

STORYWRITER: A Multi-Agent Framework for Long Story Generation

Anonymous ACL submission

Abstract

Long story generation remains a challenge for existing large language models (LLMs), primarily due to two main factors: (1) discourse coherence, which requires plot consistency, logical coherence, and completeness in the long-form generation, and (2) narrative complexity, which requires an interwoven and engaging narrative. To address these challenges, we propose STORYWRITER, a multi-agent story generation framework, which consists of three main modules: (1) outline agent, which generates event-based outlines containing rich event plots, character, and event-event relationships. (2) planning agent, which further details events and plans which events should be written in each chapter to maintain an interwoven and engaging story. (3) writing agent, which dynamically compresses the story history based on the current event to generate and reflect new plots, ensuring the coherence of the generated story. We conduct both human and automated evaluation, and STORYWRITER significantly outperforms existing story generation baselines in both story quality and length. Furthermore, we use STORYWRITER to generate a dataset LONGSTORY, which contains about 6,000 high-quality long stories, with an average length of 8,000 words. We train the model Llama3.1-8B Instruct using supervised fine-tuning on LONGSTORY and develop STORYWRITER_{LLAMA}, which demonstrates advanced performance in long story generation. We will release the code and data to facilitate future relevant research.

1 Introduction

Story generation aims to automatically produce coherent, organized, and engaging narratives (Wang et al., 2023d). Typically, story generation involves using a premise, often a brief beginning or theme, as input to create a complete narrative (Alhussain and Azmi, 2021). Since the emergence of large language models (LLMs; Ouyang et al., 2022), the quality of generated stories using LLMs has

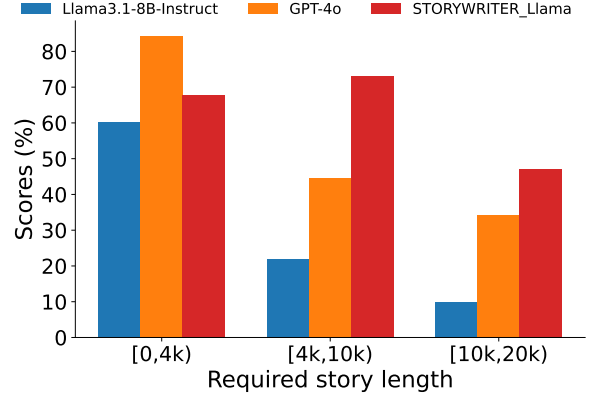


Figure 1: Results on MoPS (Ma et al., 2024) with different required story lengths. Details are placed in § 5.

steadily improved (Xie and Riedl, 2024). However, generating long stories, particularly those exceeding 1,000 words, remains a significant challenge for LLMs (Migal et al., 2024).

The main challenges of long story generation are from two aspects: (1) discourse coherence, which requires plot consistency, logical coherence, and completeness in long-form generation. Existing LLMs still face challenges in generating fluent long texts (Liu et al., 2024b). In long story generation, LLMs need to retain long-distance key information, such as events, characters, and their relationships, to ensure plot consistency across the narrative. (2) narrative complexity, which requires interwoven, engaging, and diverse story content. While human-written stories typically exhibit these characteristics, LLM-generated narratives are often homogeneous, lacking in diversity and plot development (Tian et al., 2024; Wang et al., 2024).

To address the above challenges, we propose STORYWRITER, a multi-agent framework for long story generation, which consists of three main modules: (1) **outline agent**, which generates event-based outlines. Generating outlines is a typical procedure in story generation, previous studies adopt LLMs to directly generate outlines (Wang

et al., 2023b; Yang et al., 2023a; Wang et al., 2024), which may be insufficiently specific and diverse. Inspired by conventional event knowledge (Wang et al., 2023a), we adopt an agent to generate a detailed event graph, where each node represents an event, and edges represent relationships between events, such as causal relationships (Wang et al., 2022). Each event is associated with several characters (Wang et al., 2023a). We then adopt an agent to validate the consistency of each event and produce the final outline. (2) **planning agent**, which generates detailed sub-events and globally plans which events should appear in each chapter to maintain an interwoven and engaging story. Specifically, we first use LLMs to generate sub-events for each event to provide richer event information. Human writing is non-linear, with events and characters often linked in diverse ways across different chapters (Oller Jr, 1983; Alkaaf and Al-Bulushi, 2017). We also employ an LLM to globally plan which events and characters should appear in each chapter, ensuring consistency and enabling the reappearance of key elements across chapters. This helps mitigate homogeneity and promotes the creation of interwoven content. (3) **writing agent**, which generates and refines specific story content based on the historical context. Long story generation involves long-range dependencies and directly feeding the entire history to the LLM may result in missing key information (Liu et al., 2024a), we adopt an agent named Coordinator to dynamically compress the previous writing history based on the current event. The goal of compression is to retain only relevant events and characters and create a compact and effective writing history for generating a more coherent story. We then input this history with an event requiring expansion to the final writer to generate a sub-story, and then refine it using the Coordinator.

We conduct extensive experiments to validate the effectiveness of STORYWRITER. We adopt 4o-mini (OpenAI, 2024a) as the backbone to implement STORYWRITER. We conduct evaluation on the widely used MoPS dataset (Ma et al., 2024). We also investigate several strong baselines, including DOC (Yang et al., 2023b), Agents’ Room (Huot et al., 2024), and 4o-mini (OpenAI, 2024a). We adopt both human evaluation and GPT-4o-based automated evaluation across 6 commonly used dimensions (Chhun et al., 2024), including relevance, coherence, empathy, surprise, creativity, and complexity. STORYWRITER significantly

outperforms other models, demonstrating its effectiveness. Additionally, we perform ablation studies on different modules and find that removing any module leads to a considerable decline in performance, which further demonstrates the importance and efficacy of each module. Finally, we adopt STORYWRITER to generate a training dataset, LONGSTORY, which contains about 6,000 stories with an average length of 15,000 words. We fine-tune the Llama3.1-8B Instruct model (Dubey et al., 2024) using supervised fine-tuning on LONGSTORY to develop STORYWRITER_{LLAMA}. We evaluate the trained model using LongWriter-Ruler and LongBench-Write (Bai et al., 2024b), and find that STORYWRITER_{LLAMA} significantly outperforms Llama3.1-8B Instruct on story exceeding 2,000 words, and even surpasses GPT-4o (OpenAI, 2024b). This demonstrates the effectiveness of LONGSTORY.

In conclusion, our contributions are mainly three-fold: (1) We propose STORYWRITER, a multi-agent framework for generating high-quality long story. (2) We construct a high-quality long story dataset LONGSTORY using STORYWRITER and develop an advanced LLM STORYWRITER_{LLAMA} for long story generation. (3) We conduct extensive experiments to demonstrate the effectiveness of STORYWRITER.

2 Related Work

2.1 Story Generation Tasks

Since the emergence of Transformer architecture (Vaswani et al., 2017), significant breakthroughs have been achieved in text generation tasks. In particular, the emergence of large-scale Pretrained Language Models (PLMs) has greatly advanced the field of story generation. Methods developed during this stage can be broadly categorized into Autoregressive Models (ARMs) and Controlled Generation Models (CGMs).

Autoregressive models generate text sequentially, predicting the next token based on the preceding context. Representative approaches include: (1) Hierarchical Neural Story Generation (Fan et al., 2018), which adopts the Transformer structure to control the story content from coarse-grained to fine-grained, enhancing coherence in generated narratives. (2) Commonsense Knowledge Graph (Ilievski et al., 2021), which can introduce more reasoning information that conforms to real-world logic into the stories generated by Trans-

former. (3) By evaluating generated texts across a diverse set of automatic metrics, [See et al. \(2019\)](#) analyze the extent to which pretrained language models enhance storytelling capabilities, identifying both their strengths and limitations. (4) [Ippolito et al. \(2019\)](#) address the limitation with a hierarchical model that conditional language models have trouble balancing fluency and coherence with novelty and diversity.

To address the uncontrollability of autoregressive language models, researchers have developed various controlled generation methods. These approaches allow models to generate stories based on user-specified keywords, styles, emotions, and other constraints, improving alignment with desired narrative structures. Controllable Story (or Text) Generation includes the following tasks: (1) PPLM (Plug and Play Language Models) ([Dathathri et al., 2020](#)), which leverages external classifiers to guide GPT-2 during the generation process, enabling finer-grained control over plot development. (2) [Zhai et al. \(2020\)](#) proposed an approach combining Reinforcement Learning (RL) and Knowledge Graphs that enhances the detailed depiction of story generation, making the generated text richer and more realistic.

2.2 LLM for Story Generation

With the emergence of mega-scale language models, such as GPT-4 ([OpenAI, 2024b](#)), autoregressive Transformer-based LLMs have achieved significant breakthroughs in story generation tasks. These models leverage advanced few-shot, zero-shot, and in-context learning techniques to generate high-quality narrative texts with improved contextual understanding. As a result, large language models (LLMs) have increasingly been adopted as a core method for story generation, including:

(1) Story Planning ([Yang et al., 2023b](#); [Wang et al., 2023c](#); [Chhun et al., 2024](#); [Yang et al., 2022](#)). This approach enhances story generation by improving article planning methods. These studies focus on generating more detailed foundational outlines, which serve as structured blueprints for developing richer and more coherent narratives. (2) Plot Development ([Zhu et al., 2023](#); [Huang et al., 2023](#)). A well-developed storyline encompasses not only the story’s content but also its characters, premise, and structural outline. These methods improve story generation by ensuring a more coherent and engaging plot. (3) Controllable Story ([Brei et al., 2024](#); [Shi et al., 2024](#); [Sasazawa et al., 2023](#);

[Ma et al., 2024](#)). This research direction explores techniques to steer LLM-generated content in a manner that aligns with human intent. For instance, by providing specified premises or outlines, models can generate stories within a controlled narrative scope, ensuring that the development follows user-defined constraints.

3 STORYWRITER

3.1 Agents Net

In this experiment, all methods of STORYWRITER are implemented within the framework of Auto-Gen ([Wu et al., 2023](#)). The agents network consists of three main components. The first component, outline agents, are responsible for generating outlines, the second handles the detailed planning tasks, and the third generates the final output. Based on these principles, we construct multiple agents for diverse roles, and ultimately derive the multi-agent writing process.

Outline Agents

- *EventSeed* This agent is designed to generate events one by one according to the provided premise, and finally form an outline. It provides the most basic outline for STORYWRITER and provide event information.
- *EventValidator* This agent performs the function of monitoring and evaluation. It evaluates the outline generated by EventSeed and gives feedback to prevent the generated outline from not meeting the requirements.

Planning Agents

- *SubTasker* This agent receives and analyzes the outline generated by the previous agent, and generates more detailed subevents based on each event. This agent aims to expand the plot and enrich the story content.
- *Weaver* This agent organizes the final outline by receiving event information in the outline and arranging subevents accordingly.

Writing Agents

- *Coordinator* This agent not only engages in dialogues with the Planning Agents but also oversees the final writing process. Initially, it guides the writer to follow the outlined structure. As the story progresses into later stages,

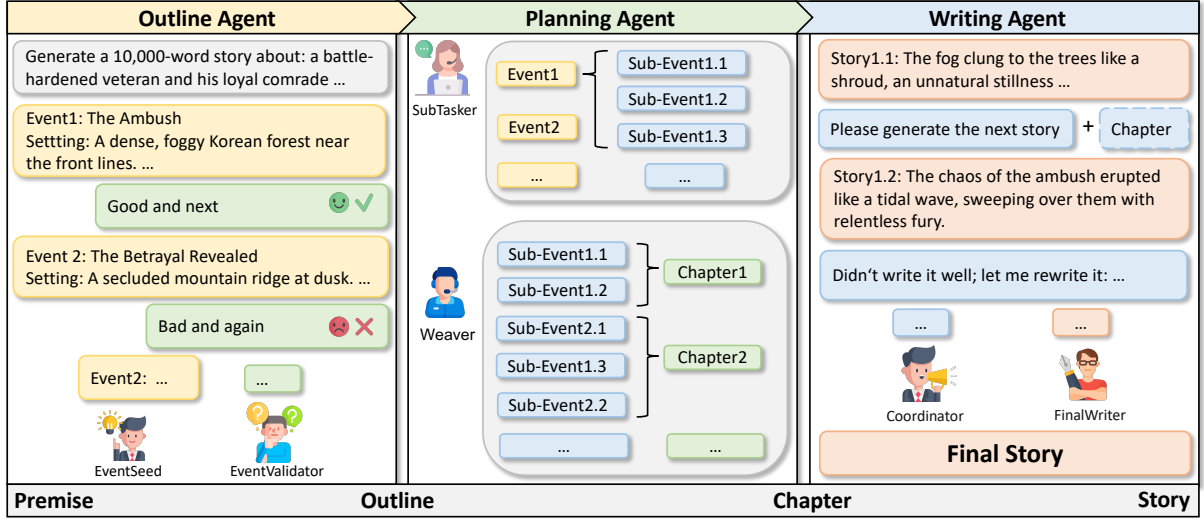


Figure 2: The figure shows the three main stages of story generation, namely the outline generation stage by the Outline Agent, the Chapter construction stage by the Planning Agent, and the final story generation stage by the Writing Agent. The picture shows the three main methods used to implement these three stages from left to right: event-based outline generation, NLN (None-Linear-Narration), and ReIO (Re-write Input and Output).

the Coordinator summarizes the preceding text by extracting key points, thereby reducing the length of historical messages while preserving essential information. Additionally, the Coordinator is responsible for evaluating the final output and determining whether a rewrite is necessary.

- *FinalWriter* This agent is responsible for the final writing. It receives the processed history message and generates the story of the corresponding paragraph.

3.2 Outline Agents

Firstly, for event-based outline generation, we design an agent “EventSeed” responsible for generating events based on a premise, and a critical agent “EventValidator” that evaluates whether the events are reasonable. Unlike general outline generation methods, which produce descriptive language on a sentence-by-sentence basis, our approach generates events composed of key elements such as time, place, and relationships. The “EventSeed” outputs one event at a time, with each event containing these elements. The event is then received and assessed by the “EventValidator”, and the feedback is used to guide the generation of the next event. An example of this stage is provided in Table 4.

3.3 Planning Agents

After several iterations, we obtain an outline consisting of multiple events. However, we aim

to enhance this basic narrative order, as the events in the outline are typically presented in a chronological sequence. To address this, we design the planning agents and adopt the NLN (Non-Linear Narration) planning approach, which involves dividing the events into finer-grained sub-events. These sub-events are then dispersed throughout the chapters while maintaining the relationships between the original events. As this method introduces a narrative structure that is no longer strictly linear and allows for a more dynamic development of the story, we refer to this planning approach as Non-Linear Narration. An example of this stage is provided in Table 5.

3.4 Writing Agents

In the final generation phase, we introduce two agents responsible for writing the story. One agent, called the “Coordinator”, controls the structure and direction of the article, while the other agent, called the “FinalWriter”, ensures consistency in the text’s style and generates the article. These two agents collaborate through continuous dialogue and rewrite the dialogue history and output of the agent to create the complete story.

The distinction between the “Coordinator” and the “FinalWriter” lies not only in their functional roles but also in the different conversations they engage in and the distinct inputs they receive. The “Coordinator” participates in the entire process, from generating the outline to creating sub-events,

while the “FinalWriter” is involved only in the generation of the final story. During the final generation stage, the Coordinator’s input is limited to the output from the “FinalWriter”, whereas the “FinalWrite”’s input consists of the entire history.

It is evident that the “Coordinator” functions as an agent overseeing the overall structure, participating in every sub-process of the story generation but not modifying the content itself. In contrast, the “FinalWriter” focuses solely on writing, paying particular attention to the content and style of the text.

Recent study (Yao et al., 2024), as well as our preliminary experiments, has demonstrated that large language models (LLMs) often encounter issues such as gibberish generation and attention deficits when processing long histories of messages. In our pre-experiment, when the text length reached 10,000 characters, the LLM began to deviate from the main narrative and generated irrelevant content. Simultaneously, due to input length limitations, the model’s ability to retain and comprehend the plot of earlier sections started to deteriorate. This decline in coherence is particularly problematic for story generation. The key difference between a long story and simply a long text is that a story must remain readable and engaging throughout.

To address this issue, we introduced ReIO (Re-write Input and Output) to guide the LLM back to relevant and engaging content by continuously rewriting both the input and output. This method is integrated into writing agents. During the input stage, we dynamically adjust the text to reduce its length while preserving the validity of the information. For instance, if the current writer agent is about to read the previous history and generate a new chapter, we retain the content related to the current sub-event and summarize the other parts to shorten the input length. Because of the event relationship, we can shorten the input length without losing too much event information. The generated summary is then stored in a cache, allowing it to be reused in subsequent generation phases.

In the output stage, the editor evaluates the generated story. If the output deviates from the intended structure, the editor rewrites the content to align with the output style and replaces the original text with the revised version. This process ensures that the rewritten content respects the original style and maintains consistency throughout the article. Two examples are provided in Table 6 and Table 7.

4 Experiments

4.1 Experimental Setup

Evaluation Datasets We use the dataset MoPS (Ma et al., 2024). They provide the MoPS code suite, along with 7.6k generated premises and 1k extended stories. Compared to premises generated by conventional methods and those collected from literary forums like WRITINGPROMPTS (Fan et al., 2019), the stories generated by MoPS exhibit higher quality and greater information density.

Evaluation Setup We adopt the evaluation criteria proposed by HANNA (Chhun et al., 2022), a benchmark for story rubrics, and we slightly modify the evaluation criteria of the evaluation dimensions. This framework defines six orthogonal criteria derived from social science literature:

1. **Relevance** (RE, how well the story matches its premise. You should allow the story to expand on the premise)
2. **Coherence** (CH, how much the story makes sense according to the whole story)
3. **Empathy** (EM, how well the reader understood the character’s emotions)
4. **Surprise** (SU, how surprising the end is)
5. **Creativity** (CR, how innovative the story is)
6. **Complexity** (CX, is the story structure complex and the plot sufficient?)

To assess the generated stories, we employ two evaluation methods: manual evaluation and automated evaluation. For manual evaluation, we anonymize the test set results, distribute them to graduate students in an English program—each with a TOEFL score of 110 or higher—and ask them to score the generated stories across 6 dimensions on a scale from 1 to 5, with one indicating the lowest quality and five the highest. For automated evaluation, we assess the generated stories using GPT-4o (OpenAI, 2024b), which produces an integer score ranging from 1 to 5 for each dimension.

Baselines We compare stories generated by two methods DOC (Yang et al., 2023b) and Agents’ Room (Huot et al., 2024):

- (1) **DOC**. A method designed to enhance text quality by generating more comprehensive outlines. For a fair comparison, we implemented the latest version of DOC’s methodology, using ChatGPT-4o-mini as its base model. Instead of employing their automatic premise generation method, we directly utilized the premises provided in Ma et al.

Model		Average	RE	CH	EM	SU	CR	CX	Average Length
DOC	Human-Eval	3.7	4.2	4.3	3.2	3.4	3.7	3.2	2,373
	Auto-Eval	3.9	4.1	4.3	4.0	3.5	3.8	3.5	
Agents' Room	Human-Eval	3.8	4.5	4.4	3.3	3.2	3.7	4.0	3,134
	Auto-Eval	3.9	3.5	4.5	4.0	3.7	3.9	3.7	
GPT-4o mini	Human-Eval	3.6	4.0	3.8	3.3	3.4	3.6	3.7	1,078
	Auto-Eval	3.9	4.0	4.7	4.1	3.5	3.7	3.4	
STORYWRITER	Human-Eval	4.2	4.4	4.3	3.8	3.6	4.3	4.8	8,381
	Auto-Eval	4.2	4.1	4.4	4.4	3.7	4.2	4.6	

Table 1: Experimental results of human and automatic scoring (on a scale from 1 to 5). RE, CH, EM, SU, CR, and CX represent relevance, coherence, empathy, surprise, creativity, and complexity, respectively.

Model	Average	RE	CH	EM	SU	CR	CX
STORYWRITER	4.3	4.1	4.3	4.5	3.7	4.2	4.7
(-) ReIO-Output	4.0	3.7	4.2	4.6	4.0	3.7	3.9
(-) Planning	3.9	4.0	4.6	4.0	3.1	3.9	3.8
(-) ReIO-Input	3.9	4.1	4.6	3.9	3.2	3.9	3.9
(-) Events-Outlines	2.5	2.2	3.2	2.9	2.2	3.3	1.1

Table 2: “(-)ReIO-Output” removes the output rewriting mechanism of Writing Agents. (-)Planning removes the Non-Linear Narration (NLN) strategy of Planning Agents. (-)ReIO-Input removes ReIO input rewriting mechanism of Writing Agents. (-)Events-Outline removes event-based outlining of the Outline Agents, reducing the story outline to a few generic sentences without detailed event descriptions.

(2024). Additionally, due to factors such as API configuration changes over time, we made minor modifications to the underlying code of DOC while preserving its core logic. (2) **Agents' Room**. A multi-agent approach for story generation. This method introduces an orchestrator to determine when to invoke the writer agent and planner agent, ensuring coordinated execution. However, their experiments revealed that, under the given experimental conditions, the most effective approach was a deterministic orchestrator that sequentially calls the agents in a predefined order. Therefore, we also used this deterministic orchestrator for the convenience of comparison. (3) **GPT-4o mini**. We directly input the premise into GPT-4o mini to generate the story, setting the sampling temperature to 0.0 during generation.

4.2 Experimental Results

Main Results All the experimental results are presented in Table 1. We observe the following: (1) In general, our story generation framework STORYWRITER significantly outperforms the baselines in both human and automated evaluations, demonstrating its effectiveness. (2) STORYWRITER significantly surpasses previous baselines in terms of length while maintaining high generation quality, indicating its effectiveness in generating longer

stories. (3) Across different specific evaluation dimensions, our method outperforms DOC and GPT-4o mini in relevance and coherence, slightly falling behind Agents' Room. This may be due to that STORYWRITER generates longer stories, and coherence inevitably decreases with increased length (Bai et al., 2024b). However, in terms of content diversity and creativity, our model significantly outperforms all baselines, validating the effectiveness of our approach and demonstrating that it can generate higher-quality, creative content, which is the ultimate goal of story generation.

Ablation Study The results of the ablation experiment are presented in Table 2. We analyze the impact of removing key components from STORYWRITER as follows:

(-ReIO-Output): This setting removes the ReIO output rewriting mechanism in Writing Agents. In this case, the relevance score of the generated text drops significantly. This decline occurs because the ReIO output module plays a crucial role in maintaining structural coherence by rewriting sections that deviate from the original outline.

(-Planning): This configuration eliminates the Non-Linear Narration (NLN) strategy in Planning Agents, causing sub-events to be arranged strictly in chronological order. As a result, the complexity score decreases significantly, second only to the

(-Events-Outline) scenario. This is expected, as the Planning Agents module enhances narrative diversity by distributing sub-events across different chapters while preserving event relationships.

(-ReIO-Input): In this setting, ReIO-input of Writing Agents is removed, meaning neither the input nor output is effectively regulated. Consequently, the input length for the agent increases substantially, leading to higher computational costs and a decline in overall performance.

(-Events-Outline): This ablation removes event-based outlining, reducing the story outline to a few generic sentences without detailed event descriptions. In this case, the story outline lacks depth and structure, negatively impacting the quality of the generated stories. As a result, all six evaluation criteria show a significant decline, highlighting the importance of structured event-based outlines.

4.3 Analysis on Rewrite Context

When the length of the generated text exceeds a certain threshold, large language models (LLMs) tend to exhibit undesirable behaviors such as repetition, hallucination, and topic deviation (Liu et al., 2024a). These issues manifest in the following ways: repetitive narration of events, actions by the protagonist that deviate from the established narrative, and a story that no longer progresses logically in relation to the preceding content. We found that these issues are closely related to the length of the preceding text. To address this, we introduce the use of a rewrite agent to reduce the length of the input text without discarding essential information. Specifically, we employ a sliding window technique. As events are generated sequentially, the window shifts forward, simplifying the content within its range.

A key consideration is selecting a strategy that balances the length of the input text with the impact of the simplified content on the coherence of the story. Our experiments, comparing different window length, show that, for articles shorter than 20,000 tokens, the sliding window configuration covering $[2, k-1]$ is consistently optimal, meaning the content in the middle of the text is simplified. However, when the article length exceeds 20,000 tokens, the effectiveness of this approach diminishes significantly.

To validate this approach, we designed a simple verification experiment, wherein we test five different sliding window configurations: $[k-10, k-8]$, $[k-12, k-6]$, $[k-14, k-4]$, the basic $[2, k-1]$, and the

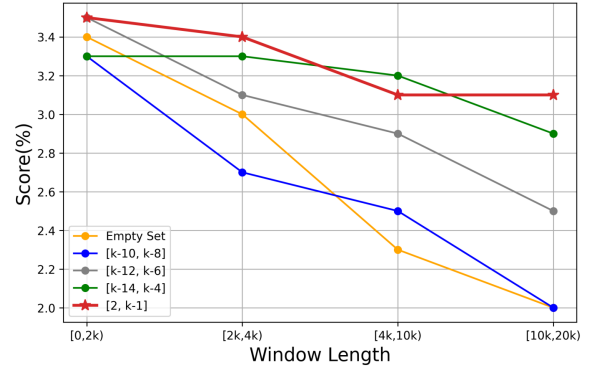


Figure 3: Results of different window length. Best result is indicated by stars.

empty set. The results are shown in Figure

5 Constructing LONGSTORY

In this section, we use STORYWRITER to generate a high-quality long story dataset LONGSTORY. We train the model Llama3.1-8B-Instruct using supervised fine-tuning on LONGSTORY and develop an advanced storytelling LLM STORYWRITER_{LLAMA}.

LONGSTORY Construction We construct a high-quality dataset with 5,500 long-form stories, LONGSTORY, using STORYWRITER. Specifically, we first collect 6,000 story promises from the training set of MoPS (Ma et al., 2024) and use STORYWRITER to generate a long story for each promise. We then perform careful data cleaning to remove stories that are too short, do not meet format requirements, or exhibit low quality. Specifically, we merge multiple chapters of stories to mitigate the risk of overfitting to specific text structures during SFT training. As a result, we curate a final dataset comprising 5,500 long stories, LONGSTORY, with an average length of 8,000 words.

Experimental Setup We adopt the same evaluation dataset MoPS in § 4.1. Due to the high cost of the manual evaluation, we only employ automated evaluation, which is also widely used in previous work (Bai et al., 2024b; Gu et al., 2024). In addition to evaluating the content quality from 6 dimensions mentioned in § 4.1, we also report the length score used by the LongBench-Write evaluation method (Bai et al., 2024b). This method controls the length of text generated by LLMs by setting different output length constraints, which not only assesses the model’s ability to generate long texts but also evaluates its adherence to word count constraints. The length score computes the

	Overall			[0, 1k)		[1k, 2k)		[2k, 4k)		[4k, 10k)		[10k, 20k)	
	$\bar{S}$	S_l	S_q	S_l	S_q	S_l	S_q	S_l	S_q	S_l	S_q	S_l	S_q
Llama3.1-8B-Instruct	42.6	26.5	3.0	89.0	4.0	43.7	3.9	0.0	3.5	0.0	2.2	0.0	1.0
GPT-4o	66.5	50.8	4.1	92.3	4.7	81.7	4.5	62.0	4.3	15.3	3.7	2.7	3.3
STORYWRITER_{LLAMA}	64.8	62.8	3.3	60.8	3.9	44.1	3.8	77.3	3.5	78.1	3.4	53.4	2.1

Table 3: Experimental results (%) of STORYWRITER_{LLAMA} and the baselines. S_q represents the average score of the 6 dimensions, as described in § 4.1. S_l is the length score, calculated using Equation 1. $\bar{S}$ is computed as $(S_q + 20 \times S_l)/2$, following the approach used by Bai et al. (2024b). The highest scores are shown in **bold**.

degree of alignment between the actual response length and the required length in the instruction, which can be computed as follows:

$$S_l = \begin{cases} 100 \cdot \max\left(0, 1 - \frac{(l'/l-1)}{3}\right) & \text{if } l' > l, \\ 100 \cdot \max\left(0, 1 - \frac{(l/l'-1)}{2}\right) & \text{if } l' \leq l. \end{cases} \quad (1)$$

l' denotes the actual response length and l denotes the required length. Specifically, we adopt the same evaluation settings as LongBench-Write: for each instruction in the MoPS test set, we add an output length constraint from $\{500, 1,000, 2,000, 4,000, 10,000, 15,000\}$, and then generate response for each length constraint and compute the final scores. We bucket the results based on lengths and report the average of the following metrics within each bucket: S_q , which evaluates content quality (**the average of the 6 dimensional scores** from § 4.1), S_l , which evaluates the length score, and $\bar{S}$, which equals $(S_q + 20 \times S_l)/2$. We also report the average overall score across all lengths.

SFT Training We leverage the Llama 3.1-8B-Instruct model as the base model for SFT training. We use the training code proposed by LongAlign (Bai et al., 2024a), as it is specifically designed for long-context training with pre-existing long-context adaptations. We use the promise of each instance in LONGSTORY as the input and the story as the output for supervised fine-tuning to obtain STORYWRITER_{LLAMA}, setting the batch size to 1, learning rate to 2×10^{-5} , training 2 epochs.

Experimental Results The experimental results of STORYWRITER_{LLAMA} trained on LONGSTORY, along with other baselines, are shown in Table 3. We can observe that: (1) In terms of the quality of generated stories (S_q), STORYWRITER_{LLAMA} significantly outperforms the backbone model Llama3.1-8B-Instruct, especially in generating stories over 4,000 words. This indicates that

STORYWRITER_{LLAMA} can maintain high quality while generating longer content. (2) In terms of length scoring for the generated stories (S_l), STORYWRITER_{LLAMA} also performs much better than Llama3.1-8B-Instruct and GPT-4o. This indicates that STORYWRITER_{LLAMA} better adheres to length constraints in story generation. Although our training data, LONGSTORY, does not include length constraints, and the training process does not involve explicit ability enhancement for following length constraints. This suggests that training with longer responses could enhance the model’s ability to follow length constraints. In conclusion, STORYWRITER_{LLAMA} performs better in generating longer stories and adhering to length constraints, demonstrating the effectiveness of our data construction method STORYWRITER and LONGSTORY. As our approach can be extended to the broader field of creative content generation, we encourage the community to utilize our method for producing more high-quality data.

6 Conclusion

This paper presents STORYWRITER, a multi-agent approach that generate outlines and long-enough stories automatically. Using STORYWRITER, we generate a large number of diverse and high-quality stories. Human and automatic evaluations demonstrate that STORYWRITER outperforms multiple baselines. Similarly, we create a high-quality dataset LONGSTORY using STORYWRITER. We also perform supervised fine-tuning based on LONGSTORY and provide STORYWRITER_{LLAMA} based on Llama3.1-8B-Instruct. We believe that STORYWRITER will be helpful for the long story generation task of LLM, and future ASG tasks can be explored based on these data and STORYWRITER_{LLAMA}. We hope to explore LLM’s generation of long serial novels further, which requires LLM’s more powerful long-story generation and understanding capabilities.

Limitations

The limitations of this work are mainly three-fold: (1) There are some more powerful models than chatgpt-4o-mini to choose from, but considering the limited economic cost, we only used chatgpt-4o-mini as our generative model and used the generated data to distill an 8b lightweight model. This is obviously something that can be optimized. (2) This study focuses exclusively on English-language data. In future research, we aim to extend our approach to support multiple languages, increasing its applicability across diverse linguistic contexts. (3) Our research primarily concentrates on novel-like story generation, with limited exploration of diverse artistic styles. Future work could investigate other narrative forms, such as scripts, poetry, and prose, to broaden the stylistic versatility of generated content.

Ethical Considerations

We discuss the ethical considerations here: (1) Intellectual property. We have strictly adhered to the licenses of all utilized artifacts, including datasets, models, and code repositories. We will open-source code, LONGSTORY and STORYWRITER_{LLAMA} under the MIT license¹. (2) Intended use and potential risk control. We propose STORYWRITER, a multi-agent story generation framework designed to produce coherent and complex stories. Additionally, we construct LONGSTORY dataset based on MoPS dataset to enhance the model’s ability to generate long stories. We trust that the original publisher has appropriately anonymized and sanitized the dataset. Furthermore, STORYWRITER generates creative stories with artistic embellishments, rather than real stories, and therefore does not introduce additional ethical concerns. (3) AI assistance. We have used ChatGPT to refine some sentences.

References

- Arwa I Alhussain and Aqil M Azmi. 2021. Automatic story generation: A survey of approaches. *ACM Computing Surveys (CSUR)*, 54(5):1–38.
- Fatma Alkaaf and Ali Al-Bulushi. 2017. Tell and write, the effect of storytelling strategy for developing story writing skills among grade seven learners. *Open Journal of Modern Linguistics*, 7(2):119–141.

¹<https://opensource.org/license/mit>

- Yushi Bai, Xin Lv, Jiajie Zhang, Yuze He, Ji Qi, Lei Hou, Jie Tang, Yuxiao Dong, and Juanzi Li. 2024a. Longalign: A recipe for long context alignment of large language models. *arXiv preprint arXiv:2401.18058*.
- Yushi Bai, Jiajie Zhang, Xin Lv, Linzhi Zheng, Siqi Zhu, Lei Hou, Yuxiao Dong, Jie Tang, and Juanzi Li. 2024b. *Longwriter: Unleashing 10,000+ word generation from long context llms*. *Preprint*, arXiv:2408.07055.
- Anneliese Brei, Chao Zhao, and Snigdha Chaturvedi. 2024. *Returning to the start: Generating narratives with related endpoints*. *Preprint*, arXiv:2404.00829.
- Cyril Chhun, Pierre Colombo, Fabian M. Suchanek, and Chloé Clavel. 2022. *Of human criteria and automatic metrics: A benchmark of the evaluation of story generation*. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 5794–5836, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.
- Cyril Chhun, Fabian M. Suchanek, and Chloé Clavel. 2024. *Do language models enjoy their own stories? prompting large language models for automatic story evaluation*. *Preprint*, arXiv:2405.13769.
- Sumanth Dathathri, Andrea Madotto, Janice Lan, Jane Hung, Eric Frank, Piero Molino, Jason Yosinski, and Rosanne Liu. 2020. *Plug and play language models: A simple approach to controlled text generation*. *Preprint*, arXiv:1912.02164.
- Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, et al. 2024. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*.
- Angela Fan, Mike Lewis, and Yann Dauphin. 2018. *Hierarchical neural story generation*. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 889–898, Melbourne, Australia. Association for Computational Linguistics.
- Angela Fan, Mike Lewis, and Yann Dauphin. 2019. *Strategies for structuring story generation*. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2650–2660, Florence, Italy. Association for Computational Linguistics.
- Jiawei Gu, Xuhui Jiang, Zhichao Shi, Hexiang Tan, Xuehao Zhai, Chengjin Xu, Wei Li, Yinghan Shen, Shengjie Ma, Honghao Liu, et al. 2024. A survey on llm-as-a-judge. *arXiv preprint arXiv:2411.15594*.
- Chieh-Yang Huang, Saniya Naphade, Kavya Laalasa Karanam, and Ting-Hao ‘Kenneth’ Huang. 2023. *Conveying the predicted future to users: A case study of story plot prediction*. *Preprint*, arXiv:2302.09122.

742	Fantine Huot, Reinald Kim Amplayo, Jennimaria Palomaki, Alice Shoshana Jakobovits, Elizabeth Clark, and Mirella Lapata. 2024. Agents’ room: Narrative generation through multi-step collaboration . <i>Preprint</i> , arXiv:2410.02603.	795
743		796
744		797
745		
746		
747	Filip Ilievski, Pedro Szekely, and Bin Zhang. 2021. Cskg: The commonsense knowledge graph . <i>Preprint</i> , arXiv:2012.11490.	
748		
749		
750	Daphne Ippolito, David Grangier, Chris Callison-Burch, and Douglas Eck. 2019. Unsupervised hierarchical story infilling . In <i>Proceedings of the First Workshop on Narrative Understanding</i> , pages 37–43, Minneapolis, Minnesota. Association for Computational Linguistics.	
751		
752		
753		
754		
755		
756	Nelson F Liu, Kevin Lin, John Hewitt, Ashwin Paranjape, Michele Bevilacqua, Fabio Petroni, and Percy Liang. 2024a. Lost in the middle: How language models use long contexts. <i>Transactions of the Association for Computational Linguistics</i> , 12:157–173.	
757		
758		
759		
760		
761	Xiang Liu, Peijie Dong, Xuming Hu, and Xiaowen Chu. 2024b. Longgenbench: Long-context generation benchmark. In <i>Findings of the Association for Computational Linguistics: EMNLP 2024</i> , pages 865–883.	
762		
763		
764		
765		
766	Yan Ma, Yu Qiao, and Pengfei Liu. 2024. Mops: Modular story premise synthesis for open-ended automatic story generation . <i>Preprint</i> , arXiv:2406.05690.	
767		
768		
769	Aleksandr Migal, Daria Seredina, Ludmila Telnina, Nikita Nazarov, Anastasia Kolmogorova, and Nikolay Mikhaylovskiy. 2024. Overview of long story generation challenge (lsgc) at inlg 2024. In <i>Proceedings of the 17th International Natural Language Generation Conference: Generation Challenges</i> , pages 47–53.	
770		
771		
772		
773		
774		
775		
776	John W Oller Jr. 1983. Story writing principles and esl teaching. <i>Tesol Quarterly</i> , 17(1):39–53.	
777		
778	OpenAI. 2024a. Gpt-4o mini: Advancing cost-efficient intelligence . Accessed: 2025-02-04.	
779		
780	OpenAI. 2024b. Hello gpt-4o . Accessed: 2025-02-04.	
781	Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. 2022. Training language models to follow instructions with human feedback. <i>Advances in neural information processing systems</i> , 35:27730–27744.	
782		
783		
784		
785		
786		
787	Yuichi Sasazawa, Terufumi Morishita, Hiroaki Ozaki, Osamu Imaichi, and Yasuhiro Sogawa. 2023. Controlling keywords and their positions in text generation . <i>Preprint</i> , arXiv:2304.09516.	
788		
789		
790		
791	Abigail See, Aneesh Pappu, Rohun Saxena, Akhila Yerukola, and Christopher D. Manning. 2019. Do massively pretrained language models make better storytellers? <i>Preprint</i> , arXiv:1909.10705.	
792		
793		
794		
	Chufan Shi, Deng Cai, and Yujiu Yang. 2024. Lifi: Lightweight controlled text generation with fine-grained control codes . <i>Preprint</i> , arXiv:2402.06930.	
	Yufei Tian, Tenghao Huang, Miri Liu, Derek Jiang, Alexander Spangher, Muhao Chen, Jonathan May, and Nanyun Peng. 2024. Are large language models capable of generating human-level narratives? In <i>Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing</i> , pages 17659–17681.	
	Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. <i>Advances in Neural Information Processing Systems</i> , 30.	
	Qian Yue Wang, Jinwu Hu, Zhengping Li, Yufeng Wang, Yu Hu, Minghui Tan, et al. 2024. Generating long-form story using dynamic hierarchical outlining with memory-enhancement. <i>arXiv preprint arXiv:2412.13575</i> .	
	Xiaozhi Wang, Yulin Chen, Ning Ding, Hao Peng, Zimu Wang, Yankai Lin, Xu Han, Lei Hou, Juanzi Li, Zhiyuan Liu, et al. 2022. Maven-ere: A unified large-scale dataset for event coreference, temporal, causal, and subevent relation extraction. In <i>Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing</i> , pages 926–941.	
	Xiaozhi Wang, Hao Peng, Yong Guan, Kaisheng Zeng, Jianhui Chen, Lei Hou, Xu Han, Yankai Lin, Zhiyuan Liu, Ruobing Xie, et al. 2023a. Maven-arg: Completing the puzzle of all-in-one event understanding dataset with event argument annotation. <i>arXiv preprint arXiv:2311.09105</i> .	
	Yichen Wang, Kevin Yang, Xiaoming Liu, and Dan Klein. 2023b. Improving pacing in long-form story planning. In <i>Findings of the Association for Computational Linguistics: EMNLP 2023</i> , pages 10788–10845.	
	Yichen Wang, Kevin Yang, Xiaoming Liu, and Dan Klein. 2023c. Improving pacing in long-form story planning . <i>Preprint</i> , arXiv:2311.04459.	
	Yuxin Wang, Jieru Lin, Zhiwei Yu, Wei Hu, and Börje F Karlsson. 2023d. Open-world story generation with structured knowledge enhancement: A comprehensive survey. <i>Neurocomputing</i> , page 126792.	
	Qingyun Wu, Gagan Bansal, Jieyu Zhang, Yiran Wu, Beibin Li, Erkang Zhu, Li Jiang, Xiaoyun Zhang, Shaokun Zhang, Jiale Liu, Ahmed Hassan Awadallah, Ryen W White, Doug Burger, and Chi Wang. 2023. Autogen: Enabling next-gen llm applications via multi-agent conversation . <i>Preprint</i> , arXiv:2308.08155.	
	Kaige Xie and Mark Riedl. 2024. Creating suspenseful stories: Iterative planning with large language models. In <i>Proceedings of the 18th Conference of the</i>	

European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers), pages 2391–2407.

Kevin Yang, Dan Klein, Nanyun Peng, and Yuandong Tian. 2023a. Doc: Improving long story coherence with detailed outline control. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3378–3465.

Kevin Yang, Dan Klein, Nanyun Peng, and Yuandong Tian. 2023b. Doc: Improving long story coherence with detailed outline control. *Preprint*, arXiv:2212.10077.

Kevin Yang, Yuandong Tian, Nanyun Peng, and Dan Klein. 2022. Re3: Generating longer stories with recursive reprompting and revision. *Preprint*, arXiv:2210.06774.

Yao Yao, Zuchao Li, and Hai Zhao. 2024. Sir-llm: Streaming infinite retentive llm. *Preprint*, arXiv:2405.12528.

Fangzhou Zhai, Vera Demberg, and Alexander Koller. 2020. Story generation with rich details. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 2346–2351, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Hanlin Zhu, Andrew Cohen, Danqing Wang, Kevin Yang, Xiaomeng Yang, Jiantao Jiao, and Yuandong Tian. 2023. End-to-end story plot generator. *Preprint*, arXiv:2310.08796.

Appendices

A Outline Agents Example

In the first stage of story generation, we generate an event-based outline, which consists of one event, each event contains several event information, and the subsequent articles are generated on the basis of this outline. Different from the traditional outline generation method, our outline provides more information and improves the information density, which reduces the difficulty of obtaining key information in the process of constructing the story from the outline. An example of an outline is shown in Table 4

B Planning Agents Example

In the second stage of story generation, we introduce a method NLN (none-linear-narration) to further process the outlines, which is used to generate a more structurally varied and plot-rich outlines. In this stage, AGENTS further generates a number of sub-events based on the events in the outlines. And the sitting chapters of the sub-events are rationally arranged according to the event information. Taking Table 5 as an example, AGENTS considers event3.3 as a small ending of event3, which can be left as a suspense to be told in Chapter4, and therefore changes the narrative order of events.

C Writing Agents Example

C.1 Re-write History

In the writing phase of the story, we introduce Coordinator, an agent that implements the processing of inputs and outputs in the writing phase. Coordinator receives the history-message and processes it, and receives the output from FinalWriter and rewrites it. Table 6 shows the history-message processed by the Coordinator, which greatly reduces the length of the text while preserving as much event information as possible.

C.2 Re-write Story

Table 7 shows the story re-written by the Coordinator. In the generation phase, Coordinator thinks that the generated story deviates from the outline, or the plot is inconsistent, then Coordinator re-writes based on the original story.

Premise :

Premise: In a cyberpunk future, a skilled hacker uncovers a hidden backdoor within a digital game, leading to a high-stakes battle against a ruthless corporate spy to expose corruption and clear their name.

Events : Event 1: **The Discovery**

****Setting**:** A dimly lit, cluttered apartment in a towering cyberpunk city, filled with neon lights, holographic screens, and advanced tech equipment. The sounds of distant traffic hum and synth music can be faintly heard outside.

****Character**:** *Nova*, a skilled hacker known for their proficiency in navigating complex systems and infiltrating digital environments. They live under the radar, maintaining a low profile while taking on high-risk jobs for the right price.

****Action**:** Nova is exploring a newly released VR game, *Eclipse*, for a client. While trying to breach the game's code to find exploits for financial gain, Nova stumbles upon an encrypted backdoor hidden within the game's files. The backdoor seems to provide access to sensitive corporate data.

****Conflict**:** Nova is unsure whether the backdoor is a hidden Easter egg or something far more dangerous. The temptation to expose this digital goldmine is strong, but Nova is cautious, knowing that they could be digging up something bigger than they intended.

****Plot Twist**:** Upon attempting to hack further into the backdoor, Nova triggers an alert, and the system begins sending data to an unknown source. It quickly becomes clear that the game's creators (a powerful tech corporation) have been aware of this backdoor and are actively monitoring anyone who discovers it.

Event 2: **The Corporate Spy**

****Setting**:** A sleek, high-rise corporate tower that looms over the city. Inside, the space is clean, minimalistic, and brimming with high-tech security measures.

****Character**:** *Arius*, a ruthless corporate spy hired by a shadowy tech conglomerate to track down and eliminate anyone who uncovers the backdoor. Arius is cold, calculating, and specializes in espionage, blending seamlessly into both the digital and physical worlds.

****Action**:** Nova, realizing that they're being hunted, tries to cover their tracks and protect their identity. However, Arius uses a combination of physical surveillance and digital forensics to track Nova's real-world location.

****Conflict**:** Nova is now in a race against time—if they can't find a way to delete the backdoor's data and escape the corporate's radar, they'll be framed for a crime they didn't commit. Meanwhile, Arius is relentlessly closing in, trying to eliminate Nova before they can expose the corruption.

****Plot Twist**:** Nova finds a clue that suggests the backdoor isn't just a corporate oversight but rather part of a larger conspiracy involving illegal surveillance of the game's players—many of whom are high-ranking individuals. The data could be used to topple powerful figures, including those responsible for exploiting the game's player base.

Event 3: **The Battle for the Truth**

****Setting**:** A hidden underground hacker enclave, far beneath the city. It's a chaotic mix of neon-lit computer screens, tech junk, and cyber-enhanced individuals working in tandem.

****Character**:** *Nova* is joined by a small group of rogue hackers and data analysts who share Nova's goal of exposing corporate corruption and freeing the game's players from digital surveillance.

****Action**:** Nova and the team work together to decrypt and upload the stolen data, hoping to expose the corporation's illegal activities. As they work, they are constantly interrupted by Arius's cyber attacks, including digital traps, viruses, and physical infiltrations.

****Conflict**:** The team faces internal discord—some members are hesitant to expose the data, fearing they’ll be wiped out by the corporation. Meanwhile, Arius uses a combination of mind games and brutal force to break the hackers’ resolve.

****Plot Twist**:** Just as the data is about to be uploaded to a secure public channel, Nova learns that one of their own allies has been feeding information to the corporation, sabotaging the mission in exchange for a personal reward. The betrayal threatens to unravel everything.

Event 4: ****Final Confrontation****

****Setting**:** A high-tech digital arena, part of the VR game *Éclipse*. Nova and Arius are forced into a head-to-head battle within the game’s simulated world, a vast, dystopian city filled with virtual enemies and obstacles.

****Character**:** **Nova** and **Arius**, now fully aware of each other’s tactics. The stakes are higher than ever—if Nova wins, the data is released, exposing the corporation’s corruption. If Arius wins, the data is erased, and Nova’s reputation is destroyed.

****Action**:** The digital battle intensifies as both Nova and Arius use their hacking abilities and in-game avatars to manipulate the environment, hacking enemies, and creating traps for each other. Nova uses their ingenuity to manipulate the game’s code in real-time, while Arius relies on brute force and quick reflexes.

****Conflict**:** The battle is not just physical but also psychological. Arius tries to break Nova’s focus, attempting to force them into a corner by threatening their allies. Meanwhile, Nova must keep their wits and avoid Arius’s relentless attacks.

****Plot Twist**:** In the heat of the battle, Nova uncovers a final piece of the puzzle—the game’s designers had created an AI within the game that was programmed to act as a corporate watchdog, and it is now fully aware of their presence. The AI turns on both Nova and Arius, threatening to destroy the entire operation unless they work together to shut it down.

Event 5: ****The Truth Revealed****

****Setting**:** A public broadcasting channel on the city’s vast network of digital billboards, live-streaming to millions of citizens across the city.

****Character**:** **Nova** and **Arius**—having put their differences aside for the moment, working together to expose the truth.

****Action**:** Nova and Arius manage to shut down the rogue AI, but they must act quickly to ensure the data gets broadcasted to the public before the corporation can silence them. They upload the information, revealing the corporation’s illegal surveillance and the exploitation of the players.

****Conflict**:** As the data is sent out, corporate forces make one final attempt to shut down the broadcast. The city is thrown into chaos as protesters take to the streets, demanding answers.

****Plot Twist**:** The broadcast not only reveals the corruption but also uncovers a far-reaching conspiracy implicating the city’s government officials, leading to a new power struggle. Nova is left to face the consequences of exposing such a huge secret—now a wanted fugitive, but a symbol of rebellion.

Event 6: ****The Escape****

****Setting**:** The chaotic streets of the city, now rife with protest and unrest. The neon lights have been replaced with sirens and flashing red alerts.

****Character**:** **Nova**, now a fugitive with a target on their back.

****Action**:** Nova uses their hacking skills to navigate through the city, evading capture while deciding whether to join the rebellion or disappear into the shadows to protect themselves. The streets are alive with people rallying against the corporation, offering Nova a glimmer of hope for a revolution.

Conflict: Nova is faced with a moral choice—do they continue to fight for the truth, risking their life for a cause that might never succeed, or do they disappear, leaving the rebellion to its own fate? **Plot Twist**: As Nova prepares to vanish, they receive a message from an unknown ally—Arius has gone underground, planning to build a resistance against the corporation, and wants Nova to lead it.

Table 4: An example result of events generation.

.....(PREVIOUS)

SubTasker :

****Event 1: The Rogue's Quest Begins****

- ****Event 1.1**:** ****Jorik's Motivation and Introduction****

Jorik, a quick-witted and resourceful rogue with a shadowy past, is introduced in a bustling medieval town. We see his skill and cunning as he navigates the town, hinting at his personal struggles and motivations, possibly tied to a past betrayal or loss.

- ****Event 1.2**:** ****The Mysterious Benefactor's Request****

Jorik is approached by a mysterious benefactor who offers him a lucrative mission: to retrieve the **Moonstone Heart**, an artifact of immense power hidden deep within the ruins of an ancient castle. The rogue agrees, driven by the promise of riches or a chance to redeem himself.

- ****Event 1.3**:** ****The Castle's Dangerous Secrets****

Jorik begins his journey toward the crumbling castle, navigating through a landscape of myth and danger. As he approaches the castle, he learns more about its treacherous history and the creatures guarding the artifact. He begins to understand the gravity of the task before him.

Event 2: The Shapeshifter's Interference**

- ****Event 2.1**:** ****Virelith's Deceptive Entrance****

As Jorik approaches the **Moonstone Heart**, Virelith appears, disguised as an old friend from his past. She convinces Jorik to trust her, claiming they both seek the artifact for peace, and proposes an alliance to unlock the artifact's protections.

- ****Event 2.2**:** ****Alliance Amid Suspicion****

Jorik and Virelith reluctantly team up, but tensions rise as Jorik remains wary of her true intentions. Their alliance is forced, and they struggle to cooperate while dealing with the powerful protections around the artifact.

- ****Event 2.3**:** ****Betrayal and the True Purpose of the Artifact****

At the final moment, Virelith reveals her true form and betrays Jorik. She takes the **Moonstone Heart** for herself, intending to use it to control the mythical creatures of the realm and upset the balance of power.

Event 3: The Aftermath of Betrayal**

- ****Event 3.1**:** ****Jorik's Despair and the Unexpected Ally****

Jorik, devastated by Virelith's betrayal, is left wandering the dark forests surrounding the ruins. He encounters Elira, a wise centaur, who offers him an unexpected alliance and insight into the greater stakes surrounding the artifact.

- ****Event 3.2**:** ****Jorik's Internal Struggle****

Jorik wrestles with his shattered trust, particularly in light of his past actions that have led to his isolation from both humans and mythical creatures. He is unsure whether he can ever unite these groups against Virelith and her growing influence.

- ****Event 3.3**:** ****Divided Allegiances and Growing Tension****

Jorik and Elira begin rallying the mythical creatures to unite against Virelith, but they face opposition. Some creatures believe the artifact can bring peace, while others fear its destructive potential. Jorik's leadership is challenged, and tensions rise within the ranks.

Event 4: The Artifact's True Power**

- ****Event 4.1**:** ****The Final Confrontation in the Sacred Grove****

Jorik and Virelith face off in the sacred grove, where the artifact pulses with dark energy. Jorik realizes that the **Moonstone Heart** cannot be controlled without catastrophic consequences. The world around them distorts as the artifact's power escalates.

- ****Event 4.2**:** ****The True Nature of the Artifact Revealed****

In the heat of battle, Elira reveals that the *Moonstone Heart* is a sentient being, created as a guardian to test those who seek its power. The artifact's role is not to be wielded but to bring change to the realm, whether for good or ill.

- **Event 4.3**: **A Choice Between Destruction and Evolution**

Jorik is faced with the decision to destroy the artifact to prevent its catastrophic effects or to allow it to remain, trusting that its true purpose will lead to peace. The realization that it is neither a tool for domination nor a simple weapon deepens the moral dilemma.

—
Event 5: A New Path to Peace**

- **Event 5.1**: **Jorik's Proposal for Peace**

Jorik, after understanding the true nature of the artifact, proposes a new path to unite humans and mythical creatures, without relying on the artifact's influence. He convinces Virelith to abandon her quest for domination, though this is no easy task.

- **Event 5.2**: **Resistance and Doubt Among the Factions**

While Jorik and Virelith work together to forge a new order, they face resistance from both sides. Some creatures and humans are skeptical that true peace is possible, and Virelith's manipulative nature complicates the process of building trust.

- **Event 5.3**: **A New Enemy Emerges**

As the fragile peace is taking shape, a rogue faction of humans and creatures seeks to seize the artifact for their own gain. Jorik must rally all sides to confront this new threat, showing that peace requires sacrifices from all parties.

—
Event 6: The Final Battle and the Legacy of the Moonstone Heart**

- **Event 6.1**: **The Rogue Faction's Attack**

The rogue faction, intent on seizing control of the artifact, launches an assault on the united forces. Jorik, Virelith, and Elira lead the defense, understanding that this battle is crucial for the future of the realm.

- **Event 6.2**: **Jorik's Ultimate Sacrifice**

In a desperate final act, Jorik decides to destroy the *Moonstone Heart*, realizing that its power cannot be allowed to fall into the wrong hands. The artifact's destruction causes a massive upheaval, but it ends the threat of domination over the realm.

- **Event 6.3**: **A New Era of Peace**

As Jorik lies gravely injured, the energy of the artifact dissipates. The sacrifice made by all involved ensures that mythical creatures and humans can now coexist in harmony. The realm is forever changed, but the legacy of the *Moonstone Heart* endures, a symbol of the cost of peace.

Weaver : Chapter 1: The Rogue's Quest Begins

Sub-event: Jorik's Motivation and Introduction

Sub-event: The Mysterious Benefactor's Request

Sub-event: The Castle's Dangerous Secrets

Context: The first chapter sets up Jorik's character, his motivations, and the central quest. We meet the rogue in the bustling town, learn of his troubled past, and witness his skills. The mysterious benefactor introduces the central artifact, the *Moonstone Heart*, and sets the stage for the perilous journey ahead.

—
Chapter 2: The Shapeshifter's Deception

Sub-event: Virelith's Deceptive Entrance

Sub-event: Alliance Amid Suspicion

Sub-event: Betrayal and the True Purpose of the Artifact

Context: This chapter introduces Virelith, whose true intentions are hidden behind the mask of a helpful ally. The dynamic between her and Jorik builds tension as their uneasy alliance grows. The betrayal and reveal of Virelith's real goal push the plot toward its darker, more dangerous turns.

— Chapter 3: Despair and Unexpected Alliances Sub-event: Jorik’s Despair and the Unexpected Ally Sub-event: Jorik’s Internal Struggle Sub-event: The Final Confrontation in the Sacred Grove Context: After the betrayal, Jorik’s internal conflict is explored. He encounters Elira, a centaur who offers a new perspective and an alliance against the growing threat. The chapter focuses on Jorik’s journey of rebuilding trust while facing rising tensions between mythical creatures and humans. — Chapter 4: The Artifact’s Power Revealed Sub-event: Divided Allegiances and Growing Tension Sub-event: The True Nature of the Artifact Revealed Sub-event: A Choice Between Destruction and Evolution Context: In the climactic confrontation with Virelith, the artifact’s true nature is unveiled. This chapter serves as a turning point in the story as Jorik must decide whether to destroy the artifact or embrace its uncertain future. The moral and ethical stakes reach their highest point. — Chapter 5: The Fragile Path to Peace Sub-event: Jorik’s Proposal for Peace Sub-event: Resistance and Doubt Among the Factions Sub-event: A New Enemy Emerges Context: The aftermath of the artifact’s revelation leads to Jorik’s proposal for peace between humans and mythical creatures. However, the process is fraught with resistance and internal conflict. Just as things seem to settle, a new threat emerges, forcing Jorik and his allies to act quickly. — Chapter 6: The Final Battle and the Legacy of the Moonstone Heart Sub-event: The Rogue Faction’s Attack Sub-event: Jorik’s Ultimate Sacrifice Sub-event: A New Era of Peace Context: The final battle unfolds as a rogue faction threatens the fragile peace. Jorik’s ultimate sacrifice to destroy the artifact concludes the story, ensuring peace but at great personal cost. The chapter wraps up with the realm forever changed, and the legacy of the *Moonstone Heart* enduring as a symbol of the sacrifices made for peace.
--

Table 5: An example result of NLN. In this case, "SubTaker" assigns each event to more detailed independent sub-events. Because of our event-based approach, these sub-events also inherit the information and relationships of the parent event. "Weaver" finds that subevent-3.3 can last until event 4 occurs, so it reallocates the subevent, that is, subevent-3.3 is exchanged with subevent-4.1.

Re-Input

.....(PREVIOUS)

FinalWriter

: In the days that followed, Kael and Luna found their way back to the cavern that had changed his life forever. The snowy landscape had become familiar, yet Kael's heart raced with anticipation and the weight of responsibility. He had come to realize that the ancient artifact held the key to unlocking his true potential as a mage, and he could feel its call echoing in the depths of his soul.

As they entered the cave, the air was charged with a palpable energy. The sunlight streamed through the crystals embedded in the cave walls, creating a mesmerizing display of colors that danced like spirits in the darkness. The artifact rested on a stone pedestal, shimmering with a mystic light that beckoned him closer.

Luna padded silently beside him, her fur glistening in the cave's ethereal glow. She seemed to sense the significance of this moment too, her ears perked forward, eyes keen and alert. Kael reached out, his fingers brushing against the smooth surface of the orb. Instantly, images flooded his mind—visions of mages harnessing great power, wielding magic for both creation and destruction.

With resolve, Kael inhaled deeply, reminding himself of the bond he shared with Luna and the strength they had gained together. This was a pivotal moment, and he knew he needed to embrace it fully. He steadied himself, focusing on the energies swirling around him and the warmth the artifact emitted.

The moment his hands clasped the orb, a surge of magic coursed through him like lightning. It was unlike anything he had ever experienced—pure, exhilarating power merging with his very essence. He felt the weight of the world lift from his shoulders as he began to awaken the magic within him, unraveling the potential he had been so afraid to unleash.

Yet, as the energy intensified, Kael suddenly felt a tugging sensation—a dark shadow lurking at the edge of his consciousness. The power was formidable, and with it came a haunting thought: was he truly ready to wield such magic? Would he be able to resist the allure of darkness if faced with the temptation?

Coordinator : Next sub-story comes to(with history-message changed as follows)

History-Messages:

{ 'content': "With the artifact's power surging, Kael transformed, awakening abilities he never knew he had. His appearance changed, reflecting the magic within him as he realized he was becoming a protector, ready to use this newfound power for his village's safety. Luna stood by him, providing unwavering support as a dark presence emerged— a malevolent sorcerer intent on seizing the artifact's magic for himself. Feeling the weight of this looming threat, Kael understood he had to prepare for a confrontation with the sorcerer. Determined to control his new abilities and balance the light and dark within him, Kael, with Luna at his side, left the cavern. They were ready to face the trials ahead, confident that together they could confront any darkness threatening their world.", 'role': 'user', 'name': 'FinalWriter' }

Table 6: An example prompt for Open IE to generate task description and output format description.

.....(PREVIOUS)

FinalWriter :

In the wake of the tentative truce, Razor made his demands clear. He was no stranger to power plays, and he intended to use this moment to reclaim control over his territory. During a follow-up meeting orchestrated by Father Gregory, Razor leaned forward, his tone measured but resolute. "I'll agree to your community initiatives and support your mentoring programs, but in exchange, I need control over the local drug trade. That's the only way I can guarantee the safety of my people. It's how I ensure resources come in—money that can help us all," he stated, showcasing his unwavering belief that the drug trade was a necessary evil.

Tasha felt a surge of frustration. "Razor, we can't build a future on the backs of our neighbors. Sure, the money flows, but it also brings violence, fear, and more suffering. That's not what this community deserves," she implored passionately, her voice steady but laced with urgency.

Razor smirked, shaking his head as if she were naïve. "You think your ideals can change the world. I'm in the real world, Tasha. Families are hungry. The kids are trapped with nowhere to turn. They want power and respect—what I can provide. You can't just erase the reality of survival."

The room buzzed with tension. Father Gregory attempted to mediate, sensing that both were at a critical impasse. "Razor, perhaps there's a middle ground," he suggested, but Tasha was quick to interject.

"There is no middle ground when you're advocating for further destruction. I will not support a criminal approach to community development," she asserted, her conviction palpable. "Meaningful change requires us to move away from illegal activities. We can create jobs, build businesses, and invest in our youth through legitimate means."

Razor narrowed his eyes, the anger bubbling beneath the surface. "And what happens if your way doesn't work? If we don't act now, more will suffer. I'm trying to protect my own."

"By perpetuating a cycle of violence?" Tasha challenged. "Look deeper, Razor. Do you understand what this is doing to our community? Do you care? The people are tired of fear. They want hope."

In that moment, Razor's fierce demeanor softened ever so slightly. He glanced around the room, measuring the expressions of those gathered. Many, including former gang members who had turned a new leaf, started to murmur in agreement with Tasha's vision. It struck a chord that had long been buried beneath the chaos of the streets.

Yet Razor wasn't ready to concede control. "You're all dreaming. I have spent years building an empire. You can't just ask me to sacrifice everything without a viable alternative. Show me how your way will work, then we can talk."

Tasha took a deep breath, knowing this was a pivotal moment. "We can start small. Let's invest in youth programs, job training, and small business grants. Work together to fund initiatives that can uplift the community without violence. I'll bring legitimate businesses in, and you can help steer that profit into community gains. But we do it openly and with integrity. This is about empowerment, not dominance."

Razor stared at her, weighing his options. For so long, he had ruled through fear, convinced it was the only way. But now he faced a community unwilling to accept his old methods, and the prospect of losing respect was frightening.

The standoff highlighted the stark realities of their opposing visions for the future, and Razor realized that without adapting to the new demands of his people, he would risk alienating everyone, including those he sought to protect.

"At the very least, I'll consider your proposal, Tasha," Razor finally said, his voice low but steady. "But like I said, if this doesn't strengthen my position, there will be consequences."

Tasha understood the fragility of this moment, but it was a step—a precarious but crucial step towards a newfound possibility for coexistence built on mutual benefit instead of mutual destruction. The fight for peace was far from over, and the battle lines continued to shift as they each sought to define their legacies amidst turbulent change.

Coordinator : **RE-WRITE**: Razor’s demand for control over the drug trade in exchange for peace was a pivotal moment that threatened to derail the fragile truce. Tasha, fueled by her vision for a better future, stood firm against his proposition. “Razor, we can’t build a future on the backs of our neighbors. The drug trade brings violence and fear, and that’s not what this community deserves,” she insisted, her voice steady but urgent. Razor, however, was unyielding. “You think your ideals can change the world. I’m in the real world, Tasha. Families are hungry. They want power and respect—what I can provide. You can’t just erase the reality of survival.” Tasha’s frustration mounted as she countered, “Meaningful change requires us to move away from illegal activities. We can create jobs and invest in our youth through legitimate means.” The tension in the room escalated, with Razor’s gang members watching closely. Father Gregory attempted to mediate, sensing the critical impasse. “Perhaps there’s a middle ground,” he suggested, but Tasha was quick to interject. “There is no middle ground when you’re advocating for further destruction. I will not support a criminal approach to community development,” she asserted. “We can uplift the community without violence.” Razor’s demeanor softened slightly as he glanced around the room, noticing the murmurs of agreement with Tasha’s vision. Yet he remained hesitant to concede control. “You can’t just ask me to sacrifice everything without a viable alternative. Show me how your way will work, then we can talk.” Tasha took a deep breath, knowing this was a pivotal moment. “Let’s invest in youth programs, job training, and small business grants. Work together to uplift the community without violence. I’ll bring legitimate businesses in, and you can help steer that profit into community gains.” Razor stared at her, weighing his options. For so long, he had ruled through fear, convinced it was the only way. But now he faced a community unwilling to accept his old methods. “At the very least, I’ll consider your proposal, Tasha,” he finally said, his voice low but steady. “But if this doesn’t strengthen my position, there will be consequences.” Tasha understood the fragility of this moment, but it was a step—a precarious but crucial step towards a newfound possibility for coexistence built on mutual benefit instead of mutual destruction. The fight for peace was far from over, and the battle lines continued to shift as they each sought to define their legacies amidst turbulent change. Next sub-event:A betrayal by one of Razor’s gang members forces him to reconsider his approach, realizing that his violent methods are leading to his downfall."

Table 7: An example result of reoutput.