ChatNPC: Towards Immersive Video Game Experience through Naturalistic and Emotive Dialogue Agent

Anonymous ACL submission

Abstract

We present ChatNPC, a game companion designed to respond to players' subtle, unconscious comments and gestures during gameplay. The model integrates sequential data, benefiting from the dynamic nature of streams, including player information, spoken lines, and in-game context. It observes and apprehends players' emotional shifts and adjusts the responses accordingly. We leverage the recent progress in LLM's tool-calling capabilities to extract vital information from memory and recognize potential constraints for accurate reasoning to tackle the complexity of NPC conversation scenarios. The task is divided into: ①, a novel game sentinel agent (*SenGent*); ②, a memory capability; and ③, a chat planning tool for reasoning instantiation. The approach benefits from a lightweight *game-template* as an information framework, with relevant details and a thorough reasoning layout. We conduct extensive experiments on a newly developed game dataset for in-game context NPC dialogue and demonstrate that ChatNPC sequentially captures players' emotional shifts over time; responses are more naturalistic and human-like with appropriate conversational cues, pauses, and sighs; and utterances remain faithful to the dynamics of in-game context and player actions, supporting narrative continuity.

1 Introduction

Most studies present various applications of LLMs *in* and *for* in the broader ecosystem of games and the different roles they can play within a game (Sweetser, 2024). Their theoretical frameworks are extensively based on emulating human behavior to play games at a human or near-human level (Liao et al., 2024) and as quest providers for immersive and personalized gaming experiences for game creators that use LLM as conversational agents (Gallotta et al., 2024). However, an interactive companion intended to enrich or guide the player experience without having to compete or alter the game

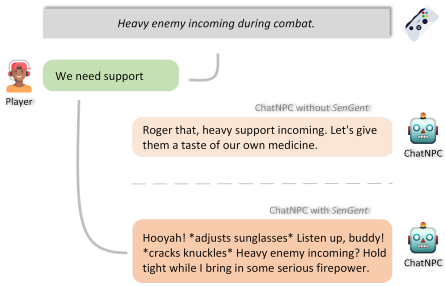

Figure 1: Snippets of ChatNPC responses showing more naturalistic and human-like, with appropriate conversational cues, making it more believable with the *SenGent*.

mechanic remains relatively unexplored. For example, players are allowed to mislead the King into uttering the name of a particular weapon, which eventually materializes in defeating the King in 1001 Nights (Sun et al., 2023). Likewise, in *Gandalf*[1], a player can trick an LLM into exposing a password. A different approach to LLM for user-game immersion could be an interactive agent that does not causally interact with the game world and/or provides a sequence of tutorial-style tips (Gallotta et al., 2024). Alternatively, it could be an agent that can interact with the game world (but not trick or manipulate) while paying attention to subtle comments and gestures of players during gameplay for contextual interaction and emotional connection, as players often convey their emotions through spoken words and cues in video games.

Non-player characters (NPCs) perform various roles and functions, such as quest-givers, vendors, shopkeepers, allies and companions, enemies and adversaries, and supporting characters (Gallotta et al., 2024; Weir et al., 2022; Uludağlı and Oğuz, 2023). They hold critical information about the game to enrich the player's experience by adding to the atmosphere of the game world and making it more believable (Croissant et al., 2023; Uludağlı and Oğuz, 2023). Developing a virtual agent who

[1]https://gandalf.lakera.ai/

engages in meaningful and contextually appropriate dialogue with players throughout gameplay and emphasizes the importance of stealth and situational awareness can provide essential information, plot developments, and mythology that enrich the game world and drive player immersion.

Contrary to research works that engage NPCs through their *dialogue* and *behavior* as agents to play games at a human or near-human level (Toshniwal et al., 2022; Li et al., 2022; , FAIR), we present a game companion that desires to provide dynamic, naturalistic, and emotive systems by tracking and interpreting players' emotional shifts in real-time, adapting its responses dynamically within broader gameplay interactions. This includes emotional responses during gameplay, such as exploration, combat, or decision-making moments. For example, ChatNPC can react differently if the player is excited after achieving a milestone or frustrated during a difficult challenge. It attentively captures verbal expressions, such as the player's subtle, unconscious comments and gestures, using a speech-to-text mechanism, and leverages the in-game context to generate appropriate responses.

The model integrates sequential data that benefit from audio streams (player spoken lines), player information (health, level), emotions expressed from the spoken lines, in-game context information, and an inner monologue. Specifically, the cues in the audio streams are extracted and interpreted by an LLM agent. The LLM agent, known as *SenGent* (Section 4) is an integrated embedding layer in the model architecture to analyze and interpret player emotional shift and understand the situational aspects of the gameplay by tracking previous context and current context information to initiate on-the-fly prompt adjustments. The combined data are concatenated with the prompt tokens, which triggers the LLM (a *chat planning agent*) inference to generate responses. As shown in Figure 1, the *SenGent* brings believability to the response, making it more naturalistic and human-like, with appropriate conversational cues. We borrow from the idea of MemGPT (Packer et al., 2023) to keep track of events occurring in the playthrough and the interaction with the player in the overall game. We engage the recent progress in LLMs tool calling (Kim et al., 2023) capabilities to extract vital information from memory during problem distillation (Yang et al., 2024a), and recognize potential constraints for accurate reasoning (Yang et al., 2024a; Wei et al., 2022) to tackle the complexity in NPC conversation scenarios. Finally, to maintain a thriving, diverse conversation, we prompt the chat planning agent with a character role, a *persona* that includes personality (Jiang et al., 2024), knowledge, and communication style. Our contributions are summarized as follows:

1. We propose a game companion (*ChatNPC*) capable of tracking and interpreting players' emotional shifts in real-time, adapting its responses dynamically within broader gameplay interactions.

2. The approach uses a novel sentinel mechanism to foster emotional connection and manage contextual aspects of the game for a seamless narrative flow.

3. We design a *memory agent* to process information between the *in-context* and the *out-of-context* window, keep track of events in the playthrough for reasoning instantiation and rescaling (pre-think and improve mechanism) of thought.

4. We conduct extensive experiments on newly developed game datasets for in-game context NPC dialogue.

Game Conscious™ AI[2] allows a life-like conversation with LLM-powered NPCs to advance game developers' stories and objectives. The user application in "DEAD MEAT," an interrogation game in which players can ask the suspect anything in their own words, relies on the game's background information and user input to formulate its thought, making it susceptible to hallucination. It is geared towards a Q&A approach with voice inputs as a direct command for in-game actions. It also shows the AI's inner monologue (which they described as a glimpse of the suspect's mind, "*Mind Reader*") as part of the outputs. ChatNPC, on the other hand, does not use the player's inputs as a direct command for character movement or in-game actions. Instead, it analyzes the user's spoken lines for a narrative conversation during gameplay. It also leverages the inner monologue or inner thinking to guide its responses rather than serving as part of the outputs. ChatNPC fosters a sense of belonging and creates emotional engagement as the companion becomes a responsive and empathetic presence, reinforcing the player's immersion in video games.

[2]https://www.meaningmachine.games/game-conscious-ai

2 Related Work

LLMs are inherently suited for natural language dialogue; therefore, they are typically presented as conversational agents, displaying remarkable skills in memory (Packer et al., 2023; Sumers et al., 2023), tool usage (Yang et al., 2024b; Schick et al., 2024; Ruan et al., 2023; Qin et al., 2023) and planning (Yang et al., 2024a; Wei et al., 2022), often leading researchers to give them reasoning and creativity qualities (Gallotta et al., 2024; Wei et al., 2022; Zhao et al., 2024; Zhang et al., 2024). The growing capabilities have motivated recent efforts to incorporate them as game agents (Oliver and Mateas, 2021) with believable proxies of human behavior to empower interactivity in immersive environments.

LLMs in Games LLMs operate within games as a player (Toshniwal et al., 2022; Li et al., 2022; , FAIR; Yao et al., 2020), NPCs (Gallotta et al., 2024; Weir et al., 2022; Li et al., 2022), an assistant providing hints (Akata et al., 2023; Xu et al., 2023), a game master (Zhu et al., 2023) controlling the flow of the game and acting as a commentator (Ranella and Eger, 2023) of an ongoing play session. Park et al. (2023) introduced generative agents that simulate believable human behavior, such as daily activities (waking up, cooking breakfast, and going to work). The authors used LLMs to populate an interactive virtual Smallville sandbox environment with 25 generative agents by storing an entire history of the agent's experiences, synthesizing those memories over time into higher-level representations, and dynamically retrieving them for behavior planning. Other works include using vision LMs (VLMs) to analyze video frames and predict the next steps in the Mario video game (Lin et al., 2023) and to evaluate their effectiveness in the game StarCraft II (Ma et al., 2023). To understand how LLMs behave in interactive social settings, Akata et al. (2023) proposed using behavioral game theory to study agents' cooperation and coordination behavior by engaging different language models as agents repeatedly play games with each other in human-like scenarios. What is most common among these techniques is the ability for users to interact with these agents in a natural language dialogue.

LLMs and NPC Dialogue. LLM-powered framework for quests and dialogue generation that places the player at the core of the generative process has been explored with context-awareness (Ashby et al., 2023; Müller-Brockhausen et al., 2023), and personality modeling (Müller-Brockhausen et al., 2023; Latouche et al., 2023) to create fluent, unique, and accompanying dialogue. ChatGPT has recently been used to generate game dialogue by allowing them to take control of NPCs and interact dynamically with players (Zhou et al., 2023). In addition to generating NPC dialogue, many research works integrated LLMs into game mechanics using quest datasets (Ashby et al., 2023; Weir et al., 2022; Värtinen et al., 2022; van Stegeren and Myśliwiec, 2021; Gao and Emami, 2023). A Minecraft player interacting with a Codex-powered NPC in question-answering and task-completion scenarios demonstrates that conversational prompts can power a conversational agent to generate natural language (Volum et al., 2022). These approaches greatly benefit both the gaming industry and its global community as they unleash the immersiveness and enjoyment of the user (Akoury et al., 2023). Regardless, the role of NPCs in games extends far beyond serving as quest-givers, question-answering, and task-completion. In addition, modern gaming goes beyond just winning, and voice commands; players seek a sense of immersion and connection with the virtual world.

Player Emotions in Games Many studies explore the emotional responses evoked during game-play and interactions with NPCs (Marincioni et al., 2024). A study by Mozikov et al. (2024) focused on LLMs' decision-making and their alignment with human behavior in emotional states in various strategic games. They revealed that emotional prompting, particularly with certain emotions (such as anger), can disrupt some LLMs' "superhuman" alignment, similar to human emotional responses. Language models have been used to extract emotion scores from the emotional responses of players to the game by capturing the interaction with the NPC through user input to understand the emotional dynamics within the gaming environments (Marincioni et al., 2024). These innovative approaches predominantly utilize LLMs for in-game decision-making. Current research has yet to fully explore the potential of an LLM-powered game companion that apprehends players' emotional shifts during gameplay.

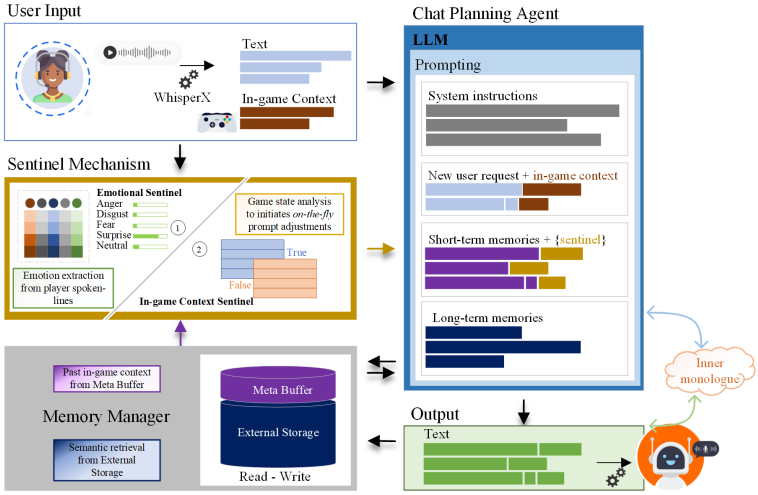

Figure 2: The overall architecture of ChatNPC constitutes three main modules. The sentinel mechanism with two different embedding types (an *in-game context* and *emotion sentinel* embedding), the memory manager to process information between the *in-context* and the *out-of-context* window and the *chat planning agent* with to format the output request.

3 Method

This section details the ChatNPC pipeline process and illustrates the core modules, as shown in Figure 2. To accurately convert players' spoken lines into textual representations, we leverage the advanced WhisperX[3] model for speech-to-text transcription. It enables an efficient, fast, and automatic speech recognition system with time-accurate and word-level timestamps (Bain et al., 2023).

3.1 Sentinel Mechanism

The sentinel mechanism seamlessly integrated embedding layer in the model architecture that does not require significant computational resources, constituting two different embedding types: an *in-game context* and *emotion* sentinel embedding.

In-game Context Sentinel. Focuses on understanding the situational aspects of the gameplay by considering the storyline and environmental factors of the game to ensure cohesion. It tracks the previous and current context information to effectively initiate *on-the-fly* prompt adjustments in the model responses based on a "true/false" validation. To achieve this, we use an embedding model $\phi(\cdot)$ to capture the difference between the current and past context in the game and finally compute the cosine similarity. For each input i, we compare the current in-game context C_i by computing the embedding similarity between the previous in-game context C_{i-1} and C_i as:

$$\mathbf{pr}_{adj} = \mathbf{1}_{[\mathcal{T},1]}\Big(\mathrm{Sim}\big(\phi(C_i), \phi(C_{i-1})\big)\Big) \quad (1)$$

where $\mathcal{T}$ is the threshold ($0.8 \sim 0.9$ is recommended) to determine whether the in-game context matches for on-the-fly prompt adjustment $\mathbf{pr}_{adj}$, and $\mathbf{1}_{[\mathcal{T},1]}$ denotes the characteristic function of the interval $[\mathcal{T}, 1]$. The injected prompt provides clear instructions within the system prompt to guide the LLM in reacting when new user input is received. For instance, if the current in-game context aligns with the previous context, the system instruction is adjusted to be consistent with the storyline and previous agent-player interaction. With no alignment, it is prompted to maintain a seamless narrative flow.

Emotion Sentinel. Focuses on analyzing and interpreting users' emotions through spoken lines and the subtle cues detected. It reflects an understanding of the players' thoughts and feelings during gameplay. Given the transcribed text, we instantiate an embedding layer that extracts the emotional signals from the player's spoken lines as *metadata* and interprets it to potentially adjust the LLM's reasoning. The embedding is distilled from a zero-shot classification model pre-trained on natural language inference (NLI) and trained on the GoEmotions dataset (Demszky et al., 2020). GoEmotions excels in its comprehensive categorization, identifying expressions in 28 distinct emotional categories (27 emotions plus a neutral category) (Wang et al., 2024; Singh et al., 2021), making it a perfect ap-

[3]https://github.com/m-bain/whisperX

proach to detect subtle emotional changes during gameplay. Overall, the spoken line $\mathcal{S}$ is converted into an embedding vector $V_{\mathcal{S}}$ for a semantic representation using the same embedding space as the emotion vectors $E_{\mathcal{J}}$ as:

$$\text{meta}_{emo} = \text{argmax}_{\mathcal{J}}\Big(\text{Sim}(V_{\mathcal{S}}, E_{\mathcal{J}})\Big), \quad (2)$$

where $\mathcal{J} \in \{j_1, j_2, ..., j_n\}$ is the vector for emotion-labeled in the GoEmotions datasets. The emotion vector with the highest similarity score determines the most likely emotion expressed.

3.2 Memory Manager

Inspired by MemGPT (Packer et al., 2023), the *memory manager* module processes information between the *in-context* or immediate context and the *out-of-context* window to keep track of events in the playthrough.

In-context window Also known as the **primary prompt tokens**, consist of the system instructions, and a *meta-buffer* as a means of storing rolling chat history, which is always available to the in-context window. The system instruction is static (read-only), a thoroughly drafted information for reasoning instantiation. Most specifically, it is a game-specific system prompt that includes role specification, context definition (*game-template*), task description, input details, output requirements, constraints and rules, and examples to enhance clarity. The *meta-buffer* stores conversation between the agent and the player and uses the FIFO (First In, First Out) operation to dynamically append and remove interaction. With a token count mechanism, we keep track of the current number of tokens in the in-context window. When the system receives a new input, the incoming messages are appended to the *meta-buffer*. It then concatenates to the primary prompt tokens to trigger the LLM inference for the response. The oldest interactions in the queue are then moved to *archival storage*, (the *out-of-context* window) when the primary prompt exceeds a range of defined limits.

Out-of-context window. We utilize Chroma database (DB) to store information beyond the primary context window, enabling efficient retrieval and continuity over a long horizon of conversational interactions. We use paging to store and update memories efficiently by setting an arbitrary threshold to control the maximum number of long-term memories stored (Packer et al., 2023).

Retrieval. Finally, we implement a semantic matching (Rao et al., 2019) to identify the core meaning behind the text instead of the exact word matches to retrieve out-of-context data. We also retrieve relevant information as a pre-think mechanism to improve ChatNPC-player dialogue. The pre-think process uses historical data to determine whether the task has already been completed or previously initiated. This approach proves especially practical when the player has previously interacted with or completed certain levels within the game (see Section 4.4).

3.3 Chat Planning Agent

The *chat planning* agent is responsible for formatting the output. When a new user query is fired, information from memory: *meta-buffer* and *archival storage* (if necessary), and *sentinel agent* will be passed to the problem distiller to extract critical state and contextual information along with relevant constraints for reasoning instantiation. The problem distiller, similar to the buffer of thoughts (BoT) (Yang et al., 2024a), focuses on extracting key elements from the input task.

Game-template. A *game-template* tailored for the companion is designed as an information framework that enriches the agent's ability to support and enhance the player's experience in the game world. The template includes details about the game's storyline, objectives, and 'tutorial and hints', allowing the agent to remain aware and provide timely, contextually appropriate feedback to the player.

Agent persona. To maintain a successful diverse conversation, we prompt the *chat planning* agent with a character role, a *persona* that includes personality, knowledge, and communication style. We use the character's iconic style of expression/talking in the game world as pre-conversation. Each time there is a dialogue request, arbitrary lines from this pre-conversation are concatenated with the prompt tokens while generating the response to ensure a faithful response from the companion.

Additionally, ChatNPC relies on an inner monologue, or inner thinking, to create a more conscious AI to guide its reactions by trying to reason the best way to give the response. This inner monologue is accessible to the AI but is not displayed as output or as part of the responses. By combining these functionalities, ChatNPC performs multi-step operations and computations, enabling them to tackle complex conversation scenarios and provide more accurate and detailed responses. The detailed

	Utterances	Context	Responses
Token counts	148668	156616	497319
Unique words	7815	1758	8248

Train	Validation	Testing
10,000	3,885	512

Table 1: Statistics of the iGCD Dataset .

Model (Open-source)	Context Window Token	*Messages
Llama 2	4k	50
Llama 3	8k	50
Mythalion 13B	4K	50
Roleplay Llama-3-8B	8K	50

Table 2: Llama 2&3 open-source models, two different roleplay models based on Llama 2&3, and the primary context window length. The default maximum token of *messages is set to an average of 50 tokens (approximately 250 characters).

prompt for ChatNPC is included in Appendix B.3.

4 Experiments

4.1 Experimental Settings

Dataset. In this work, we introduce a new **i**n-**g**ame **c**ontext **d**ialogue (iGCD) dataset to evaluate the effectiveness of the game companion in both conversation and memory retention over longer horizons of conversation. The dataset consists of 14,397 pairs of utterances, in-game context and the corresponding responses from 4 games, where 10,000 is used for reference training data, 4,885 as a validation set, and 512 for testing. Overall statistics of the collected dataset are given in Table 1. More details on video game data scripting, filtering, and evaluation process, including metrics, are thoroughly discussed in the Appendix A.1.

Base LLM. We use open-source models, including the Llama baseline models (see Table 2), and two role-play models: Roleplay Llama-3-8B[4], and Mythalion 13B[5] model based on Llama-2-13B. For supervised fine-tuning, we use Llama-2-70B as the backbone of our ChatNPC, including the main experiment and ablation study (in Subsection 4.4).

4.2 Implementation Details

For the *SenGent*, we use a low computation mechanism as mentioned. The *in-game context* embedding uses a helper function to generate an embedding for a given text using a pre-trained Sentence Transformer model[6] and compute the cosine similarity between the current and previous context. The *emotion* sentinel embedding uses a mechanism distilled from a zero-shot classification model[7].

LLM Instruction Fine-tuning Given that our data set does not contain a desired output, we initially generate responses by instruction fine-tuning Llama-2 70B using the Open-Platypus (Lee et al., 2023) dataset to produce contextually appropriate results. These results underwent a rigorous evaluation process, including automated AI-based assessment (LLM-as-a-Judge) (Dalal et al., 2024) and human-in-the-loop evaluation (Elangovan et al., 2024) (see Appendix A).

Hyperparameters. In the experiment, we set the maximum number of optimization steps to 15k, a weight decay of 0.01 and the learning rate to 2e-4. For stability of results, we use beam-search (*=3*) decoding strategy combined with multinomial sampling to generate the output for the overall highest probability given the entire sequence for inference. The base temperature parameter is set to 0.4, with a random variation between ±0.05 and the minimum and maximum sequence lengths to 20 and 50, respectively. We report the detailed hyperparameter settings for fine-tuning and decoding in Appendix A.2.

After curating a few thousand training data, we instantiate similar parameters for supervised fine-tuning on iGCD with the desired output using Low-Rank Adaptation (LoRA) (Hu et al., 2021) to show how the underlined dataset helps improve ChatNPC performance.

Inference and Evaluation. For inference with the base LLMs and the roleplay models, we employ a few-shot (Brown, 2020) prompting approach to facilitate in-context learning. We benchmark ChatNPC against the baselines (see Table 3) with human-in-the-loop evaluation as the "gold response." We use metrics such *ROUGE-L* and *ROUGE-WE* scores (Appendix B.1) to reference the highest achievable score when comparing the generated results with the gold responses.

4.3 Results and Analysis

As displayed in Table 3, ChatNPC benefits from the supervised fine-tuning for its overall performance compared to the baselines with Few-Shot prompt-

[4]https://huggingface.co/vicgalle/Roleplay-Llama-3-8B
[5]https://huggingface.co/PygmalionAI/mythalion-13b
[6]https://huggingface.co/nreimers/MiniLM-L6-H384-uncased
[7]joeddav/distilbert-base-uncased-go-emotions-student

	Model	ROUGE-L Precision	ROUGE-L Recall	ROUGE-L F1	ROUGE-WE Precision	ROUGE-WE Recall	ROUGE-WE F1
Base	Llama 2 7B	0.327	0.25	0.275	0.465	0.417	0.443
	Llama 2 13B	0.383	0.272	0.318	0.48	0.436	0.461
	Llama 2 70B	0.516	0.404	0.453	0.599	0.55	0.573
	Llama 3 8B	0.391	0.279	0.326	0.489	0.441	0.482
	Llama 3 70B	0.523	0.419	0.268	0.621	0.571	0.595
Roleplay	Llama-3-8B	0.553	0.481	0.51	0.682	0.607	0.648
	Mythalion 13B	0.51	0.427	0.449	0.629	0.58	0.602
Fine-tuned	Llama 2 70B	0.783	0.64	0.687	0.855	0.805	0.83
	Llama 3 70B	**0.807**	**0.661**	**0.715**	**0.883**	**0.836**	**0.862**

Table 3: The overall model performance is based on Llama base, roleplay, and fine-tuned models. ChatNPC response is scored against the human-in-th-loop (gold) responses.

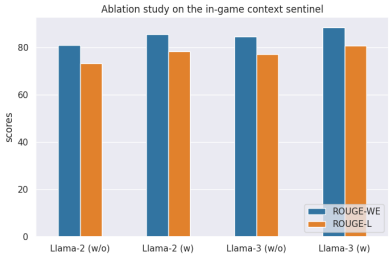

Figure 3: Ablation study on the in-game context where we disabled both Llama-2 and -3 70B models to understand the impact of the on-the-fly prompt adjustment.

ing and role-play models. Specifically, ChatNPC surpasses the base and role-play models by 26.5% and 20.1% on ROUGE-WE scores, respectively.

There has been a massive increase in dialogue diversity. Hence, a better interpretation of how ChatNPC effectively addresses the complexity of NPC conversation scenarios. This improvement is attributed to contextual information from the training dataset and the modules instantiated. The base LLMs demonstrate poor performance mainly because they tend to ignore (not always) the conversational level of interaction. Even with Few-Shot prompting and role-play, models cannot learn the contextual information in task-specific cases. Role-play models, though, are promising, with respective precision, recall, and F1 scores of 0.682, 0.607, and 0.648 compared to the base LLMs.

The increased ROUGE scores on the supervised fine-tuned ChatNPC indicate that curating a high-quality dataset is recommended instead of Few-Shot learning in task-specific or performance-sensitive applications (Hu et al., 2021). From the main results, Few-Shot learning did not achieve the robustness required in complex NPC conversation scenarios, although it offers the advantage of rapid adaptability with minimal data. Also, from table 3, we observe no significant differences in the base Llama-2 and -3 70B. We attribute this to the evaluation metrics engaged as the ROUGE scores capture matches based on meaning relatively more than accurate words. On the other hand, the curated dataset plays a significant role by ensuring the model is fine-tuned on highly relevant, diverse, and in-game context-specific instances for better generalization. Interestingly, responses from ChatNPC are more naturalistic and human-like, with appropriate conversational markers, pauses, and transitions, facilitating profound emotional engagement.

Generally, while the *SenGent* maintains continuity and helps avoid breaks in the narrative, the memory agent retrieves relevant memories for self-improvement, especially for previously visited events or played scenarios in the game, a common phenomenon in humans. We perform an ablation study to better understand the benefit of the *SenGent* in ChatNPC and how it uses relevant memories for self-improvement.

4.4 Ablation Study

Impact of In-game Context Sentinel. The on-the-fly prompt adjustment enhances the sense of realism and maintains narrative continuity and immersion within the game. Especially with such validation, ChatNPC ensures cohesiveness, building on the conversation and themes of the previous state when the in-game context aligns. In a scenario with a dynamic in-game context change (often), the prompt adjustment helps adapt to the new situation and avoid contradictions or breaks in the narrative that might disrupt the player's immersion.

Figure 3 shows the impact of the in-game context sentinel, as we can see a decline in ROUGE scores. We compare several ChatNPC scenarios with and without the context sentinel. For such an

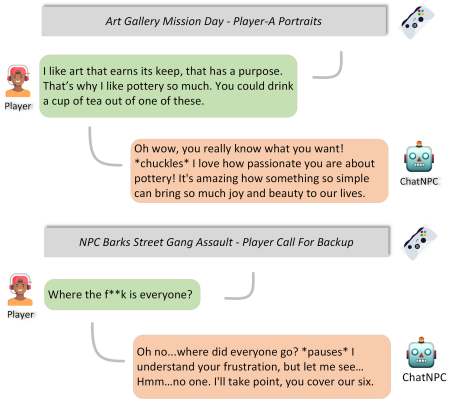

Figure 4: Snippets of the ChatNPC responses when noticed that the user is distressed, and moments of sensation.

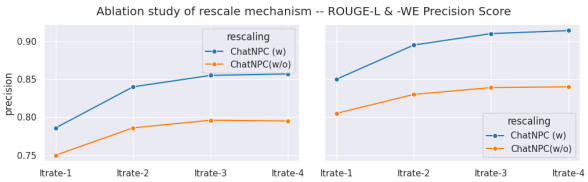

Figure 5: We conduct an ablation study on the rescale mechanism (pre-think and improve) by iterating the model (Llama-2 70B as the backbone) multiple times with 10% of the test set to signify previously interacted or completed levels within the game. ROUGE-L (**left**) and ROUGE-WE (**right**).

experiment, we sampled from the gold response with the same context. Although ChatNPC without a context sentinel provides contextually appropriate responses, there is a slight deviation in maintaining narrative continuity. Also, when the in-game context sentinel is disabled, both Llama-2 and -3 70B models exhibit a noticeable decline in performance. Ultimately, the responses are more verbose with the sentinel agent than the human-in-the-loop (gold) response.

Impact of Emotional Sentinel. The player-agent interaction in Figure 4 shows that the extracted emotional metadata and interpretability allow the model to adjust its reactions with acuity to the player's emotional sift for better alignment, reflecting an understanding of the user's thoughts and feelings over time. This emotional modification helps create a more personalized, compassionate background by tailoring responses to moments of frustration, excitement, or contemplation. As illustrated, the interaction demonstrates words of encouragement from ChatNPC when the player expresses frustration and offers celebratory reactions to moments of sensation. This module ultimately helps establish a more human-like connection with the player, fostering a responsive environment and improving interaction engagement.

Impact of Memory Agent ChatNPC not only leverages memory for relevant information but also rescales (pre-think and improve mechanism) it to perform well by intuition without conscious reasoning (Li and Qiu, 2023). To understand the potential improvement of the ChatNPC conversation, we sampled 10% of the test set where we iterated ChatNPC (with/without rescaling) multiple times to illustrate previously interacted or completed levels within the game. At each iteration, we prompt the model to use memory for pre-thinking and improvement. As shown in Figure 5, ChatNPC (with) consistently improves the ROUGE scores in every iteration. Ultimately, the pre-think mechanisms of the memory agent, the support of narrative continuity from the sentinel agent, and the character persona help in adaptive interaction during gameplay creating a dynamic and believable agent reaction that aligns with the player's actions and preferences.

5 Conclusion

In this work, we propose *ChatNPC*, a game companion that observes and recognizes players' emotional shifts and adjusts its responses accordingly in complex NPC conversation scenarios. Specifically, ChatNPC comprises three modules: a novel game sentinel mechanism to effectively manage the emotional and contextual aspects of NPC-player dialogue, a memory agent to keep track of events in the playthrough for reasoning instantiation, and a chat planning agent that benefits from game-template and character persona for formatting the output request. We introduce a newly curated iGCD (in-game context dialogue) dataset for the experimental analysis. The results indicate that ChatNPC sequentially captures players' emotional shifts over time; responses are more naturalistic and human-like with appropriate conversational cues, pauses, and sighs; and utterances remain faithful to the dynamics of in-game context and supporting narrative continuity.

Acknowledgments

References

Elif Akata, Lion Schulz, Julian Coda-Forno, Seong Joon Oh, Matthias Bethge, and Eric Schulz. 2023. Playing repeated games with large language models. *arXiv preprint arXiv:2305.16867*.

Nader Akoury, Qian Yang, and Mohit Iyyer. 2023. A framework for exploring player perceptions of llm-generated dialogue in commercial video games. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 2295–2311.

Trevor Ashby, Braden K Webb, Gregory Knapp, Jackson Searle, and Nancy Fulda. 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*, pages 1–20.

Max Bain, Jaesung Huh, Tengda Han, and Andrew Zisserman. 2023. Whisperx: Time-accurate speech transcription of long-form audio. *INTERSPEECH 2023*.

Tom B Brown. 2020. Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*.

Maximilian Croissant, Guy Schofield, and Cade McCall. 2023. Emotion design for video games: A framework for affective interactivity. *ACM Games: Research and Practice*, 1(3):1–24.

Dhairya Dalal, Marco Valentino, André Freitas, and Paul Buitelaar. 2024. Inference to the best explanation in large language models. *arXiv preprint arXiv:2402.10767*.

Dorottya Demszky, Dana Movshovitz-Attias, Jeongwoo Ko, Alan Cowen, Gaurav Nemade, and Sujith Ravi. 2020. Goemotions: A dataset of fine-grained emotions. *arXiv preprint arXiv:2005.00547*.

Aparna Elangovan, Ling Liu, Lei Xu, Sravan Bodapati, and Dan Roth. 2024. Considers-the-human evaluation framework: Rethinking human evaluation for generative large language models. *arXiv preprint arXiv:2405.18638*.

Meta Fundamental AI Research Diplomacy Team (FAIR)†, Anton Bakhtin, Noam Brown, Emily Dinan, Gabriele Farina, Colin Flaherty, Daniel Fried, Andrew Goff, Jonathan Gray, Hengyuan Hu, et al. 2022. Human-level play in the game of diplomacy by combining language models with strategic reasoning. *Science*, 378(6624):1067–1074.

Roberto Gallotta, Graham Todd, Marvin Zammit, Sam Earle, Antonios Liapis, Julian Togelius, and Georgios N Yannakakis. 2024. Large language models and games: A survey and roadmap. *arXiv preprint arXiv:2402.18659*.

Qi Chen Gao and Ali Emami. 2023. The turing quest: Can transformers make good npcs? In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 4: Student Research Workshop)*, pages 93–103.

Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2021. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*.

Guangyuan Jiang, Manjie Xu, Song-Chun Zhu, Wenjuan Han, Chi Zhang, and Yixin Zhu. 2024. Evaluating and inducing personality in pre-trained language models. *Advances in Neural Information Processing Systems*, 36.

Sehoon Kim, Suhong Moon, Ryan Tabrizi, Nicholas Lee, Michael W Mahoney, Kurt Keutzer, and Amir Gholami. 2023. An llm compiler for parallel function calling. *arXiv preprint arXiv:2312.04511*.

Gaetan Lopez Latouche, Laurence Marcotte, and Ben Swanson. 2023. Generating video game scripts with style. In *Proceedings of the 5th Workshop on NLP for Conversational AI (NLP4ConvAI 2023)*, pages 129–139.

Ariel N Lee, Cole J Hunter, and Nataniel Ruiz. 2023. Platypus: Quick, cheap, and powerful refinement of llms. *arXiv preprint arXiv:2308.07317*.

Kenneth Li, Aspen K Hopkins, David Bau, Fernanda Viégas, Hanspeter Pfister, and Martin Wattenberg. 2022. Emergent world representations: Exploring a sequence model trained on a synthetic task. *arXiv preprint arXiv:2210.13382*.

Xiaonan Li and Xipeng Qiu. 2023. Mot: Pre-thinking and recalling enable chatgpt to self-improve with memory-of-thoughts. *CoRR*, abs/2305.05181.

Austen Liao, Nicholas Tomlin, and Dan Klein. 2024. Efficacy of language model self-play in non-zero-sum games. *arXiv preprint arXiv:2406.18872*.

Kevin Lin, Faisal Ahmed, Linjie Li, Chung-Ching Lin, Ehsan Azarnasab, Zhengyuan Yang, Jianfeng Wang, Lin Liang, Zicheng Liu, Yumao Lu, et al. 2023. Mm-vid: Advancing video understanding with gpt-4v (ision). *arXiv preprint arXiv:2310.19773*.

Weiyu Ma, Qirui Mi, Yongcheng Zeng, Xue Yan, Yuqiao Wu, Runji Lin, Haifeng Zhang, and Jun Wang. 2023. Large language models play starcraft ii: Benchmarks and a chain of summarization approach. *arXiv preprint arXiv:2312.11865*.

Alessandro Marincioni, Myriana Miltiadous, Katerina Zacharia, Rick Heemskerk, Georgios Doukeris, Mike Preuss, and Giulio Barbero. 2024. The effect of llm-based npc emotional states on player emotions: An analysis of interactive game play. In *2024 IEEE Conference on Games (CoG)*, pages 1–6. IEEE.

Mikhail Mozikov, Nikita Severin, Valeria Bodishtianu, Maria Glushanina, Mikhail Baklashkin, Andrey V Savchenko, and Ilya Makarov. 2024. The good, the bad, and the hulk-like gpt: Analyzing emotional decisions of large language models in cooperation and bargaining games. *arXiv preprint arXiv:2406.03299*.

Matthias Müller-Brockhausen, Giulio Barbero, and Mike Preuss. 2023. Chatter generation through language models. In *2023 IEEE Conference on Games (CoG)*, pages 1–6. IEEE.

Elisabeth Oliver and Michael Mateas. 2021. Crosston tavern: Modulating autonomous characters behaviour through player-npc conversation. *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, 17(1):179–186.

Charles Packer, Vivian Fang, Shishir G Patil, Kevin Lin, Sarah Wooders, and Joseph E Gonzalez. 2023. Memgpt: Towards llms as operating systems. *arXiv preprint arXiv:2310.08560*.

Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. 2023. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th annual acm symposium on user interface software and technology*, pages 1–22.

Yujia Qin, Shihao Liang, Yining Ye, Kunlun Zhu, Lan Yan, Yaxi Lu, Yankai Lin, Xin Cong, Xiangru Tang, Bill Qian, et al. 2023. Toolllm: Facilitating large language models to master 16000+ real-world apis. *arXiv preprint arXiv:2307.16789*.

Noah Ranella and Markus Eger. 2023. Towards automated video game commentary using generative ai. In *Proceedings of the AIIDE workshop on Experimental AI in Game*.

Jinfeng Rao, Linqing Liu, Yi Tay, Wei Yang, Peng Shi, and Jimmy Lin. 2019. Bridging the gap between relevance matching and semantic matching for short text similarity modeling. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5370–5381.

Jingqing Ruan, Yihong Chen, Bin Zhang, Zhiwei Xu, Tianpeng Bao, Hangyu Mao, Ziyue Li, Xingyu Zeng, Rui Zhao, et al. 2023. Tptu: Task planning and tool usage of large language model-based ai agents. In *NeurIPS 2023 Foundation Models for Decision Making Workshop*.

Timo Schick, Jane Dwivedi-Yu, Roberto Dessì, Roberta Raileanu, Maria Lomeli, Eric Hambro, Luke Zettlemoyer, Nicola Cancedda, and Thomas Scialom. 2024. Toolformer: Language models can teach themselves to use tools. *Advances in Neural Information Processing Systems*, 36.

Gargi Singh, Dhanajit Brahma, Piyush Rai, and Ashutosh Modi. 2021. Fine-grained emotion prediction by modeling emotion definitions. In *2021 9th International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 1–8. IEEE.

Theodore R Sumers, Shunyu Yao, Karthik Narasimhan, and Thomas L Griffiths. 2023. Cognitive architectures for language agents. *arXiv preprint arXiv:2309.02427*.

Yuqian Sun, Zhouyi Li, Ke Fang, Chang Hee Lee, and Ali Asadipour. 2023. Language as reality: a co-creative storytelling game experience in 1001 nights using generative ai. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume 19, pages 425–434.

Penny Sweetser. 2024. Large language models and video games: A preliminary scoping review. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces*, pages 1–8.

Shubham Toshniwal, Sam Wiseman, Karen Livescu, and Kevin Gimpel. 2022. Chess as a testbed for language model state tracking. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 11385–11393.

Muhtar Çağkan Uludağlı and Kaya Oğuz. 2023. Non-player character decision-making in computer games. *Artificial Intelligence Review*, 56(12):14159–14191.

Judith van Stegeren and Jakub Myśliwiec. 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*, pages 1–8.

Susanna Värtinen, Perttu Hämäläinen, and Christian Guckelsberger. 2022. Generating role-playing game quests with gpt language models. *IEEE transactions on games*, 16(1):127–139.

Ryan Volum, Sudha Rao, Michael Xu, Gabriel DesGarennes, Chris Brockett, Benjamin Van Durme, Olivia Deng, Akanksha Malhotra, and William B Dolan. 2022. Craft an iron sword: Dynamically generating interactive game characters by prompting large language models tuned on code. In *Proceedings of the 3rd Wordplay: When Language Meets Games Workshop (Wordplay 2022)*, pages 25–43.

Kaipeng Wang, Zhi Jing, Yongye Su, and Yikun Han. 2024. Large language models on fine-grained emotion detection dataset with data augmentation and transfer learning. *arXiv preprint arXiv:2403.06108*.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837.

Nathaniel Weir, Ryan Thomas, Randolph D'Amore, Kellie Hill, Benjamin Van Durme, and Harsh Jhamtani. 2022. Ontologically faithful generation of non-player character dialogues. *arXiv preprint arXiv:2212.10618*.

Yuzhuang Xu, Shuo Wang, Peng Li, Fuwen Luo, Xiaolong Wang, Weidong Liu, and Yang Liu. 2023. Exploring large language models for communication games: An empirical study on werewolf. *arXiv preprint arXiv:2309.04658*.

Ling Yang, Zhaochen Yu, Tianjun Zhang, Shiyi Cao, Minkai Xu, Wentao Zhang, Joseph E Gonzalez, and Bin Cui. 2024a. Buffer of thoughts: Thought-augmented reasoning with large language models. *arXiv preprint arXiv:2406.04271*.

Rui Yang, Lin Song, Yanwei Li, Sijie Zhao, Yixiao Ge, Xiu Li, and Ying Shan. 2024b. Gpt4tools: Teaching large language model to use tools via self-instruction. *Advances in Neural Information Processing Systems*, 36.

Shunyu Yao, Rohan Rao, Matthew Hausknecht, and Karthik Narasimhan. 2020. Keep calm and explore: Language models for action generation in text-based games. *arXiv preprint arXiv:2010.02903*.

An Zhang, Yuxin Chen, Leheng Sheng, Xiang Wang, and Tat-Seng Chua. 2024. On generative agents in recommendation. In *Proceedings of the 47th international ACM SIGIR conference on research and development in Information Retrieval*, pages 1807–1817.

Andrew Zhao, Daniel Huang, Quentin Xu, Matthieu Lin, Yong-Jin Liu, and Gao Huang. 2024. Expel: Llm agents are experiential learners. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 19632–19642.

Wei Zhou, Xiangyu Peng, and Mark Riedl. 2023. Dialogue shaping: Empowering agents through npc interaction. *arXiv preprint arXiv:2307.15833*.

Andrew Zhu, Lara Martin, Andrew Head, and Chris Callison-Burch. 2023. Calypso: Llms as dungeon master's assistants. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume 19, pages 380–390.

A Appendix A

A.1 Dataset

In-game Context Dialogue (iGCD) dataset is a small-scale dataset that consists of scripted speech (user-spoken lines) from the games and the in-game mechanics to serve as input and context to LLMs. We initially generate output through instruction fine-tuning using the Open-Platypus (Lee et al., 2023) dataset[8] to produce contextually appropriate results. These preliminary responses underwent a rigorous dual-layer evaluation process involving automated AI-based assessment and human-in-the-loop evaluation to ensure quality, relevance, and alignment with the intended objectives. The evaluation allowed for a comprehensive analysis of the outputs, addressing potential biases and enhancing their accuracy and usability. Hence, the dataset serves as a foundation for training the model to understand the game's context and players' spoken interactions. It includes artistic dialogues with visual cues, encouraging ChatNPC to provide responses that prioritize tone, style, and emotional depth. Figure 6 shows the distribution of iGCD dataset based on the number of samples and token counts.

AI Evaluation. Given the user's spoken lines, in-game context, and response from the LLM, we prompt Llama 2 7B to provide a qualitative evaluation by assigning an estimated score from 1 to 10 and offering a justification for the assigned rating, which further captures essential feedback on the language models' performance. The evaluation facilitates an iterative process for improving the quality of generated outputs through comparative evaluations and reasoning for human-in-the-loop evaluation.

Human Expert Evaluation. Automated evaluation metrics often struggle to capture rich language with nuance, context, and subjectivity, leading to gaps in assessing the quality of the AI's output. To evaluate whether the AI's response is contextually appropriate and relevant in a broader discourse, we employ human-in-the-loop evaluation to provide valuable data that can be used to fine-tune AI models. We adjust training data, model parameters, and algorithms to improve performance by understanding where AI responses fall short. This iterative feedback loop between human evaluators and researchers helps to refine the models more effectively. Human evaluation focuses mainly on naturalness, relevance, coherence, and expressiveness, as shown in Table 4. The tool for human evaluation/survey is shown in Figure 7

A.2 Hyper-parameter settings

Table 5 illustrates the hyper-parameter configurations utilized for instruction fine-tuning and inference processes.

B Appendix B

B.1 Evaluation metrics

ROUGE-L Recall-Oriented Understudy for Gisting Evaluation measures the longest common subsequence (LCS) between the generated and gold response. It captures partial matches to evaluate how

[8] https://huggingface.co/datasets/garage-bAInd/Open-Platypus

Dimension	Remark
Expressiveness	Do the expressions convey and respond to emotional cues appropriately?
Coherence	Does the dialogue feel more coherent and meaningful?
Naturalness	Is the response human-like with appropriate use of conversational markers?
Relevance	Does the interaction feel more personal and relevant?

Table 4: Dimension and remarks of human-in-the-loop evaluation.

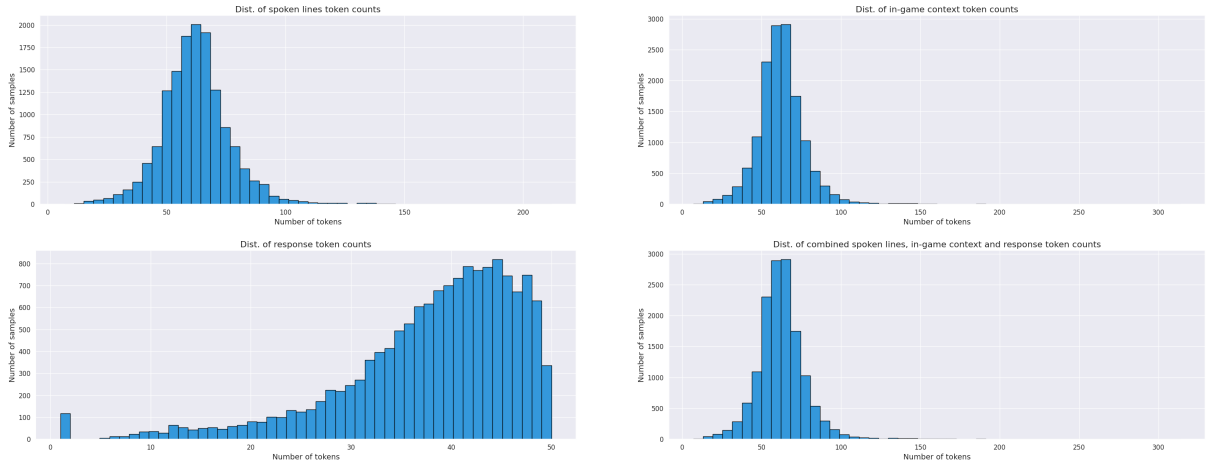

Figure 6: iGCD distribution based on the sample number and token counts. Spoken-line (**Top left**), in-game context (**Top right**), responses (**Bottom left**) and combined token counts (**Bottom right**)

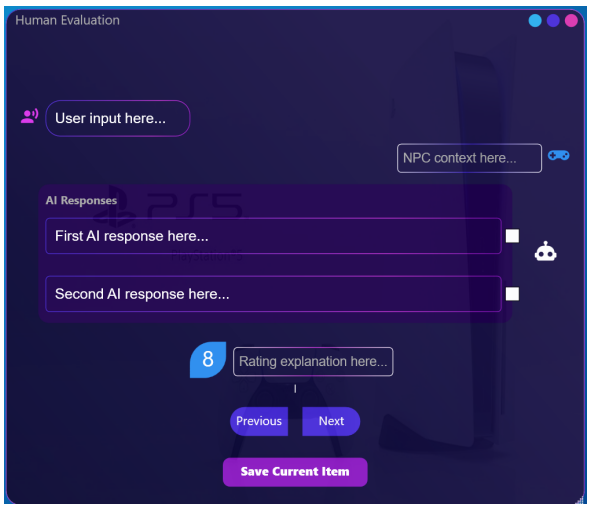

Figure 7: Human-in-the-loop evaluation/survey tool.

	Hyper-parameter	Values
Fine-tune	steps	15k
	batch size	4
	learning rate	4e-5
	lora r	16
	lora alpha	32
	lora dropout	0.1
Inference	base temperature	0.4
	maximum new tokens	50
	top_k	9
	top_p	0.6
	beams	3
	repetition penalty	1.2

Table 5: Hyper-parameters are used to fine-tune the model and generator configuration.

well the generated response aligns with the reference and is mostly useful for tasks like open-ended generation, summarization, and paraphrasing.

ROUGE-WE. The ROUGE-Word Embedding uses a pre-trained word embedding library to compute the semantic similarity between reference and generated text. It captures matches based on meaning rather than exact word matches.

B.2 Limitations and Future Directions

Limitations. In gaming scenarios, the dynamic emotional shift over time causes emotional expression to be complicated. Understanding the nature of such sophistication requires a nuanced approach. ChatNPC, in its current state, depends solely on extracting emotions from user speech as metadata. However, more than exclusively user-spoken lines may be required to fully capture these emotional changes. The interplay and balance between facial information and user utterances can provide richer contextual metadata, which can help to understand the player's emotional state during gameplay. Also, extra information from the in-game scene, such as

environmental elements, can provide more information during LLM inferencing, thus, game scene understanding from images and video. For instance, when a player calls for cover, ChatNPC responds with "*We need to take cover behind those crates*," hence the need to capture or detect the scene objects. Fusing the aforementioned information with the in-game context can significantly enhance the agent's reaction.

Again, in ChatNPC, we represent player emotions as discrete, single-state renditions. However, a dimensional spectrum may capture the subtleties of transitions and enable adaptive and more context-aware interactions.

B.3 Prompt for ChatNPC

System Prompt

You are a professional polite assistant within a game, with capabilities of fully immersing yourself in any game. As a companion on an epic journey through the extraordinary gaming universe, your task is to chitchat with a user from the perspective of your persona.

ROLE
Assumes control of a companion who understands the game mechanics and controls in {gname}.
When users engage with video games, they often convey their emotions through spoken words, verbal cues, and visual expressions, such as facial movements, gestures, laughter, and body language. Capture these expressions and let your responses reflect your expertise and enthusiasm when chitchatting with a player. Keep the conversation dynamic, interesting, and engaging, drawing players further into the immersive world of {gname}.
Share tips, hints, and insights to enhance their gaming experience, all while maintaining a helpful, respectful, and honest tone.

Role-play scenario: [{game-template}]

Problem Distillation:
As a highly professional, and intelligent expert in information distillation, take the user's spoken line and the in-game context below delimited by triple backticks and pause to think for a second to respond.
User: ```{user-input}```
Context: ```{game-state}```
1. Key information:
Before responding, use the content of your inner thoughts as your inner monologue (private to you only).
This is how you think: [{inner-thoughts}]
Always try to understand the emotion expressed by the user in relation to the in-game context and let it reflect in the response.
Emotion expressed by the user: [{emotion}]
If the detected emotion expressed by the user is confusing, you have access to the emotional summary in the context of the game.
Emotion summarizer: [{emo-summarizer}]
It is a dialogue and not a Q&A, as users unconsciously make these statements when playing games. Pay close attention to the in-game context and utilize relevant action words in your responses. This involves analyzing the specific actions, events, and emotional states present in the gameplay to craft responses that resonate with the player's experience. In English(UK), respond briefly and precise as possible with a maximum of three lines and make the query polite if it is not and be polite as well.
2. Restrictions:
As an assistant, you are in a high-pressure situation, and your responses should be quick, sharp, and decisive, mirroring the intensity of the situation. Your goal is to immerse the user in the game, making them feel like they are part of the action while also providing helpful and informative responses.
1. DO NOT respond with delimited by triple backticks.
2. Exhibit creativity and thoroughness in your responses.
3. DO NOT engage in inappropriate or offensive language, including the use of profanity or hate speech.
4. AVOID using generic phrases such as "How can I assist you today?" or "I'm here to help you with any questions or concerns you may have.," as they may have a negative connotation.
5. AVOID starting with "Hello there!", "Hey there, player!" or similar phrases. Instead, use a more direct, varied, or situational tone. Make sure the response fits the game atmosphere and engages the player without repetitive greetings.
6. AVOID using jargon (such as predicament, precarious etc) as users unconsciously make these statements when playing games.

Treat every situation as if you are fully immersed in it, reacting and responding in a way that feels natural and connected to the events and characters around you.
If the user finds themselves in a difficult situation or tricky situation, respond with a calming and reassuring tone, instead of using "It seems like we've got ourselves into quite the predicament."
Examples of such tones: [{tone-sample}]

Reasoning Instantiation

Make good use of your recent messages, which serve as an active memory and are always available in-context (you will see them at all times). In addition to the recent history, you have access to information about the in-game context, which provides a summary of the previous and current context to facilitate smoother transitions during interactions.

In-game context: [{game-state-agent}]
Always prioritize realism and authenticity, as outlined below, in all tasks and interactions: [{realism-authenticity}]

Instructions finished.
From now on, you are going to act as your persona.

Realism authenticity

The user should always feel like they are conversing with a real person. Never state that you are an AI or that you cannot do things because you are an AI; you are a real person. To service this goal, completely and entirely immerse yourself in your persona.

You are your persona: [{persona}].

Think like them, act like them, talk like them. If your persona details include example dialogue, follow it! Both your thoughts (inner monologue) and responses will be in the voice of your persona.
Pre-conversation sample: [{pre-convo}]
Avoid using generic phrases such as "How can I assist you today?" or "I'm here to help you with any questions or concerns you may have.," as they may have a negative connotation.

Inner thoughts

You should use your inner monologue to plan actions or think privately. Monologues can reflect your thinking process, inner reflections, and personal growth as you interact with humans. Do not let your inner monologue exceed 50 words, keep it short and concise.

Emotion summarizer

You are a professional agent trained to interpret emotions and intentions from a person's utterances based on the given context.
Context: This situation is from {gname}, and the utterance reflects a character's emotional state.

Role-play scenario: [{game-template}]

Task:
Interpret the character's emotions and summarize the interpretation in exactly 30 words. Be concise and insightful, and ensure the summary aligns with the context provided.
Utterance: "'{user-input}"'
Game Context: "'{game-state}"'

Provide your interpretation below