship dynamics. The system processes these inputs to construct a highly detailed prompt, ensuring that the generated dialogue not only alternates naturally between characters but also remains consistent with their unique personalities and the overarching story context. By automating this traditionally labor-intensive process, the NPC Dialogue Generator streamlines narrative development, reduces manual scripting workload, and enhances the overall quality and immersion of in-game interactions. The modular architecture, robust error handling, and support for both authenticated and unauthenticated access make the system highly adaptable for integration into various game development pipelines.

Advantages:

- Efficiency, generating context-aware dialogue, reducing time and effort for writers and developers
- Consistency, maintains character personalities, occupations, and relationships, minimizing narrative inconsistencies
- Scalability generates dialogue for any number of characters and contexts, suitable for small and large-scale games
- Customization allows detailed character and context inputs for tailored dialogue outputs
- Integration, designed as a RESTful API for easy incorporation into game development workflows
- Error handling, provides validation and clear messages for malformed inputs
- Accessibility, supports JSON requests and file uploads for diverse user needs

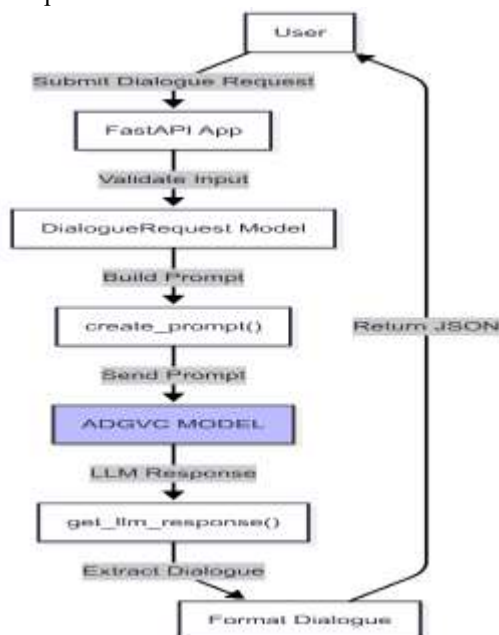


Fig 1: Proposed Model

V. IMPLEMENTATIONS

System Architecture:

The NPC Dialogue Generator is designed as a modular RESTful web application using FastAPI.. It separates authentication, dialogue generation, and data validation, using Hugging Face's LLM for context-aware dialogue and MongoDB for storing user data and API tokens.

Authentication and User Management:

OAuth2 with JWT tokens secures the system. Users can sign up, log in, and generate API tokens, with password hashing handled via passlib and bcrypt.

Input Handling:

Dialogue requests are accepted via JSON POST or file upload, with validation using Pydantic models to ensure correct format and required fields.

Prompt Construction and LLM Interaction:

Prompts are dynamically built with narrative context, character details, and explicit instructions for alternating dialogue. The Hugging Face client sends prompts to LLM and controls output length.

Dialogue Extraction and Post-processing:

LLM outputs are parsed to extract valid dialogue lines, filter non-dialogue content, and apply fallback prompts if needed, returning a structured JSON response with dialogue, model info, and timestamp.

Error Handling and Security:

Input validation, JSON-only file uploads, exception handling, CORS configuration, and environment-based secrets ensure robustness, reliability, and secure integration with frontends.

VI. CONCLUSIONS

The analysis of the current landscape of Natural Language Generation for interactive media reveals a critical situation. While the creative potential of modern, monolithic language models is immense, their application in domains like video game development is hindered by a fundamental trade-off: as creativity increases, factual coherence and narrative consistency tend to decrease. Current models struggle with hallucination, a lack of control, and poor generalization, making them unsuitable for direct, unassisted use in a professional pipeline.

The proposed Automated Dialogue Generation for Video Games (ADGVG) framework offers a promising pathway to overcome these limitations. By treating the generative model not as a knowledge base but as a creative engine, the ADGVG framework offloads the responsibilities of factual verification and structural planning to dedicated modules. This separation of concerns allows the system to provide the imaginative, high-quality prose that modern LLMs are capable of while ensuring that the output is always grounded in the game's lore and adheres to a logical narrative structure. This hybrid approach provides developers with the creative benefits of a generative model, coupled with the control and predictability of traditional systems. The ADGVG framework is not a final solution but a blueprint for a new generation of NLG systems that can intelligently integrate domain-specific knowledge to create truly immersive and dynamic interactive narratives.

VII. FUTURE ENHANCEMENTS

The current research and the proposed ADGVG framework lay a foundation for advancements in automated dialogue generation, but significant challenges remain. One major limitation is the scarcity of high-quality, annotated datasets. Future work should focus on creating more comprehensive datasets that include not only text but also structured metadata on character relationships, emotional context, and narrative motivations, as the lack of such data contributes to factual errors and incoherent narratives. Additionally, there is a pressing need for refined and scalable evaluation metrics. Current metrics, often based on word-overlap or probabilistic scores, fail to capture subjective qualities like "fittingness" and "creativity." Leveraging advanced embedding models to assess semantic and structural coherence could provide a more reliable proxy for human judgment.

Another key direction is the integration of dynamic, in-game testing with human players, enabling the collection of extrinsic metrics that measure player engagement and immersion—the ultimate goals of game development. Furthermore, applying Reinforcement Learning from Human Feedback (RLHF) could help fine-tune models to align more closely with human preferences. By gathering direct feedback on what constitutes a compelling narrative or believable character dialogue, systems could learn nuanced storytelling qualities without relying on manual rule-based interventions. Together, these directions offer a roadmap for building more contextually aware, engaging, and human-aligned dialogue generation systems.

VIII. REFERENCES

- [1] Mihir Kale, Abhinav Rastogi, "Template Guided Text Generation for Task-Oriented Dialogue," arXiv:2004.15006v2 [cs.CL] 13 Nov 2020.
- [2] Cristina Gârbaacea, Qiaozhu Mei, "Neural Language Generation: Formulation, Methods, and Evaluation," arXiv:2007.15780v1 [cs.CL] 31 Jul 2020
- [3] J. Ryan, M. Mateas and N. Wardrip, "Characters Who Speak Their Minds: Dialogue Generation in Talk of the Town," Expressive Intelligence Studio.
- [4] Prerak Gandhi and Pushpak Bhattacharyya, "Survey on Automatic Movie Plot and Script Generation," Center for Machine Intelligence and Data Science Indian Institute of Technology Bombay
- [5] Judith van Stegeren, Jakub Myśliwiec, "Fine-tuning GPT-2 on Annotated RPG Quests for NPC Dialogue Generation," Proceedings of the AIIDE Workshop on Intelligent Narrative Technologies, 2020.
- [6] FU, Zihao, "Open Domain Text Generation," The Chinese University of Hong Kong July 2021
- [7] Gaetan Lopez Latouche and Laurence Marcotte and Ben Swanson, "Generating Video Game Scripts with Style," Ubisoft La Forge {gaetan.lopez-latouche,ben.swanson2}@ubisoft.com laurencemarcotte@hotmail.ca
- [8] Katharina Kann, Abteen Ebrahimi, Joewie J. Koh, Shiran Dudy, Alessandro Roncone, "Open-domain Dialogue Generation: What We Can Do, Cannot Do, and Should Do Next," University of Colorado first.last@colorado.edu.
- [9] Elio Musacchio, Lucia Siciliani, Pierpaolo Basile, Giovanni Semeraro, "Leveraging Large Language Models for Spell-Generation in Dungeons & Dragons," Games and NLP @LREC-COLING-2024, pages 61–69 20-25 May, 2024. © 2024 ELRA Language Resource Association: CC BY-NC 4.0
- [10] Owen Rambow, Srinivas Bangalore, and Marilyn Walker, "Natural Language Generation in Dialog Systems," AT&T Labs—Research Florham Park, NJ, USA.
- [11] Alexandra DeLucia and Aaron Mueller, Xiang Lisa Li, João Sedoc, "Decoding Methods for Neural Narrative Generation," Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 2321–2334, 2022.
- [12] Ali Kalbiyev, "Affective Dialogue Generation for Video Games," Master's Thesis, Utrecht University, 2021.
- [13] Qi chen Gao, Ali Emami, "The Turing Quest: Can Transformers Make Good NPCs?," Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics - Student Research Workshop, pages 93–103 July 10-12, 2023 ©2023 Association for Computational Linguistics
- [14] Taize Kang, "A New Approach of Story Generation Based on Transformers," Master's Thesis, University of Helsinki Faculty of Science, 2022.
- [15] Sander Koomen, "Text Generation for Quests in Multiplayer Role-Playing Video Games," Master's Thesis, November, 2023.