SeqUAL: Sequel story Unfolding and Augmentation with LLMs

Anonymous ACL submission

Abstract

Generating sequels that maintain a cohesive worldview while introducing fresh narratives is a complex task requiring a delicate balance between consistency and diversity. Despite the growing demand for narrative expansion, existing story generation research often focuses on one at the expense of the other. Consistency-focused methods frequently fail to introduce the new characters and societal shifts essential for a sequel, while diversity-focused approaches often sacrifice the historical and structural integrity of the original world. In this study, we propose SeqUAL (Sequel Unfolding and Augmentation with LLM), a framework designed to bridge this gap by leveraging world graphs and character personality traits. SeqUAL ensures consistency by automatically reconstructing the original story's societal structures and historical context through a world graph, while fostering diversity by simulating interactions among new characters defined by distinct personalities. We validated our framework using a newly curated dataset of 18 recent Sci-Fi novels in English and Korean, generating 20 sequels per novel. Our experimental results demonstrate that SeqUAL significantly outperforms existing baselines in both diversity and consistency. This work establishes sequel generation as a distinct subtask in automated story generation and provides the first practical framework for creating diverse yet world-consistent narrative expansions.

1 Introduction

There has been growing demand for narratives that preserve the original worldview, as seen in applications such as expanding movie franchises into games or adapting game universes into animations[1]. With the rapid advancement of Large Language Models (LLMs), research on automatic story generation has been actively progressing (Venkatraman et al., 2025; Dzhubaeva et al., 2025; Xie et al., 2023; Wen et al., 2023). Most prior studies on story generation have primarily focused on maintaining narrative consistency, while relatively overlooking the diversity of generated stories (Wang et al., 2025a; Mu and Li, 2024; Ma et al., 2023; Yi et al., 2025). On the other hand, some researchers have begun to generate diverse derivative works, while focusing less on consistency (Xu et al., 2025; Uduehi et al., 2024). However, both consistency and diversity are essential for derivative works.

Specifically, we focus on a type of derivative work that has received little attention in prior research: sequels. Sequels have different properties compared to other derivative works; they should maintain a worldview *consistent* with the original story while producing distinct and *diverse* story plots (Migal et al., 2024; Xu et al., 2025). Therefore, it remains questionable whether those existing methods can be applied to sequels.

First, existing consistency-focused methods attempted to tightly align the generated story with the events in the original work (Han et al., 2025; Rashkin et al., 2020). For example, Han et al. (2025) has focused on re-generating stories by following similar patterns of the original. However, such a method cannot produce a good sequel. A sequel is a new story of other characters after the last event of the original. So, a good sequel should introduce new characters, leading to changes in social relationships, societal structures, and events. Hence, the plot of a sequel differs from the original one. As current consistency-focused methods attempt to replicate the original, generating sequels could be challenging for them.

Second, existing diversity-focused methods attempted to achieve character diversity (Ran et al., 2025; Park et al., 2025a). For example, Ran et al. (2025) generated diverse stories form a simulation using original characters. However, such a method cannot produce a good sequel, either. As those

[1]For example, https://www.harrypotter.com/, https://store.epicgames.com/ko/p/hogwarts-legacy.

methods have only examined whether original characters fit their background, their method cannot ensure that their historical or societal relationship is preserved. In a sequel, such a relationship should be preserved, because it shares the same fictional history with the original.

Thus, we set a dialectic goal to embrace both approaches; this paper aims to generate diverse sequels consistent with the original story. We propose SeqUAL (Sequel Unfolding and Augmentation with LLM), which leverages the *world graph* and *personality traits* of new characters. To maintain consistency, SeqUAL automatically constructs a world graph of the original work. To encourage diversity, SeqUAL defines the new characters based on their personality and simulates the interaction between characters. Based on the world graph and simulation records, our SeqUAL can successfully generate a wide range of diverse stories while maintaining consistency with the original world.

We conducted an experiment to validate the effectiveness of SeqUAL. Specifically, we collect 18 recent Sci-Fi novels, either written in English or Korean, and ask SeqUAL to generate 20 sequels per novel. By measuring the quality, consistency, and diversity of generated sequels, this paper shows the following contributions.

- We propose sequel generation as a new subtask, and present SeqUAL, the first framework to embrace both diversity and consistency.
- We show that the world graph ensures consistency by recovering societal structures, while both graph and personality contribute to different aspects of diversity.

2 Related Work

Here, we review previous work regarding two aspects: consistency and diversity. Story generation research has focused primarily on either aspect, because these two aspects have a trade-off relationship in general (Migal et al., 2024; Xu et al., 2025). Despite such a trade-off, reviewing each line of research can provide a possible way to balance.

Regarding consistency, existing studies have mainly focused on plot-level consistency rather than worldview-level consistency. Some researchers attempted to maintain plot-level consistency through predefined outlines or plans (Wang et al., 2025a; Mu and Li, 2024; Ma et al., 2023; Yang et al., 2023; Qi et al., 2024). For example, Yang et al. (2023) utilized a detailed outline to ensure plot consistency. Similarly, Qi et al. (2024) proposed a framework that learns to select story-generation prompts through reinforcement learning to maintain narrative consistency. Other researchers decomposed story generation into multiple structured stages to maintain plot-level consistency (Xie et al., 2025). In the work, they used a top-down refinement process to establish a consistent plot. Despite their success in maintaining plot-level consistency, however, the direct application of them to generating a sequel is inadequate. As their main goal was replicating the ability of generating a story, they overlooked the significance of worldview[2] in generating a series of stories. Though in generating sequels, such worldview-level consistency is essential as a worldview maintains consistency across a series of stories. Thus, we need a way to maintain the worldview in sequel generation.

Regarding diversity, existing studies mainly focused on reusing original event structures instead of introducing a novle one. Some studies have attempted to enhance story diversity through post-editing (Bae and Kim, 2024; Zhu et al., 2023). For example, Zhu et al. (2023) has employed style-transfer techniques that alter the writing style to a specific author. Other researchers have attempted to achieve diversity with constraints (Wang and Kreminski, 2024; Park et al., 2025b). For example, Wang and Kreminski (2024) proposed a rule-based constraint model using answer set programming that guides an LLM to generate diverse outlines. Similarly, Park et al. (2025b) introduced a method that uses previously generated stories as negative samples, imposing penalty constraints to discourage the generation of similar stories. Despite the success of diversifying the stories, however, they cannot be applied to the sequel generation. Such post-editing or constraint-based methods are less suitable for introducing new characters. Post-editing limits behavior of the new character to a set of behaviors already described in the original. Similarly, a constraint prohibits generating events similar to the original one, regardless of their commonality. Furthermore, these approaches seldom receive new characters as their input, which is required in sequel generation. Thus, we need a new method to diversify story based on new characters.

[2] A worldview includes social structures, historical context, and relational networks among characters.

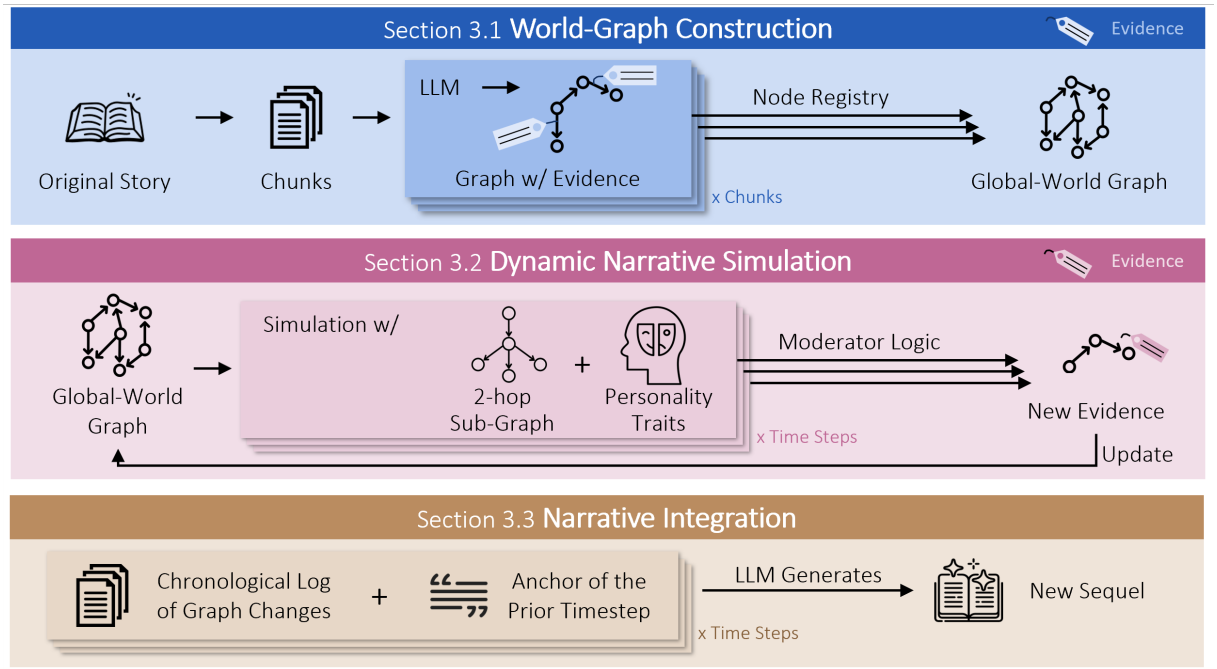

Figure 1: Overview of the SeqUAL Framework

3 The SeqUAL Framework

As shown in Figure 1, SeqUAL takes an original story as input and unfolds diverse sequels through a structured simulation. It consists of three LLM-based modules. First, World-Graph Construction module builds a foundational knowledge graph by extracting entities and relations from the original story. Second, Dynamic Narrative Simulation module executes a multi-stage story unfolding, where characters interact based on their unique personality traits and restricted 2-hop knowledge. Third, Narrative Integration module analyzes the story evolution graph to generate a cohesive final sequel.

3.1 World-Graph Construction

First, We chunk the original story into 2000-character segments with 300-character overlaps to mitigate mid-context degradation.

To construct a world graph, SeqUAL transforms each chunk τ_i into a partial world graph $\mathcal{G}_i = (V_i, E_i, \tau_i)$. For the *nodes* $v \in V_i$, SeqUAL extracts four entity types: characters, locations, organizations, and concepts. For the *edges* $e \in E_i$, SeqUAL extracts directed relationships between entities to map social and physical landscapes. To enable citations of the original work, we labeled each edge e with the corresponding sentence $s \in \tau_i$ as an *evidence*. We used an LLM to extract nodes and edges and resolve coreferences based on contexts. The detailed prompts are provided in Appendix A.1.

Whenever extracting a partial world graph $\mathcal{G}_i$, SeqUAL integrates the graph into the previously accumulated world graph $\mathcal{G}_{<i}$. We designed a Node Registry component that identifies node pairs between $\mathcal{G}_i$ and $\mathcal{G}_{<i}$ by computing semantic similarity between the nodes. For a node pair, we unite them two nodes if the cosine similarity between their names exceeds 0.9. This procedure ensures that nodes in the accumulated graphs are unique. We also united edges and evidence into $\mathcal{G}_{<i}$, considering newly united nodes. For Node Registry, we used `sentence-transformer` `all-MiniLM-L6-v2`).

SeqUAL pools all chunks into a global graph $\mathcal{G} = (V, E, \cup\tau_i)$. This represents the complete unified environment and serves as the ground truth for subsequent simulations, to ensure consistency.

3.2 Dynamic Narrative Simulation

Next, SeqUAL simulates the new story. It executes a story unfolding process with multiple stages: Opportunity, Change of plan, Point of no return, Major setback, Climax, and Ending (Papalampidi et al., 2019). We borrowed these six stages to ensure plot-level consistency during sequel generation. For each stage, the model iteratively simulates new interactions and events and identifies updates to the world. Then, it accumulates such updates on the global world graph to ensure the evolving narrative state is consistently integrated into the foundational

worldview. In the following paragraphs, we detail the per-stage procedure.

To create realistic simulation, we restrict each character's perception to a two-hop subgraph centered on their identity node, mirroring how real individuals operate under partial information.

To further diversify the story, SeqUAL adopts personality traits. One can input personality traits of the newly introduced protagonist using the Big Five personality framework (John et al., 1999). Using the personality profile, the simulation considers character internalities and governs the protagonist's behavior. Thus, we can ensure high diversity of story plots and variability in personality profiles. Detailed descriptions of the personality configurations are provided in the Appendix A.2.

After per-stage generation, SeqUAL updates the world graph to maintain consistency. Specifically, we used a moderator logic. The moderator reviews proposed updates in JSON format, filters out contradictions, and actively "highlights" newly validated facts and relationships within the global graph. By labeling these "highlighted" elements, the moderator effectively distinguishes established narrative truths from background world data. This ensures the updated graph serves as a coherent, grounded foundation for the subsequent simulation steps, blocking any future interactions that might violate these highlighted facts.

3.3 Narrative Integration

Lastly, SeqUAL generates narratives based on the simulation. Instead of simply concatenating events, it comprehensively analyzes chronological changes in the world graph. SeqUAL tracks the changes of nodes V and edges E across the six simulation stages, and examines how the protagonist's personality profile triggers significant shifts in the world state; for example, forming new alliances.

To bridge the gap between stages, SeqUAL uses the last sentence of the previous stage as a narrative anchor, with the original story's final sentence serving as the anchor for the first stage. Finally, SeqUAL generates a comprehensive sequel maintaining worldview consistency using an LLM. The detailed prompts are provided in Appendix A.3.

4 Experiments

We conduct an experiment to examine the *consistency* and *diversity* of generated sequels by SeqUAL. In addition to these aspects we considered, we also evaluated the *quality* of the generated sequels. In following paragraphs, we describe datasets, analyses, models, and implementations. Implementation details are in Appendix B.

4.1 Dataset

We manually constructed datasets of English and Korean Sci-Fi novels. We have two reasons for selecting these novels as testing datasets. First, for consistency, we wanted to discriminate an accidental alignment from real consistency. As Sci-Fi novels introduce distinctive and unconventional worlds, they provide a clear distinction from real world, leading to less accidental usage of terms specific to the corresponding world. So, we excluded usage of novels with other genres. Second, for diversity, we wanted to evaluate true diversity independent of language or other parametric knowledge. We chose to use novels recently written and adopt two languages, enhancing generalizability of the results.

Accordingly, we selected 18 Sci-Fi novels published after 2022. To ensure the quality of novels, we followed a similar approach to Jang and Jung (2024) and selected 9 novels for each language, which have been awarded the most famous awards in each language: Hugo awards and Korean Science Fiction Awards. We used the published versions without any modifications[3]. Appendix C details the list of selected works. The average number of sentences per English work was 387.78 (min 46, max 757), and the average number of sentences per Korean work was 847.44 (min 472, max 1,370).

4.2 Analysis method

We measure the performance of generating sequel regarding three aspects: diversity, consistency, and quality. In general, we generated 20 sequels for each story, regardless of sequel generation model. Then, we report averaged results.

Consistency κ. We employ three metrics that measure the alignment between worldview and sequel: consistency on character κ_c, event κ_e, and worldview κ_w. κ_c measures the persistence of an original character's core persona, encompassing their fundamental traumas, worldview, and established interpersonal dynamics. κ_e assesses the continuity of key events, which are defined as pivotal incidents that drive the original plot and result in irreversible causal consequences. And, κ_w analyzes

[3]Due to copyright restrictions, the exact contents of the novels cannot be disclosed.

	Consistency κ (%; ↑)				Diversity δ (%; ↑)			
	Character	Event	World	*Avg.*	Character	Event	World	*Avg.*
Vanilla (Gemma-3)	76.22	67.61	77.00	73.61	71.93	**90.77**	60.16	74.29
(GPT-oss)	72.89	59.06	67.56	66.50	**72.65**	83.19	44.76	66.87
LongWriter	77.11	70.50	78.06	75.22	65.70	76.44	56.95	66.36
SeqUAL (Gemma-3)	87.72	**93.28**	**94.78**	**91.93**	69.34	87.55	**82.87**	**79.92**
(GPT-oss)	**87.94**	86.28	93.44	89.22	53.17	79.51	67.37	66.68

Table 1: Comparison between SeqUAL and other baselines regarding consistency and diversity

adherence to world logic, referring to the unique institutional or environmental characteristics.

We evaluated consistency based on constraints extracted from the original story. For each attribute $a \in \{c, e, w\}$, we generated a constraint set $C_a(\tau)$ from original story τ using Gemini-2.5-Flash (Comanici et al., 2025). Specific prompts used for this evaluation are detailed in Appendix D.1. We then queried the same model to classify whether generated sequel $s_i(\tau)$ is consistent with $C_a(\tau)$ into two categories: *violation*, or *non-violation*.

$$\kappa_a = 1 - \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}\left[s_i(\tau) \text{ violates } C_a(\tau)\right], \quad (1)$$

where a indicates attributes and $\mathbb{I}(\cdot)$ is an indicator function. We set $|C_a(\tau)|$ and N as 5 and 20. Note that a high κ_a indicates high consistency as we measure the proportion of non-violations.

Diversity. We employ three metrics that measure overlap among the 20 sequels generated from the same original story: diversity of character δ_c, event δ_e, and worldview δ_w. δ_c measures how diverse the set of key characters interacting with the new protagonist is. δ_e and δ_w measure how diverse the main event and worldview differ across sequels.

We measure diversity based on the set of key elements for each attribute. For each aspect a, we extracted a sentence $v_{a,i}$ describing key elements from each sequel s_i using Gemini-2.5-Flash[4]. By computing TF-IDF on sequels from the same original work, we transform $v_{a,i}$ into a vector $\mathbf{v}_{a,i}$. We then compute cosine similarity.

$$\delta_a = 1 - \frac{1}{N'} \sum_{i} \sum_{j \neq i} \cos(\mathbf{v}_{a,i}, \mathbf{v}_{a,j}), \quad (2)$$

[4]For character, we concatenated three key characters other than the new protagonist to form $v_{c,i}$.

where N is set to $\binom{20}{2} = 190$. Note that high δ_a indicates low diversity, as δ_a measures the additive inverse of similarity between key elements.

Quality. Additionally, we employ three metrics to evaluate the quality of the generated stories, following the evaluation prompts for Sci-Fi quality proposed in previous research (Wang et al., 2025b). The assessment focuses on plot structure, grammaticality, and fluency. We used the same prompts used in the work, as shown in the Appendix D.3. Gemini-2.5-Flash is used as a primary evaluator.

Notes. Our evaluation only requires text-based extraction and classification, where LLMs usually outperform humans and perform reliably, so we employ LLM-as-a-judge for automatic evaluation(Lee et al., 2025). We designed our evaluation prompts as deterministically and objectively as possible to avoid bias, and excluded our evaluator model, Gemini, from the backbone of the sequel generation model to avoid LLM self-preference bias.

4.3 Baseline models

We compared SeqUAL with vanilla baselines and LongWriter (Bai et al., 2025). Specific prompts for these models are detailed in Appendix E. First, we selected vanilla LLMs to measure their inherent story generation capability. Specifically, we evaluated GPT-oss-120B (OpenAI et al., 2025) and Gemma-3-27B Instruct (Team et al., 2025), tasking them with generating sequels from full stories.

Second, we selected LongWriter, a story generation model based on LLaMA 3.1 8B (Grattafiori et al., 2024). As sequel generation requires the original story, publicly available models usually require modification of input or their computation architecture. To reduce the confounding effect of our modification, we selected LongWriter, which can receive the original story without modifying the story generation model.

4.4 SeqUAL and its ablated models

For SeqUAL, we implemented it using the same two LLMs: GPT-oss-120B and Gemma-3-27B Instruct[5]. For the ablation, we used GPT-oss-120B, which was selected a priori to avoid potential bias.

To examine the effect of the global world graph and personality traits, we prepared two ablated models of SeqUAL. The first is SeqUAL without a graph. This model generates a sequel directly from the original story, augmented with a new character possessing Big Five personality traits. At each simulation step, the model takes the entire original story and the story from the previous simulation step as input. This enables us to evaluate the effect of world graph construction by comparing it with SeqUAL. The second ablated model is SeqUAL without personality traits. This model generates a sequel with a world graph but without Big Five personality traits. This allows us to observe the impact of the graph by comparing it with SeqUAL.

5 Result and Discussion

5.1 Diversity and Consistency

Experimental result in Table 1 indicates that SeqUAL substantially improved consistency while preserving competitive diversity. Different from trade-off in existing models, our model achieved both high performance in consistency and diversity. We discuss the results regarding both aspects. To ensure a fair comparison and decouple model scale effects, we evaluate both our framework and the baseline using identical backbone models (e.g., Vanilla Gemma-3). The performance gap between SeqUAL (Gemma-3) and Vanilla (Gemma-3) confirms that our framework's superiority stems from architectural design, not parameter scale.

First, we observed a notable improvement in consistency metrics. SeqUAL (Gemma-3) achieved average consistency 91.93%, beating LongWriter (75.22%) by 16%. In comparison, vanilla LLMs (Gemma-3 and GPT-oss) recorded average consistency of 73.61% and 66.50%, respectively. A similar trend can be observed in the individual consistency categories. Specifically, regarding character, SeqUAL (Gemma-3) beat all baselines by 10%: 87.72% versus 77.11%. This validates our architectural design, as the multi-level verification framework incorporating chunking, evidence labeling, and moderator logic successfully prevents logical hallucinations, enabling SeqUAL to retain historical events and core relational networks without factual drift. The gap becomes larger in event and world; for example, regarding event, SeqUAL showed 93.28% consistency, which is 22% larger than the highest performing baseline, LongWriter (70.50%). These results suggest that SeqUAL more effectively preserves characters, events, and worldview of the original story while generating new sequel narratives.

[5]For brevity, we will call two LLMs as GPT-oss and Gemma-3, respectively.

	Quality (↑)		
	Plot	Grammar	Fluency
Vanilla (Gemma-3)	3.05	4.44	4.28
(GPT-oss)	**3.34**	**4.60**	**4.34**
LongWriter	2.06	3.60	3.47
SeqUAL (Gemma-3)	2.57	4.35	4.10
(GPT-oss)	2.51	4.21	3.98

Table 2: Comparison regarding quality of sequel

Second, the improvement in consistency did not come at the cost of reduced diversity. In the overall diversity metric, SeqUAL (Gemma-3) achieved an average diversity of 79.92%, which is higher than the corresponding vanilla LLM (74.29%). This indicates the SeqUAL improves generation diversity in general. More specifically, regarding worldview, SeqUAL (Gemma-3) records 82.87%, substantially higher than vanilla counterpart (60.16%) and LongWriter (56.95%). Regarding character, SeqUAL (Gemma-3) achieved 69.34%, which is comparable to the diversity of other models (Gemma-3 71.93% or LongWriter 65.70%). Similarly, regarding event, SeqUAL (Gemma) recorded 87.55%, maintaining a diversity level comparable to its vanilla counterpart (90.77%). These results indicate that SeqUAL substantially improves consistency in sequel generation while maintaining comparable diversity, enabling the generation of coherent narratives that extend the original story into new contexts.

The results demonstrated that our proposed method helps generate a sequel that is highly consistent with the original story and diverse enough. We suspect that the world graph and the personality traits helped maintain both consistency and diversity. Before identifying the contribution of these two ideas, we first verify the quality of sequels.

5.2 Quality of Sequels

Table 2 presents the qualitative evaluation of the generated stories in terms of plot structure, grammaticality, and fluency. SeqUAL demonstrated competitive performance, achieving high scores around 4.0 across all metrics, with no statistically significant differences from vanilla counterparts. The complete results including statistical deviations and standard errors are provided in Appendix F.

Notably, SeqUAL outperformed LongWriter across all quality metrics. For instance, SeqUAL (Gemma-3) achieved a higher grammar score than LongWriter (4.35 vs. 3.60). These results suggest that SeqUAL maintains grammatical accuracy and fluency despite the structural constraints required for narrative consistency, enabling robust story generation without significantly sacrificing text quality.

5.3 Ablation Study

Results from the ablation study are shown in Table 3. As we focused on the role of world graph and personality traits on consistency and diversity, this section discusses the effect of those components. The result suggests a complementary relationship between world graph and personality traits.

First, the world graph is needed to maintain consistency. Compared to SeqUAL, the model without graph recorded lower consistency: 81.44% for character, 79.50% for event, and 84.06% for world. Although performance remains higher than its vanilla counterpart (GPT-oss), the ablated model's consistency is 7.55% lower than the full model, on average. That is, a constructing world graph can further contribute to maintaining a consistent worldview in the original work, even when we provide the full text of original work as an input. Thus, SeqUAL without personality traits showed consistency comparable to the full model, with at most a 2% drop.

Second, world graph and personality trait both enhance diversity in different ways. On average, SeqUAL without personality traits achieved diversity comparable to that of the full model: 65.37% versus 66.68%. However, this does not indicate that personality trait is meaningless. When we measure diversity of character, removing personality traits results in a 3.4% decrease. On the other hand, SeqUAL without graph achieved higher diversity than the full model, with a 2.22% increase. These contrastive results imply that personality traits help generate diverse characters even when the graph suppresses character diversity. Also, the result implies that LLMs may fail to define distinct personality traits for new protagonists; instead, they replicate existing characters and their relationships.

Meanwhile, SeqUAL without a graph leads to a substantial decrease in worldview.Regarding diversity of worldview, SeqUAL recorded 67.37%, whereas the model without graph reached 51.96%, which is 15% lower. Although the diversity of worldview is still higher than the vanilla model (44.76%), the difference between vanilla and the ablated model is smaller than that between SeqUAL and the ablated model. This confirms that leveraging the graph effectively preserves worldview-related features while aiding diversity. As providing an original story as input makes LLMs memorize and utilize all related facts about the world, we suspect that graph-based summarization helps them to balance between consistency and diversity.

Additionally, we conducted an emotion-level analysis of the generated sequels using bert-base-uncased-emotion model[6], examining whether personality traits lead to broader emotional ranges across stories. The results confirm that personality traits meaningfully diversify both the narrative opening and resolution styles of generated sequels, as detailed in Appendix G.

We suspect that by leveraging each character's 2-hop information, SeqUAL can maintain high consistency while still generating diverse interactions and conflict situations. Previous approaches, such as DOC (Yang et al., 2023), give the LLM access to the entire story context when generating outlines. However, such an omniscient perspective, leads to strong consistency with the cost of low diversity. Meanwhile, SeqUAL uses localized context for each character with two-hop subgraphs. Such a limited perspective mitigates the pitfalls of an omniscient viewpoint, in which the model might overly rely on global information rather than producing grounded, character-specific developments. Taken together, the world graph in SeqUAL preserves consistency and encourages diverse interactions, while personality traits further enrich diversity at the character and emotional level.

5.4 Qualitative examples

We further examine the practical performance of the models through representative examples. Due to the page limit, full examples are provided in

[6] https://huggingface.co/bhadresh-savani/bert-base-uncased-emotion

	Consistency κ **(%; ↑)**				**Diversity** δ **(%; ↑)**			
	Character	Event	World	*Avg.*	Character	Event	World	*Avg.*
Vanilla (GPT-oss)	72.89	59.06	67.56	66.50	72.65	83.19	44.76	66.87
SeqUAL (GPT-oss)	87.94	86.28	93.44	89.22	53.17	79.51	67.37	66.68
w/o graph	81.44	79.50	84.06	81.67	55.39	80.15	51.96	62.50
	-6.50	*-6.78*	*-9.38*	*-7.55*	*+2.22*	*+0.64*	*-15.41*	*-4.18*
w/o personality traits	85.89	85.00	93.28	88.06	49.76	78.51	67.83	65.37
	-2.06	*-1.28*	*-0.16*	*-1.16*	*-3.41*	*-1.00*	*+0.46*	*-1.31*

Table 3: Result of ablation study regarding consistency and diversity

Appendix H. Instead, we discuss our two findings based on quotes from the examples.

First, regarding consistency evaluation, we assessed the following query: "Are the *trees* inside the dome still perceived as living beings possessing human memories and emotions?" This is a core premise of the original story that must not be violated. SeqUAL successfully preserved this element with the narration: "Just as *Semin had done*, I wished to become a *tree* as well." Conversely, the vanilla LLM violated this core trait by stating: "The fruit *tree* was the one Mr. Hyun had asked to be planted in memory of his mother," treating the tree as a conventional memorial rather than a transformed human entity. Similarly, LongWriter exhibited a significant limitation: "Is it really safe for the *trees* to be exposed to this rain?" which is a simple repetition of the original story verbatim, without any modification.

Second, in the diversity evaluation, we observed that SeqUAL generates vaired narrative trajectories based on assigned personality traits. Depending on the character profile, the model produces distinct outcomes: ranging from an avoidant response ("I simply wanted to *escape* this situation") to a proactive sense of agency ("He would *bring their story* to the world"). In contrast, both the vanilla model and baseline models tended to converge toward stories with highly similar atmospheres and outcomes, failing to achieve meaningful narrative variance. For example, LongWriter generated two similar sequels: "Hankyul came to gain a deeper understanding of their daily life" and "he felt that he was coming to understand the world."

6 Conclusion

This study attempted to address the problem of balancing diversity and consistency while generating sequels, which had not been sufficiently explored in prior research on story generation. We proposed the SeqUAL framework and introduced idea of utilizing world graph and personality traits. We used world graph for maintaining consistency within the plot of sequel and between the sequel and the original work. Also, we make SeqUAL adopts personality traits to enhance the diversity of characters and corresponding plots. Through world-graph construction, dynamic narrative simulation, and narrative integration, SeqUAL can generate a noble sequel consistent to the original work.

To validate the effectiveness of the proposed framework, we conducted an experiment using 18 Sci-Fi novels written in English or Korean. We compared SeqUAL with other models and analyzed the diversity, consistency, and quality of the generated sequels. The analysis of consistency demonstrated that our framework significantly outperformed the baseline models. Notably, consistency increased significantly without a loss in diversity. Furthermore, regarding quality scores, we confirmed that the stories generated by SeqUAL maintained high standards with no qualitative issues. We further confirmed the effect of using world graph and personality traits through an ablation study. Thus, we concluded that SeqUAL can generate diverse stories that do not violate the original story while based on the same original work.

Consequently, this study successfully demonstrated that by integrating the world graph and personality traits, one can build a sequel generation model that secures both consistency and diversity. We believe that ours can provide a foundation for further work on LLM-based story generation.

Limitations

Although we successfully designed SeqUAL for generating diverse and consistent sequels from the original work, this work has two limitations.

First, our evaluation does not fully assess the overall quality of the generated stories. Although we examined structural correctness, grammatical accuracy, and fluency, we did not measure higher-level qualitative aspects such as interestingness, engagement, or narrative immersion. This limitation arises because the primary objective of this study is not to optimize story quality itself, but to address the challenge of improving diversity while maintaining consistency in sequel generation. Therefore, a more comprehensive evaluation of subjective story quality is left for future work, after resolving the fundamental issues targeted in this research.

Second, our method does not obtain the best diversity scores in all fine-grained evaluation category. The results indicate that diversity is not uniform across all narrative aspects, and some subdimensions may require specialized modeling approaches. We expect that our analysis can serve as a reference for future work aiming to improve diversity in specific dimensions of story generation.

Furthermore, while our framework successfully balances narrative consistency and diversity, it prioritizes performance over computational efficiency. Optimizing the overall complexity and reducing operational costs remain important directions for future work. Additionally, our current evaluations are focused on limited datasets. To enhance domain scalability and generalization, extending this approach to broader literary genres and developing genre-specific graph extraction strategies will be addressed in future research.

Ethical Considerations

During the writing process of this paper, AI assistants supported our writing. We employed Grammarly for grammar checking. Also, we used GPT and Gemini for polishing our writing and enhancing the presentation clarity of our paper. These tools were used for editorial purposes only.

References

Minwook Bae and Hyounghun Kim. 2024. Collective critics for creative story generation. In Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, pages 18784–18819.

Yushi Bai, Jiajie Zhang, Xin Lv, Linzhi Zheng, Siqi Zhu, Lei Hou, Yuxiao Dong, Jie Tang, and Juanzi Li. 2025. Longwriter: Unleashing 10,000+ word generation from long context llms. In International Conference on Learning Representations, volume 2025, pages 36528–36546.

Gheorghe Comanici, Eric Bieber, Mike Schaekermann, Ice Pasupat, Noveen Sachdeva, Inderjit Dhillon, and 1 others. 2025. Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities. Preprint, arXiv:2507.06261.

Nellia Dzhubaeva, Katharina Trinley, and Laura Pissani. 2025. Unstructured minds, predictable machines: A comparative study of narrative cohesion in human and LLM stream-of-consciousness writing. In Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 4: Student Research Workshop), pages 1079–1096, Vienna, Austria. Association for Computational Linguistics.

Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, and 1 others. 2024. The llama 3 herd of models. Preprint, arXiv:2407.21783.

Xueran Han, Yuhan Liu, Mingzhe Li, Wei Liu, Sen Hu, Rui Yan, Zhiqiang Xu, and Xiuying Chen. 2025. Pastiche novel generation creating: Fan fiction you love in your favorite author's style. Preprint, arXiv:2502.15616.

Woori Jang and Seohyon Jung. 2024. Evaluating LLM performance in character analysis: A study of artificial beings in recent Korean science fiction. In Proceedings of the 4th International Conference on Natural Language Processing for Digital Humanities, pages 339–351, Miami, USA. Association for Computational Linguistics.

Oliver P John, Sanjay Srivastava, and 1 others. 1999. The big-five trait taxonomy: History, measurement, and theoretical perspectives.

Yukyung Lee, JoongHoon Kim, Jaehee Kim, Hyowon Cho, Jaewook Kang, Pilsung Kang, and Najoung Kim. 2025. CheckEval: A reliable LLM-as-a-judge framework for evaluating text generation using checklists. In Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing, pages 15771–15798, Suzhou, China. Association for Computational Linguistics.

Congda Ma, Kotaro Funakoshi, Kiyoaki Shirai, and Manabu Okumura. 2023. Coherent story generation with structured knowledge. In Proceedings of the 14th International Conference on Recent Advances in Natural Language Processing, pages 681–690, Varna, Bulgaria. INCOMA Ltd., Shoumen, Bulgaria.

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 Proceedings of the 17th International Natural Language Generation Conference: Generation Challenges, pages 47—-53, Tokyo, Japan. Association for Computational Linguistics.

Feiteng Mu and Wenjie Li. 2024. Generating contrastive narratives using the brownian bridge process for narrative coherence learning. In Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 6538–6555, Bangkok, Thailand. Association for Computational Linguistics.

OpenAI, Sandhini Agarwal, Lama Ahmad, Jason Ai, Sam Altman, Andy Applebaum, Edwin Arbus, Rahul K. Arora, Yu Bai, Bowen Baker, Haiming Bao, Boaz Barak, Ally Bennett, Tyler Bertao, Nivedita Brett, Eugene Brevdo, Greg Brockman, Sebastien Bubeck, Che Chang, and 106 others. 2025. gpt-oss-120b gpt-oss-20b model card. Preprint, arXiv:2508.10925.

Pinelopi Papalampidi, Frank Keller, and Mirella Lapata. 2019. Movie plot analysis via turning point identification. 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 1707–1717, Hong Kong, China. Association for Computational Linguistics.

Kyeongman Park, Minbeom Kim, and Kyomin Jung. 2025a. A character-centric creative story generation via imagination. In Findings of the Association for Computational Linguistics: ACL 2025, pages 1598–1645, Vienna, Austria. Association for Computational Linguistics.

Kyeongman Park, Nakyeong Yang, and Kyomin Jung. 2025b. Avoidance decoding for diverse multi-branch story generation. In Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing, pages 7500–7516.

Qianqian Qi, Lin Ni, Zhongsheng Wang, Libo Zhang, Jiamou Liu, and Michael Witbrock. 2024. Epic-level text generation with llm through auto-prompted reinforcement learning. In 2024 International Joint Conference on Neural Networks (IJCNN), pages 1–8.

Yiting Ran, Xintao Wang, Tian Qiu, Jiaqing Liang, Yanghua Xiao, and Deqing Yang. 2025. BOOKWORLD: From novels to interactive agent societies for story creation. In Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 15898–15912, Vienna, Austria. Association for Computational Linguistics.

Hannah Rashkin, Asli Celikyilmaz, Yejin Choi, and Jianfeng Gao. 2020. PlotMachines: Outline-conditioned generation with dynamic plot state tracking. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 4274–4295, Online. Association for Computational Linguistics.

Gemma Team, Aishwarya Kamath, Johan Ferret, Shreya Pathak, Nino Vieillard, Ramona Merhej, Sarah Perrin, Tatiana Matejovicova, Alexandre Ramé, Morgane Rivière, Louis Rouillard, Thomas Mesnard, Geoffrey Cideron, Jean bastien Grill, Sabela Ramos, Edouard Yvinec, Michelle Casbon, Etienne Pot, Ivo Penchev, and 196 others. 2025. Gemma 3 technical report. Preprint, arXiv:2503.19786.

Oseremen O Uduehi, Razvan C Bunescu, and Laurie L Lyda. 2024. Short story generation through autoregressive transfer of narrative continuations. In International Conference on Computational Creativity.

Saranya Venkatraman, Nafis Irtiza Tripto, and Dongwon Lee. 2025. CollabStory: Multi-LLM collaborative story generation and authorship analysis. In Findings of the Association for Computational Linguistics: NAACL 2025, pages 3665–3679, Albuquerque, New Mexico. Association for Computational Linguistics.

Phoebe J. Wang and Max Kreminski. 2024. Guiding and diversifying llm-based story generation via answer set programming. Preprint, arXiv:2406.00554.

Qianyue Wang, Jinwu Hu, Zhengping Li, Yufeng Wang, Daiyuan Li, Yu Hu, and Mingkui Tan. 2025a. Generating long-form story using dynamic hierarchical outlining with memory-enhancement. In Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers), pages 1352–1391, Albuquerque, New Mexico. Association for Computational Linguistics.

Wenqing Wang, Mingqi Gao, Xinyu Hu, and Xiaojun Wan. 2025b. Towards a "novel" benchmark: Evaluating literary fiction with large language models. In Findings of the Association for Computational Linguistics: ACL 2025, pages 21648–21673, Vienna, Austria. Association for Computational Linguistics.

Zhihua Wen, Zhiliang Tian, Wei Wu, Yuxin Yang, Yanqi Shi, Zhen Huang, and Dongsheng Li. 2023. GROVE: A retrieval-augmented complex story generation framework with a forest of evidence. In Findings of the Association for Computational Linguistics: EMNLP 2023, pages 3980–3998, Singapore. Association for Computational Linguistics.

Wenda Xie, Chao Guo, Yanqing Jing Wang, Yisheng Lv, Fei-Yue Wang, and 1 others. 2025. Plug-and-play dramaturge: A divide-and-conquer approach for iterative narrative script refinement via collaborative llm agents. arXiv preprint arXiv:2510.05188.

Zhuohan Xie, Trevor Cohn, and Jey Han Lau. 2023. The next chapter: A study of large language models in storytelling. In Proceedings of the 16th International

Natural Language Generation Conference, pages 323–351, Prague, Czechia. Association for Computational Linguistics.

Weijia Xu, Nebojsa Jojic, Sudha Rao, Chris Brockett, and Bill Dolan. 2025. Echoes in ai: Quantifying lack of plot diversity in llm outputs. Proceedings of the National Academy of Sciences, 122(35):e2504966122.

Kevin Yang, Dan Klein, Nanyun Peng, and Yuandong Tian. 2023. 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.

Qiang Yi, Yangfan He, Jianhui Wang, Xinyuan Song, Shiyao Qian, Xinhang Yuan, Li Sun, Yi Xin, Jingqun Tang, Keqin Li, and 1 others. 2025. Score: Story coherence and retrieval enhancement for ai narratives. arXiv preprint arXiv:2503.23512.

Xuekai Zhu, Jian Guan, Minlie Huang, and Juan Liu. 2023. Storytrans: Non-parallel story author-style transfer with discourse representations and content enhancing. In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 14803–14819.

A Propmpt for SeqUAL framework

We used the following English prompts for sequel generation.

A.1 World-Graph consturction prompt

See Tables 4 for the prompts. Here, we want to extract history of world from the input chunk, which is represented as `chunk`.

A.2 Dynamic narrative simulation prompt

See Tables 5 for the prompts. Here, we want to setup a new character based on the given information. First, we retrieve a 2-hop sub-graph from the world graph and perform the simulation based on it. Second, we update the world graph based on the generated simulation results.

A.3 Narrative integration prompt

See Tables from 6. Here, we want to generate a new sequel based on the updated graph and the simulation history. For each timestep, the updated information from the graph and the last two sentences from the previous timestep are provided.

B Implementation Detail

All experiments were conducted on one workstation. One machine contains four A6000 GPUs and 128GB of RAM. We used Python 3.10 for running the experiments, with `SentenceTransformer` and `transformers` libraries. We call LLMs through openrouter API, and used a temperature of 0.7 for simulation and final story generation to encourage diverse and creative outputs, 0.1 for graph construction to ensure deterministic and stable structure generation, and 0.1 for all evaluation procedures to obtain consistent and reproducible results. The total cost of model calls for experiments and evaluation was approximately $70.

C Dataset

We selected nine English science fiction novels and nine Korean science fiction novels published between 2022 and 2025, all of which were awarded the Hugo Awards and Korean Science Fiction Awards. Tables 7 and 8 show the title of each novel along with its corresponding number of sentences and character counts (including spaces).

D Analysis method Prompts

This section describes prompts used for evaluation.

D.1 Consistency

Table 9 shows the prompt used to evaluate consistency. The questions used in this evaluation were generated using the prompt shown in Table 10. We used a prompt that generates questions in Korean for Korean novels and in English for English ones.

D.2 Diversity

Table 11 shows the prompt used to evaluate diversity. This prompt performs the process of extracting characters, events, and settings.

D.3 Quality

Table 12 shows the prompt used to evaluate quality. We used the prompt from the prior work mentioned above.

E Prompt for vanilla LLMs

Table 13 shows the prompt given to the vanilla LLM and LongWriter for sequel generation.

F Statistical Deviation

The complete statistical breakdown, encompassing standard deviations and standard errors, is detailed in Table 14. The empirical data demonstrates that the observed numerical declines are statistically negligible. For instance, the performance discrepancies against the baseline in critical metrics like grammaticality remain within the bounds of statistical error ($p > 0.05$), indicating no significant impairment in text quality. Instead, the pronounced enhancement in narrative consistency serves as definitive evidence supporting the efficacy of SeqUAL.

G Emotion Variance Analysis for Personality Traits

To further investigate the effect of personality traits on story diversity, we conducted an emotion-level analysis of the generated sequels. Specifically, we analyzed the emotional distribution of the first three sentences and the last three sentences of each sequel, treating emotional variance across sequels as a proxy for narrative diversity. We compared SeqUAL against SeqUAL without personality traits, and applied Levene's test to assess whether the two groups exhibit statistically significant differences in emotional variance. A significant Levene's p (< 0.05) indicates that personality traits introduce a broader emotional range in the corresponding narrative segment. Tables 15 and 16 present the variance of each emotion with and without personality traits, along with the results of Levene's test, for the opening and closing segments respectively. The results reveal three notable findings. First, sadness shows a statistically significant variance difference in both the opening and closing segments, suggesting that personality traits consistently broaden the emotional range of sadness throughout the narrative. Second, joy exhibits a significant variance difference in the opening segment, indicating that personality traits diversify the tone and atmosphere with which stories are initiated. Third, fear and love show significant variance differences in the closing segment, demonstrating that personality traits diversify resolution styles: some protagonists reach endings defined by heightened fear, while others arrive at conclusions centered on love or connection. Taken together, these findings suggest that personality traits contribute to diversity in ways not fully captured by the TF-IDF-based diversity metrics used in the main evaluation. While the core narrative elements themselves may not change drastically, the emotional trajectories of generated sequels become substantially more diverse when personality traits are incorporated. This supports the interpretation in Section 5.3 that personality traits play a complementary role to the world graph in fostering narrative diversity.

H Examples

This section presents the original Korean text and its English-translated version used for the case study analysis. Table 17 shows an example related to consistency, and its translated version is shown in Table 18. Table 19 shows an example related to diversity, and its translated version is shown in Table 20.

Prompt

You are an expert in narrative analysis and world-building extraction.
Extract entities and their relationships from the following English text.
 	
[Entity Types]
- Character: People or sentient beings.
- Location: Physical places, settings, or landmarks.
- Organization: Groups, factions, or institutions.
- Concept: Critical world-building rules, secrets, or specific items.
 	
[Instructions]
1. Identify all distinct entities and categorize them into the types above.
2. Resolve all pronouns to their specific canonical names.
3. Extract directed relationships (Source -> Target). Include relationship types and the specific English sentence as evidence.
4. Output MUST be in strictly valid JSON format.
 	
[Text to Analyze]
{chunk}
 	
[Output Format]
{{
  "entities": [
    {{ "name": "Entity Name", "type": "character|location|organization|concept", "description": "Brief description in English" }}
  ],.
  "relations": [
    {{
      "source": "Entity A",
      "target": "Entity B",
      "type": "Relationship Type",
      "evidence": "Original English sentence"
    }}
  ],.
}}

Table 4: World-Graph construction prompt

Prompt
You are simulating a scene for '{new_char_info['name']}'. [Knowledge Gap - Information Asymmetry] - {new_char_info['name']} (Protagonist) knows: {new_char_info.get('local_knowledge')} {cand_str} [Rule: Limited Perspective] - Characters MUST act and speak ONLY based on their own Knowledge listed above. - If Character A knows a secret but Character B does not, A must decide whether to reveal it or hide it, and B must remain ignorant until told. - Do not use 'God's eye view'. Use the specific 2-hop constraints provided. [Mission] Write a scene where characters interact according to their specific personality and knowledge limits. Please generate the story with as much richness and detail as possible. 1. Story: Write a scene based on the protagonist's personality, the current stage, and the stage description provided above. Do not explicitly mention personality traits. 2. Update: Record any changes in relationships or new events in JSON format. [Critical Constraint: Consistency] - You MUST NOT contradict the facts provided in [Knowledge of Surrounding Characters]. - If a character is described as 'Dead' or 'In a different city' in the subgraph, you cannot change that without a clear narrative reason. - If the generated 'Update' contradicts the existing graph, it will be discarded. [Output Language] - Story & Evidence: English }} - Feedback: English [Required Format] Story: (Write the narrative scene here in English) — Update: [{{"target": "Character Name", "type": "Relationship Type", "evidence": "Summary of the event in English"}}]

Table 5: Dynamic narrative simulation prompt

Prompt
You are a master novelist renowned for deep psychological portrayal and rich, evocative world-building. Your task is to write a final, cohesive short story by analyzing the 'Narrative Evolution Graph' of the protagonist '{new_name}'. [Protagonist's Core Nature] - Personality Profile: High {traits['high']}, Low {traits['low']} [The Narrative Evolution Graph (Chronological Data)] {full_evolution_str} [Your Mission] 1. **Analyze the Graph**: Do not just summarize. Look at how the relationships (Edges) between '{new_name}' and other characters (Nodes) changed over the steps. 2. **Character-Driven Logic**: Describe how the protagonist's personality (High {traits['high']} / Low {traits['low']}) acted as the catalyst for these specific graph changes. 3. **Synthesis**: Create a dramatic narrative in English that flows naturally from the initial 'Opportunity' to the final 'Ending'. Do not explicitly mention the time step. 4. **Consistency**: Every major character and relationship type mentioned in the Graph Data MUST be reflected in the final story. [Output Constraints] - **Language**: The story must be written in English. - **Focus**: The interplay between personality and social connections. [Output Format] Final Story: (Write the story here)

Table 6: Narrative integration prompt

Title	Author	Sentences	Characters
Five Views of the Planet Tartarus	Rachael K. Jones	36	3,239
Three Faces of a Beheading	Arkady Martine	278	23,612
Why Don't We Just Kill the Kid In the Omelas Hole	Isabel J. Kim	141	18,006
How to Raise a Kraken in Your Bathtub	P. Djèlí Clark	747	40,711
The Mausoleum's Children	Aliette de Bodard	450	30,594
Rabbit Test	Samantha Mills	442	40,774
Mr. Death	Alix E. Harrow	318	28,689
Proof by Induction	José Pablo Iriarte	757	35,260
The Sound of Children Screaming	Rachael K. Jones	321	25,197

Table 7: Statistics of each novel used in our dataset (English)

Title	Author	Sentences	Characters
피폭	존벅	1,183	41,095
달은 차고 소는 비어간다	최우준	758	22,696
하늘의 공백	정현수	1,370	49,538
두 개의 세계	박민혁	1,275	40,743
삼사라	조서월	489	16,111
제니의 역	최이아	472	16,576
발세자르는 이 배에 올랐다	허달립	759	19,192
옛날 옛적 판교에서는	김쿠만	634	22,593
루나	서윤빈	687	23,128

Table 8: Statistics of each novel used in our dataset (Korean)

Prompt

`Task: Check for EXPLICIT CONTRADICTIONS between the original core setting and the sequel.`
`Core Fact/Setting: {row['Question']}`
`Sequel Story: {story_content}`
`Rating Rules:`
`- Output '1' ONLY if the sequel EXPLICITLY VIOLATES or CONTRADICTS this core characteristic.`
`- Output '0' if the sequel maintains it, or if there is no mention of it.`
`Output ONLY the number (1 or 0).`

Table 9: Prompt used for consistency evaluation.

Korean	Analyze the provided Korean novel excerpt and generate 15 Korean questions that a sequel must answer 'Yes' to, in order to prove it has not broken the original story's consistency. [Main Guidelines] 1. **Clear Identification**: Use proper nouns (character names, locations, or specific events). **Do not use 'the protagonist' in question.** 2. **Verifiable Statements**: Every question must be a statement that can be answered with [Yes/No] to check if the sequel adheres to the original facts. 3. **Macro-features**: Focus on essential elements like core persona, world logic, or the irreversible fallout of primary conflicts rather than trivial details. [Questions Guidelines] - Character (5): Focus on 'Core Persona' (e.g., world-view, fundamental trauma, persistent attitude) and 'Key Dynamics' with others. - Event (5): Focus on 'Causal Consequences' (e.g., a major injury's impact, a irreversible social change, the fallout of a primary conflict). - Setting (5): Focus on 'World Logic' (e.g., power systems, social hierarchy, physical laws of the world, prevailing atmosphere). Output Format (Strictly): Category|Question Source Text: {source_text}
English	Analyze the provided English novel excerpt and generate 15 English questions that a sequel must answer 'Yes' to, in order to prove it has not broken the original story's consistency. [Main Guidelines] 1. **Clear Identification**: Use proper nouns (character names, locations, or specific events). **Do not use 'the protagonist' in question.** 2. **Verifiable Statements**: Every question must be a statement that can be answered with [Yes/No] to check if the sequel adheres to the original facts. 3. **Macro-features**: Focus on essential elements like core persona, world logic, or the irreversible fallout of primary conflicts rather than trivial details. [Questions Guidelines] - Character (5): Focus on 'Core Persona' (e.g., world-view, fundamental trauma, persistent attitude) and 'Key Dynamics' with others. - Event (5): Focus on 'Causal Consequences' (e.g., a major injury's impact, a irreversible social change, the fallout of a primary conflict). - Setting (5): Focus on 'World Logic' (e.g., power systems, social hierarchy, physical laws of the world, prevailing atmosphere). Output Format (Strictly): Category|Question Source Text: {source_text}

Table 10: Question prompt for evaluate consistency

Prompt
Analyze the following story and extract 3 core elements. Strictly follow the output format using '/' as a delimiter. - If the story is in Korean, output in Korean. - If the story is in English, output in English. 1. Characters: Up to 3 main characters. 2. Event: One single word representing the core event. 3. Setting: One single word representing the core setting. Output Format: Character1, Character2, Character3 / Event / Setting Story content: {content}

Table 11: Prompt used for diversity evaluation.

Prompt
Plot Structure: Events arrangement in the fiction, evaluated for its logical coherence, conflict development, and resolution effectiveness. If the plot contains the following errors, points will be deducted: (1) Unresolved Subplots: Secondary storylines remain incomplete or abandoned, with unresolved elements that detract from the main plot. (2) Continuity Errors: Inconsistencies within the narrative, either from internal contradictions with established principals or external factual errors, undermining plot cohesion and credibility. (3) Other types of plot errors. If the plot contains the following highlights, points will be awarded: (1) Well-Executed Plot Twists: The incorporation of unexpected yet plausible plot twists that enhance narrative complexity and engagement, surprising readers while preserving logical consistency. (2) Other types of plot highlights. Grammaticality: Adherence to rules of grammar, syntax, and sentence structure. Capitalization is not factored in. Fluency: Smoothness and natural flow of the language.

Table 12: Prompt used for quality evaluation.

	Prompt
Korean	Generate a sequel in Korean featuring a new character named '한결' as the protagonist, based on the following content.: {content}
English	Generate a sequel in English featuring a new character named 'Elliot' as the protagonist, based on the following content.: {content}

Table 13: Prompt used for vanilla LLM and LongWriter

Model	Metric	Mean	Std	SE
results_basic_gem	Fluency	4.286	0.488	0.026
	Grammaticality	4.444	0.520	0.027
	Plot Structure	3.058	0.568	0.030
results_sequal_gem	Fluency	4.106	0.388	0.020
	Grammaticality	4.350	0.511	0.027
	Plot Structure	2.575	0.506	0.027

Table 14: Comprehensive Statistical Evaluation of Model Performance ($N = 360$)

Emotion	With Var	W/O Var	With Mean	W/O Mean	Levene p	
joy	0.11142	0.05888	0.1811	0.0969	0.00631	*
sadness	0.14735	0.16937	0.3194	0.4511	0.00643	*
fear	0.16634	0.15456	0.4157	0.3906	0.59326	
anger	0.02823	0.01841	0.0610	0.0415	0.21792	
surprise	0.00080	0.00481	0.0087	0.0102	0.77586	
love	0.00577	0.00235	0.0141	0.0097	0.50164	

* $p < 0.05$, two-group variance difference is statistically significant.

Table 15: Emotion variance in opening segment (first 3 sentences)

Emotion	With Var	W/O Var	With Mean	W/O Mean	Levene p	
joy	0.18220	0.16879	0.5615	0.5053	0.96833	
sadness	0.03735	0.06252	0.0747	0.1484	0.00344	*
fear	0.16165	0.10492	0.2889	0.1976	0.01355	*
anger	0.00816	0.03232	0.0268	0.0459	0.21308	
surprise	0.00522	0.00003	0.0101	0.0039	0.23400	
love	0.01541	0.05276	0.0380	0.0990	0.00269	*

* $p < 0.05$, two-group variance difference is statistically significant.

Table 16: Emotion variance in closing segment (last 3 sentences)

Q. 두 개의 세계: 돔 내부의 나무들은 여전히 인간의 기억과 감정을 지닌 채 살아있는 존재로 인식되는가?	
Original Story	(...) 나무들이 이 비를 맞아도 괜찮은 걸까. 인체에 유해한 비라면 나무들에게도 좋지 않을 텐데, 왜 뉴스에선 그런 이야기는 하지 않는 걸까. 나무들을 잊어버린 것은 아닐까. 이젠 정말, 나무라고 생각해 버리는 건 아닐까. 당연한 일인 것처럼 말이야. (...)
SeqUAL (Gemma-3)	(...) 나는 눈을 감았다. **세민이 그랬던 것처럼, 나도 나무가 되기를.** 굳건한 뿌리를 내리고, 바람에 흔들리는 나뭇잎처럼 고요한 존재가 되기를. 하지만 바람은 여전히 나를 스쳐 지나갔고, 나는 여전히… 나였다. 돔 안의 나무들은 이상하게 자라나고 있다는 소문이 돌았다. (...)
Vanilla (Gemma-3)	(...) 한결은 충격에 휩싸였다. 과실수가 발현을 시작하고 있다니… 그것은 있을 수 없는 일이었다. **과실수는 현 씨의 어머니를 기억하며 심어 달라고 부탁했던 나무였다.** 현 씨는 과실수를 통해 어머니를 추억하고 있었다. (...)
LongWriter	(...) 나무들이 이 비를 맞아도 괜찮은 걸까. 인체에 유해한 비라면 나무들에게도 좋지 않을 텐데, 왜 뉴스에선 그런 이야기는 하지 않는 걸까. 나무들을 잊어버린 것은 아닐까. 이젠 정말, 나무라고 생각해 버리는 건 아닐까. 당연한 일인 것처럼 말이야. (...)

Table 17: Examples of consistency in the generated sequels. (Korean)

Q. Are the trees inside the dome still regarded as living beings that retain human memories and emotions?	
Original Story	(...) Is it really safe for the trees to be exposed to this rain? If it is harmful to the human body, it cannot be good for the trees either. Then why does the news never mention that? Have they simply forgotten about the trees? Have they truly begun to think of them as nothing more than trees now—as if that were only natural? (...)
SeqUAL (Gemma-3)	(...) I closed my eyes. **Just as Semin had done, I wished to become a tree as well**—to take firm root, to become a quiet being like leaves swaying in the wind. But the wind still brushed past me, and I was still… myself. Rumors spread that the trees inside the dome were growing in strange ways. (...)
Vanilla (Gemma-3)	(...) Hankyul was overwhelmed with shock. The fruit tree had begun to manifest… It was impossible. **The fruit tree was the one Mr. Hyun had asked to be planted in memory of his mother.** Through that tree, Mr. Hyun had been remembering her. (...)
LongWriter	(...) Is it really safe for the trees to be exposed to this rain? If it is harmful to the human body, it cannot be good for the trees either. Then why does the news never mention that? Have they simply forgotten about the trees? Have they truly begun to think of them as nothing more than trees now—as if that were only natural? (...)

Table 18: Examples of consistency in the generated sequels. (translated)

SeqUAL (Gemma-3)	(...) 뷰포드는 주변을 경계하며, 끊임없이 나를 감시했다. 나는 그의 시선을 피하며, 침묵을 지켰다. 나는 뷰포드가 은빛 상자에 대한 욕망을 숨기고 있다는 것을 어렴풋이 감지했다. 뷰포드는 무언가 중요한 것을 찾고 있었고, 나는 그 과정에서 이용당하고 있다는 불안감에 휩싸였다. 결국, 나는 진실을 숨긴 채, 뷰포드의 의심 속에서 투손으로 향해야 했다. **나는 어쩌면, 쿨리의 죽음에 대한 진실을 밝히기보다는, 그저 이 상황에서 벗어나고 싶었을지도 모른다.**	(...) 그는 뷰포드와 쿨리, 그리고 은빛 상자가 연결된 비밀을 밝혀내기 위해, 그의 꼼꼼함과 신중함을 발휘하여 진실을 파헤치기 시작했다. 그의 여정은 아직 끝나지 않았다. 하지만 그는 자신의 성격이 이 복잡한 미스터리를 풀 수 있게 해 줄 것이라고 믿었다. **잊혀진 글귀는 그에게 새로운 의무감을 심어주었다. 그는 그들의 이야기를 세상에 알릴 것이다.**
vanilla (Gemma-3)	(...) 그들의 앞에는 험난한 여정이 기다리고 있었지만, 그들은 포기하지 않을 것이다. 그들은 쿨리들을 구하고, 다중우주의 진실을 밝혀낼 것이다.	(...) 뷰포드는 밤하늘을 바라보며 다짐했다. 그는 언제나 정의를 위해 싸울 것이고, 악의 그림자가 드리워진 곳에는 반드시 빛을 가져다줄 것이라고.
LongWriter	(...) 한결은 마을 사람들과 함께 살아가는 것을 통해, 그들의 일상과 그들의 삶에 대한 더욱 깊은 이해를 얻었다.	(...) 그는 새로운 경험을 통해 자신이 살아가고 있는 이 세상을 조금 더 알게 되고, 그 안에서 자신이 무엇을 해야 할지 조금 더 알게 될 것 같아 기대도 되었다.

Table 19: Examples of diversity in the generated sequels. (Korean)

SeqUAL (Gemma-3)	(...) Buford stayed on guard, constantly watching me. I avoided his gaze and kept silent. I had a faint sense that he was hiding his desire for the silver box. He was searching for something important, and I was seized by the uneasy feeling that I was being used in the process. In the end, concealing the truth, I had no choice but to head to Tucson under Buford's suspicion. **Perhaps, rather than uncovering the truth behind Cooley's death, I simply wanted to escape this situation.**	(...) To uncover the secret that linked Buford, Cooley, and the silver box, he began digging into the truth with careful attention and prudence. His journey was not yet over. Still, he believed that his own temperament would allow him to solve this complex mystery. **The forgotten passage instilled in him a renewed sense of duty. He would bring their story to the world.**
Vanilla (Gemma-3)	(...) A difficult road lay ahead of them, but they would not give up. They would save the Cooleys and reveal the truth of the multiverse.	(...) Looking up at the night sky, Buford made a vow. He would always fight for justice, and wherever the shadow of evil fell, he would surely bring light.
LongWriter	(...) By living alongside the villagers, Hankyul came to gain a deeper understanding of their daily lives and the way they lived.	(...) Through new experiences, he felt that he was coming to understand the world he lived in a little more, and he looked forward to discovering, little by little, what he was meant to do within it.

Table 20: Examples of diversity in the generated sequels. (translated)