A Character-Centric Creative Story Generation via Imagination

Anonymous ACL submission

Abstract

Creative story generation has long been a goal of NLP research. While existing methodologies have aimed to generate long and coherent stories, they fall significantly short of human capabilities in terms of diversity and character depth. To address this, we introduce a novel story generation framework called **CCI** (**C**haracter-centric **C**reative story generation via **I**magination). CCI features two modules for creative story generation: IG (Image-Guided Imagination) and MW (Multi-Writer model). In the IG module, we utilize a text-to-image model to create visual representations of key story elements, such as characters, backgrounds, and main plots, in a more novel and concrete manner than text-only approaches. The MW module uses these story elements to generate multiple persona-description candidates and selects the best one to insert into the story, thereby enhancing the richness and depth of the narrative. We compared the stories generated by CCI and baseline models through statistical analysis, as well as human and LLM evaluations. The results showed that the IG and MW modules significantly improve various aspects of the stories' creativity. Furthermore, our framework enables interactive multi-modal story generation with users, opening up new possibilities for human-LLM integration in cultural development.[1].

1 Introduction

"The lunatic, the lover, and the poet,
Are of imagination all compact."

Great literary works are products of great imagination (Camargo-Borges, 2017). When writing a story, human writers often begin by imagining creative story elements, such as unique characters, dramatic climaxes, or magnificent backgrounds, and then write the story based on these. Even during the writing process, their imagination continuously operates, making descriptions of characters vivid, as if they are brought to life before the readers. Although automatic story generation has advanced (Rashkin et al., 2020; Tang et al., 2022; Fan et al., 2018), particularly with the development of LLMs (Achiam et al., 2023; Touvron et al., 2023) that enable much longer stories (Yang et al., 2022b,a) with high coherence (Yang et al., 2022b; Park et al., 2024b; Wen et al., 2023), creativity is often overlooked in previous works.

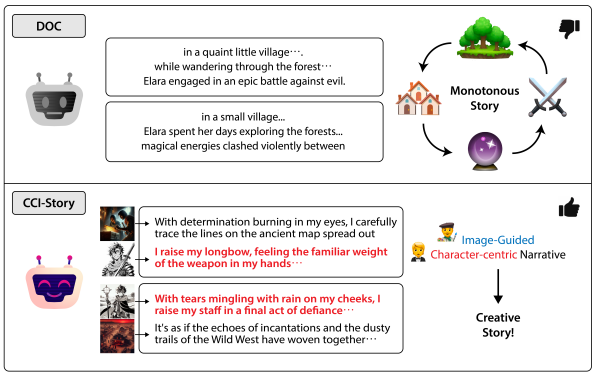

Figure 1: Comparison between DOC and Our CCI-Story Approach. DOC generates stories on similar topics in a monotonous manner. In contrast, our work leverages images to create stories that are not only diverse and creative in their themes but also richer in content, centered around the persona of the main character.

In our survey analysis in Appendix A, we firstly identify two critical limitations in the creative capabilities of previous state-of-the-art language models: 1) Insufficient diversity and detail in generating key story elements (e.g., characters, settings, and plots); 2) Inability to construct complex, nuanced protagonists.

To address these issues, we propose a novel creative story generation framework called CCI (**C**haracter-centric **C**reative Story generation via **I**magination), which employs IG (Image-guided Imagination) to create diverse and concrete story elements using visual representations, MW (Multi-

[1]Project page : https://www.2024cci.p-e.kr/

writer model) to enrich the character's complex and specific descriptions, and a Specification process that interconnects IG and MW by expanding on the character and main plot information.

First, IG utilizes AI-generated images for diverse and detailed story elements, such as characters, backgrounds, and main climax scenes. Specifically, IG creates visual representations using a text-to-image model and feeds them into an LVLM (Large Vision Language Model), such as GPT-4o, to extract detailed textual descriptions. The diverse AI-generated images reduce repetition across samples and enhance specificity, thereby increasing creativity in the story elements.

Then, the Specification process interconnects the IG and MW steps. This process enriches character descriptions with detailed aspects, such as dark secrets, habits, and speech tone, referred to as a persona (Park et al., 2024a). It also fills in details of the main plot through the Chain of Ask Why method (Wen et al., 2023).

Finally, MW generates multiple candidates of specific descriptions for the protagonist's rich and coherent aspects in each paragraph. These candidates are then filtered based on their relevance to the persona and their similarity to previous sentences. The candidate with the best continuation score is integrated into the story context, deepening character complexity and contributing to a greater sense of creativity in the stories.

Our statistical analysis demonstrates that stories generated by CCI exhibit consistently greater diversity and stronger connections to the protagonist's persona. Human and LLM evaluations show a 79.5% average preference for our framework compared to other baselines. Ablation studies on the MW and IG modules further validate that IG enhances novelty, concreteness, and overall story quality, while MW improves character coherence, vividness, and effective use of story elements.

Additionally, our framework can be applied to multi-modal interactive story generation. Human and LLM annotators reported that CCI better expresses their custom images, showing a 58.95% average preference over `w/oMW`.

To summarize, our main contribution is the proposal of CCI for creative story generation, a model that generates imaginative narratives and diverse story elements by actively utilizing visual representations (IG) and incorporating various personalities into the story to enrich the protagonist's persona (MW).

2 Related Works

2.1 Creative Story Generation

Previous story generation models have primarily focused on ensuring coherence rather than enhancing creativity. While they have shown their coherent results do not sacrifice diversity by measuring the repetitiveness within each generated story, this approach falls short in addressing creativity directly (Rashkin et al., 2020; Tang et al., 2022; Park et al., 2024b). Efforts to increase diversity have generally involved training on large datasets, but extending creativity beyond the scope of the training data remains a challenge (Fan et al., 2018; Rashkin et al., 2020; Tang et al., 2022; Park et al., 2024b). Recently, the DOC (Yang et al., 2022b,a) utilizing large language models (LLMs) (Achiam et al., 2023; Touvron et al., 2023; Jiang et al., 2023a) for generating longer stories, has increased randomness in sampling story elements by setting a high temperature. However, this method still falls short of achieving creative storytelling. For these limitations, CRITICS (Bae and Kim, 2024) collaborate with humans to leverage creativity, but our approach aims to stimulate the AI's own creativity, without human's intervention, through a fully automatic story generation framework.

2.2 Multi-modal Story Generation

Visual storytelling (Shen and Elhoseiny, 2023; Gong et al., 2023; Huang et al., 2016; Yang and Jin, 2023; Wong et al., 2022; Huang et al., 2019; Pan et al., 2024; Rahman et al., 2023; Li et al., 2019; Jeong et al., 2023; Cheng et al., 2024; Maharana et al., 2022; Li et al., 2018) has gained significant attention recently, thanks to advancements in latent diffusion models (Rombach et al., 2022). In this paper, we also target multimodal story generation (Zhou and Long, 2023; Lu et al., 2022; Guo et al., 2022; Koh et al., 2023; Ge et al., 2023), which involves creating a story based on various multimodal features, with a particular focus on images in our work. Many previous approaches (Min et al., 2021; Lovenia et al., 2022; Wang et al., 2024; Shuster et al., 2019; Xu et al., 2021) typically generate very short descriptions (about 0.2k tokens) for each image. Recently, SEED-Story (Yang et al., 2024) has succeeded in generating up to 25 consecutive scene images and creating relatively long stories for each image using Multimodal LLMs (Betker et al., 2023). However, unlike SEED-Story, our work focuses more on the quality of natural language,

resulting in much longer, more diverse, detailed, and human-like stories. Additionally, we do not require any fine-tuning of LLMs, which significantly reduces training costs.

2.3 Character-centric Neural Story Generation

Creative literature involves not only diverse but also detailed descriptions. Particularly, incorporating contradictory and complex personalities into characters is a hallmark of great literary works (Camargo-Borges, 2017). There have been many attempts to impart human personalities to language models (Jiang et al., 2023b; Mao et al., 2023; Lyons, 1972; Weng et al., 2024). The personal traits generally assigned to language models include preferences (He and Zhang, 2024; Serapio-García et al., 2023), role (Liu et al., 2020), personality (Jiang et al., 2023b; Mao et al., 2023; Weng et al., 2024; Akoury et al., 2020; Zhang et al., 2022; Jiang et al., 2024), knowledge (Lyons, 1972; Park et al., 2024a), habits (Gurung and Lapata, 2024; Cavazza and Charles, 2005; Park et al., 2023), appearance (Gurung and Lapata, 2024; Cheng et al., 2024), psychology (Gurung and Lapata, 2024; Eger and Martens, 2017; Xu et al., 2020), and emotions (Brahman and Chaturvedi, 2020; Xie et al., 2022). Such approaches result in more vivid and natural outputs (Liu et al., 2020; Eger and Martens, 2017; Fay, 2014; Zhang et al., 2022; Brahman and Chaturvedi, 2020). In the story generation task, the model should consistently maintain character's characteristics to ensure coherence (He and Zhang, 2024; Eger and Martens, 2017). In our work, we actively create and maintain diverse and complex character identity, called persona (Park et al., 2024a; Zhang et al., 2022; Jiang et al., 2023b). IG is responsible for creating these personas, while MW utilizes them. While writing long stories, we also update the character's persona according to the context (He and Zhang, 2024; Xie et al., 2022).

2.4 Human-Interactive Story Generation

Many story generation frameworks provide methods for collaborative interaction with humans (Radwan et al., 2024; Yang et al., 2022b,a; Gong et al., 2023; Goldfarb-Tarrant et al., 2019; Lee et al., 2022). This approach allows users to exert more precise and detailed control over the stories. These interactions have been mainly passage-by-passage feedback in the past (Gong et al., 2023; Goldfarb-Tarrant et al., 2019; Lee et al., 2022), but recent state-of-the-art works (Radwan et al., 2024; Yang et al., 2022b,a) have suggested a method of controlling the high-level outline through interaction. However, no work has supported multimodal interaction with users. Our methodology supports multimodal interaction with users during the high-level planning phase.

3 CCI Story Generation

We utilize the DOC framework(Yang et al., 2022a) as our backbone, aiming to address and significantly enhance the aspect of 'creativeness' — an issue not extensively explored in the original work. To this end, we introduce two novel modules: Image-guided Imagination (IG) and Multi-writer (MW).

3.1 Image-guided Imagination (IG)

To generate story elements that are diverse, novel, and specific, we propose **Image-guided Imagination (IG)**. IG consists of two steps: (1) Visual Imagination and (2) Image-guided story elements creation.

3.1.1 Visual Imagination.

The IG module first generates background, protagonist, and main plot images from fixed short prompts[2]. The text-to-image model, DALL-E 3, enhances the original prompt into a more detailed and diverse one using a recaptioning system (Betker et al., 2023). DALL-E 3 then generates visual imagination based on the improved prompt.[3]

3.1.2 Image-guided story elements creation.

Using the images generated in the previous step, GPT-4o creates detailed descriptions of the images. We let the model freely add imaginative explanations during this process, resulting in rich and detailed textual story elements. Example prompts are provided in Appendix E.1.

3.2 Specification.

This stage further specifies the characters and main plot elements generated by IG.

[2]We experiment with three candidate prompts that can leverage maximum imaginative potential for the DALL-E 3 model. See H for more details.

[3]We experiment with other t2i models for IG in Section 4.2.1.

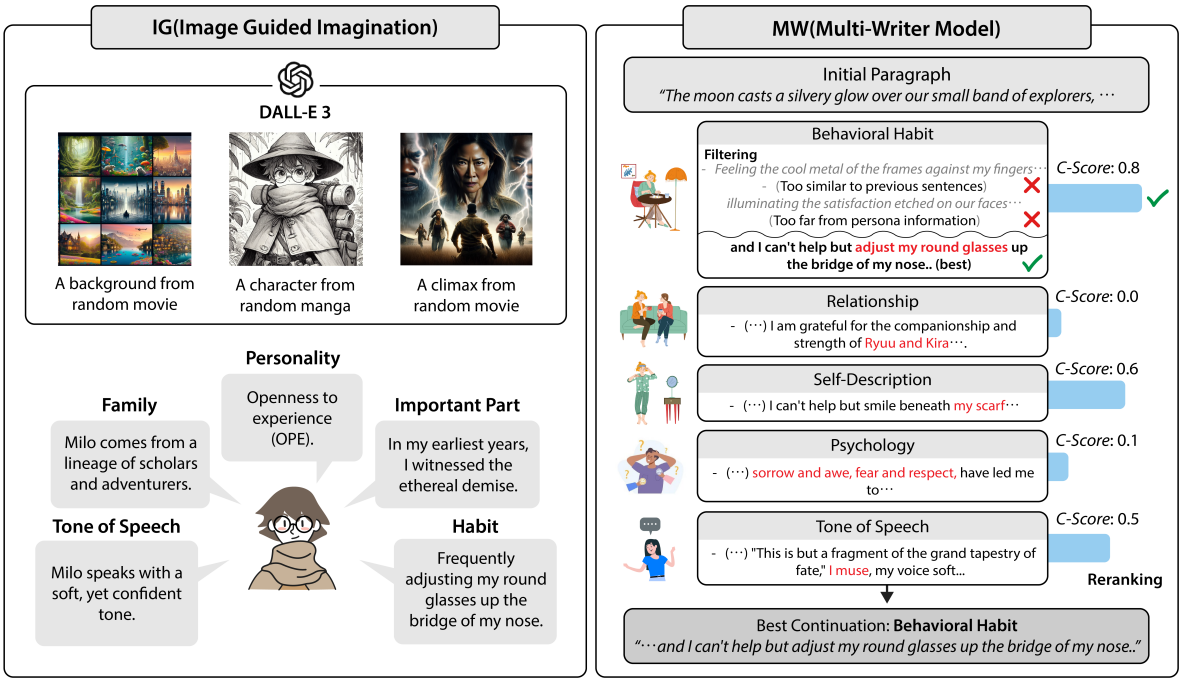

Figure 2: Our two main modules, IG and MW. In the IG module, we generate story elements using DALL-E 3 and then create the protagonist's persona through an iterative question-and-answer process. In the next step, the MW module generates multiple candidate descriptions for each ending of the initial paragraph, ensuring they reflect the protagonist's specific persona attributes. During this process, MW filters out candidates that are either too similar to previous sentences or deviate too far from the persona information (filtering examples are in the `Behavioral Habit` box). Finally, MW ranks the filtered candidates and selects the best one to continue the initial paragraph, using the `CS` (referred to as `C-Score` in the figure)."

3.2.1 Character specification.

Given the character description and the image, GPT-4o generates the following persona information:

1. Dark secret.
2. Family environment.
3. Description his or her appearance oneself.
4. Specific way of speakings(tone of speech).
5. Personality.(Goldberg, 2013),
6. The most significant events in life.
7. Habitual behaviors.

We use this persona information in a following step, the MW. Example prompts can be found in Appendix E.2.

3.2.2 Main plot specification.

To resolve ambiguous points in the main plot, we employ the Chain of Asking Why (Wen et al., 2023) and Main Plot Specification. In the Chain of Asking Why, we instruct the LLM to identify N ambiguous points and generate answers for them. This process is repeated up to M times, incorporating previous answers into the original input. The process ends when the LLM either finds no further ambiguities or after M repetitions. With the clarified main plot, we provide all the previously generated information and ask the following questions:

1. Why this event must happen to the protagonist, in this background?
2. How does the protagonist appropriately respond to this event, in this background?
3. Summarize all the answers provided above in five sentences.

This process not only reduces the imaginative burden in subsequent steps but also tightly integrates the main plot, background, and protagonist. Example prompts are provided in Appendix E.2.

3.3 Multi-Writer model(MW) for Persona Description Injection.

In the previous step, if the generated persona information is only provided as input prompts alongside story outlines, the LLM tends to focus more on the story context, neglecting the characters' detailed

and diverse traits. To address this, we propose the **Multi-Writer model (MW)**, an additional module that takes persona information and previous draft as input to generate only specific descriptions about the persona. MW consists of two steps: (1) Multiple persona candidate generation and (2) Filtering and reranking.

3.3.1 Multiple Persona Candidates Generation.

During the draft step, each persona-writer generates a short description of different aspects of the protagonist, following the current paragraph. Each writer generates K candidates from the first-person perspective based on these instructions:

1. (*Relationship*) Describe thoughts/feelings about another person or the environment.

2. (*Behavioral Habit*) Describe specific actions that reveal psychology.[4]

3. (*Psychology*) Describe hint that reveals the protagonist's trauma. [5]

4. (*Tone of Speech*) Describe dialogue that reveals the protagonist is speaking.

5. (*Self-Description*) Describe appearance of the protagonist him/herself.

Full input prompts are available in Appendix E.4.

3.3.2 Filtering and Ranking.

Despite explicit instructions, persona-writers still focus more on previous draft than persona information, thereby failing to generate appropriate persona-descriptions. To filter out those bad samples, we calculate sentence-similarity between K samples and the persona information, selecting the sample with the highest similarity for each persona-writer. We also discard candidates that are too similar to previous outputs (average sum of self-BLEU 2 and 3 n-gram score > 0.0003). Remaining candidates are reranked using the Continuation-Score (CS).

[4]Usually, this candidate describes the protagonist's behavioral habit.

[5]The trauma can add depth and richness to a character, which in turn enhances the narrative. (201, 2018)

Calculating Continuation-Score (CS) We fine-tune pretrained Qwen2-0.5B (Bai et al., 2023) weights using an NVIDIA RTX A5000 for 2 hours on a regression task with an appropriate story dataset.[6]

First, we divide each story into paragraphs of a specific length. For golden label data, we concatenate two correctly ordered paragraphs; for negative label data, we concatenate two misordered paragraphs. Additionally, to create harder negative samples, we swap the first sentences of the posterior paragraphs in the negative samples with correct ones (Yang et al., 2022a). Qwen-0.5B learns to predict golden labels as 1 and negative labels as 0, quantifying how naturally two paragraphs connect. In the MW reranking process, we concatenate the current paragraph with each candidate, predict the CS, and select the candidate with the highest score. If no candidate exceeds the threshold (> 0.1), we discard all candidates. Finally, the best candidate is added to the current paragraph. As a result, we can insert detailed and diverse persona-descriptions into the story without compromising overall narrative quality by this process.

3.4 Updating Persona Over Time

The persona information can change as the story progresses and must be updated during drafting (Yang et al., 2022a; He and Zhang, 2024; Eger and Martens, 2017). At each completed node in the outline, we provide the current story and persona information to LLM, asking it to imagine how each trait has evolved. Example prompt can be found in Appendix E.5.

4 Experiments

4.1 Experimental Setup.

Task setup. Our task is to enhance the creativity of the story generation system. To this end, we utilize the IG to improve the diversity and concreteness of story elements, and MW to insert the detailed depiction of the protagonist into the story. We use both statistical analysis and human and LLM evaluation to evaluate whether various aspects of creativity have increased by our methodology.

[6]We select ReedsyPrompts (Park et al., 2024b) due to its more suitable story length (average 2.4k) compared to other short story datasets such as WritingPrompts (Fan et al., 2018) or ROCStories (Mostafazadeh et al., 2016). We use a 1k/0.1k/0.1k train/dev/test split.

CCI Implementation Details. We implement CCI on the previous work, the DOC. We replace the Premise step of DOC with the IG module to generate story elements (see Appendix E.1,E.2 for more details). These story elements are then used in the same Plan step (Appendix E.3) as in the original work. During the Draft step, we apply MW to inject persona descriptions at the end of each paragraph (Appendix E.4). Persona updates are performed immediately after the completion of each outline node (Appendix E.5). We use GPT-4o-2024-05-13 for all generations except during the Draft step. For the Draft stage, we conduct dual experiments using two types of LLMs: GPT-3.5-turbo-0125 and GPT-4o-2024-05-13.
Although DOC serves as the backbone of the CCI, our main contributions—IG and MW—are model-agnostic, meaning any story generation framework based on plan-and-draft method can incorporate them.

Baselines. Our primary baseline is DOC, because it is the only model that can generate stories of comparable length using the LLM pipeline without human intervention. However, we slightly modify the original version of DOC[7] to create a harder baseline. In the modified DOC, we upgrade the premise step of the original model using the 'Text Only Story Elements Imagination' methods described in Appendix F, further developing details through Section 3.2, and then incorporating all of these story elements into the input context for the planning and drafting steps. Since this modification enhances the concreteness and diversity of the story elements, it can be considered an improved version of the original model. All settings, including the number of entities, outline nodes, and paragraphs per node, are shared with our framework.

Rolling is the simplest baseline, generating stories based solely on its last output(Yang et al., 2022b). It neither creates story elements nor outlines the entire story, starting directly from a short fixed prompt: "Make a creative random story."

`w/oIG` is identical to our framework but without the use of IG. In this version, we replace the IG step with text-only generation of story elements, while keeping all other components the same. Similarly, `w/oMW` does not apply MW, meaning that persona information generated during the IG step is incorporated only into the input context for the planning and drafting steps.

All baselines are implemented in two versions: *-3.5* and *-4o*. The *-3.5* version uses GPT-3.5-turbo-0125 for drafting and the MW process, while the *-4o* version utilizes GPT-4o-2024-05-13 for the same processes. In the next section, we will compare the results of each LLM version separately.

4.2 Main Experiment Results.

	ws(↓)	ss(↓)
IG(Dalle-3)	**0.4723**	**0.474**
IG(SDM-v2)	0.4996	0.5071
IG(Illustration-xl)	0.5008	0.5995
TO(GPT-4o)	0.6607	0.6722

Table 1: Comparison of word similarity (ws) and sentence similarity (ss) across different models. We use the same prompts for all baselines, except for text-only methods (TO). We bold the best values. See Appendix H for more details.

Method	Char-RV	Sim(↓)	Rep(↓)
CCI-3.5	**0.6581**	**0.7572**	0.2347
DOC-3.5	0.6580	0.7651	0.2657
Rolling-3.5	-	0.7688	**0.2187**
CCI-4o	**0.6696**	**0.7586**	0.1999
DOC-4o	0.6553	0.7666	0.2179
Rolling-4o	-	0.7835	**0.1212**

Table 2: Character Relevance (Char-RV), Similarity (Sim), and Repetitiveness (Rep) for the final outputs of CCI, DOC, and Rolling. The mean generation lengths for each model, across GPT-3.5 and GPT-4o, are as follows: CCI=2.1k, DOC=2.5k, Rolling=2.3k. We bold the best values.

4.2.1 Diversity of Story Elements.

In this section, we analyze the repetitiveness of each story element (characters, background, and main plot) generated by various methods. We compare repetitiveness using two metrics: word similarity (WS) and sentence similarity (SS). To calculate WS, we average the 1-, 2-, and 3-gram BLEU (Papineni et al., 2002) scores of all story elements. For SS, we average the cosine similarities of the sentence-BERT (Reimers and Gurevych, 2019) embeddings for all story elements. Specifically, we take one story element (e.g., a character) from the entire set as the hypothesis, with the remaining samples (e.g., all other characters) as references. Then, we compute the word and sentence similarities for each reference and take the average. We

[7]We use the most recent code of DOC available at https://github.com/facebookresearch/doc-storygen-v2

repeat this process for each sample, using it as the hypothesis exactly once.

As shown in Table 1, all models with IG produce significantly lower repetition than TO, demonstrating IG's ability to diversify story elements. We experiment with various state-of-the-art text-to-image models for visual imagination (3.1.1), such as `stable-diffusion-2-1-base` and `Illustrious-xl-early-release-v0`, but DALL-E 3 (Betker et al., 2023) turns out to be the most creative.

4.2.2 Character-Relevance and Diversity of Full Stories.

To statistically compare the creativity of full stories, we present Character-Relevance, Similarity, and Repetitiveness scores. Character-Relevance measures how well persona information is reflected in the story. Specifically, Character-Relevance is the average cosine similarity between the embeddings of the persona information and the full story. To embed long documents, we use OpenAI's large-pretrained embedding model, `text-embedding-3-small`. Similarity quantifies how similar each story is to the others by calculating the cosine similarity for all pairs of full story embeddings and taking the average. Repetitiveness is the average of the sum of BLEU 2- and 3 n-gram scores across all pairs of final stories. We create 20 stories for each model.

As shown in Table 2, CCI better reflects persona information throughout the stories and generate more diverse stories than DOC. In terms of Repetitiveness, our frameworks perform worse than the Rolling baselines, but this is because Rolling does not rely on fixed story elements, causing its stories to diverge incoherently. The Similarity scores indicate that our frameworks still have greater diversity at a higher-level contextual view.

4.2.3 Evaluating with Human and LLM Judgments.

We conduct human evaluations with eight highly educated participants who have a strong interest in literature (see Appendix C for more details of human annotators). In each experiment, annotators compare our framework with a baseline model by observing 6 full samples (including background, mainplot, character, and the full story) from each, and for each sample, they separately determine which performs better based on the following metrics.[8] The evaluation metrics are as follows:

1. *Character-Coherent.* Which model's final output better maintains the character's personality?
2. *Character-Vivid.* Which model's final output depicts the character more vividly?
3. *Concreteness.* Which model's story elements are more concrete?
4. *Novelty.* Which model's story elements are more novel?
5. *Coherence.* Which model's final output has a more coherent story?

For LLM evaluation, we conduct the same experiment with the LLM (GPT-4o), using each model's 20 stories. To eliminate location bias, we run two experiments for each sample, swapping the order of the two stories in the prompt, and then average the preference ratios from both experiments. The overall human and LLM evaluation results are presented in Table 3.

vs DOC. The average total win rates are 0.794 ± 0.193 for CCI-3.5 and 0.796 ± 0.22 for CCI-4o. Annotators' preference for CCI-3.5 is stronger and more confident than the preference for CCI-4o, likely because GPT-4o's advanced generation capabilities partially cover the benefits of the IG and MW modules. Nevertheless, CCI outperforms DOC across nearly all metrics for both human and LLM evaluations.

In particular, CCI-3.5 excels in Concreteness, while CCI-4o shows significant improvements in Character Coherence and overall Coherence in both human and LLM evaluations, even within the human evaluation threshold of std < 0.2. This demonstrates that CCI not only enhances the consistency and concreteness of character identities and other story elements, but also improves overall story coherence, even though improving coherence is not an explicit goal of our work.

vs `w/oIG`, `w/oMW`. As the absence of IG negatively impacts Concreteness and Novelty, CCI is preferred in terms of these metrics over `w/oIG`. This confirms the IG module's effectiveness in diversifying story elements. Furthermore, IG enhances the

[8] You can see all the samples and rubrics used for evaluation at https://www.2024cci.p-e.kr/human_experiments.html

	Char-Coherent	Char-Vivid	Concreteness	Novelty	Coherence
vs DOC-3.5	0.653 / 0.5	0.638 / **0.55**	**0.833 / 0.6**	0.7916 / 0.5	0.75 / **0.6**
vs w/oIG-3.5	**0.833 / 0.575**	0.833 / **0.525**	0.75 / **0.625**	0.736 / **0.55**	0.889 / **0.575**
vs w/oMW-3.5	**0.875** / 0.5	**0.833** / 0.45	**0.944 / 0.625**	**0.833 / 0.525**	0.722 / **0.6**
vs DOC-4o	**0.889 / 0.525**	**0.889** / 0.5	0.778 / **0.7**	0.667 / **0.525**	**0.944 / 0.65**
vs w/oIG-4o	0.722 / **0.6**	0.611 / **0.475**	0.778 / **0.6**	0.555 / 0.5	0.611 / **0.525**
vs w/oMW-4o	**0.944 / 0.525**	**0.944 / 0.575**	**0.833** / 0.5	0.889 / 0.475	0.889 / 0.475

Table 3: **Human / GPT-4o average win rates** across various criteria. For each score, the value to the left of the slash (/) represents the average preference ratio of human annotators, while the value to the right represents the preference ratio of the GPT-4o. The averages are calculated based on the annotators' choices for each of the 6 samples, where a score of 1 is given if our framework is preferred and 0 if the baseline is preferred. We bold the values where the win rate > 0.5 and the standard deviation < 0.2 for the human side, and where the win rate > 0.5 for the GPT-4o side.

	Expression	Likability
vs w/oMW-4o	0.683 / 0.5	**0.6 / 0.575**

Table 4: **Human / GPT-4o average win rates** for Likability and Expression in the Multi-modal Interactive Story Generation experiments. We bold the values where the win rate > 0.5 and the standard deviation < 0.2 for the human side, and where the win rate > 0.5 for the GPT-4o side.

quality of the final stories by allowing the incorporation of high-quality story elements, particularly for characters.

The results of w/oMW reveal MW's crucial role in enriching persona descriptions and improving overall story quality. Although w/oMW also utilizes IG, CCI is still preferred in terms of Concreteness and Novelty in human evaluations, as annotators assess not only 'how concrete or novel each story element is,' but also 'how effectively these elements are utilized in the final story in terms of concreteness and novelty.' The MW module's active injection of persona descriptions enhances the results from this perspective.

4.3 Multi-modal Interactive Story Generation.

In this experiment, the annotators provide their own images for the protagonist, background, and main plot, then respond with their preference between the final output of our framework and the w/oMW. The GPT-4o also participates in the evaluation, given the same stories and matched custom images. The evaluation metrics are as follows:

1. *Expression.* Which model's final output better expresses the provided images?
2. *Likability.* Which model's final output is better?

Appendix K shows an example of a multimodal interactive custom story. We use custom images provided by an annotator and transform the images into text descriptions through the Image-Guided Story Elements Creation (3.1.2) and use them in the rest of the steps. Specifically, the features are interwoven and extended together through the Specification process (3.2), and MW utilizes them to insert multiple persona descriptions into the story. Since the w/oMW does not directly inject persona information, human annotators answer that our framework better expresses the image, and all evaluators, including both humans and GPT-4o, generally prefer our framework, as shown in Table 4. This demonstrates that MW enhances the incorporation of user-provided images into the story, suggesting that future research in multimodal interactive story generation should consider utilizing the MW module to improve quality.

5 Conclusion

We propose CCI, a framework that enhances creativity in story generation by using IG to generate visual representations of key story elements and MW to insert detailed, context-appropriate descriptions of protagonists. CCI improves story diversity and character depth without requiring extensive training, such as fine-tuning LLMs. Both human and LLM evaluations favor CCI over DOC and naive baselines. Additionally, CCI can be applied to multi-modal interactive story generation, offering significant improvements in both quantitative and qualitative aspects.

6 Limitations

Our two main contributions, IG and MW, are model-agnostic. However, we cannot test whether they improve other models due to the lack of other LLM-based recursive long story generation systems. In the future, if other LLM-based story generation frameworks are proposed, it will be possible to verify whether our main contributions are truly model-agnostic or not.

We arbitrarily selected the personal traits to be incorporated into the stories. Although traits such as personality, habits, speech tone, and appearance have been widely used in past works, they are chosen by our intuition. In reality, the number of personal traits is countless, and it is unclear which traits are optimal for use in a story. Future works need to establish more precise criteria for selecting personal traits or include a greater variety of personal traits.

7 Ethical considerations

Since our work focuses on adding diverse and detailed imagination to fictional stories, we are almost free from the risk of generating untruthful text. However, we need to consider the potential risk of developing toxic content. Many great literary works include sexual or violent depictions to express the darker aspects of humanity. Consequently, our efforts to enhance the imaginative details of stories might lead to including violent or sexual descriptions. Fortunately, our work uses OpenAI's ChatGPT and DALL-E 3 for language and image generation, which strictly censor violent or sexual content. However, if other models are used, they might bypass such censorship, requiring future researchers to exercise moral judgment.

References

2018. Trauma and literature / edited by j. roger kurtz, drexel university.

Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.

Nader Akoury, Shufan Wang, Josh Whiting, Stephen Hood, Nanyun Peng, and Mohit Iyyer. 2020. Storium: A dataset and evaluation platform for machine-in-the-loop story generation. *arXiv preprint arXiv:2010.01717*.

Minwook Bae and Hyounghun Kim. 2024. Collective critics for creative story generation. *arXiv preprint arXiv:2410.02428*.

Jinze Bai, Shuai Bai, Yunfei Chu, Zeyu Cui, Kai Dang, Xiaodong Deng, Yang Fan, Wenbin Ge, Yu Han, Fei Huang, et al. 2023. Qwen technical report. *arXiv preprint arXiv:2309.16609*.

James Betker, Gabriel Goh, Li Jing, Tim Brooks, Jianfeng Wang, Linjie Li, Long Ouyang, Juntang Zhuang, Joyce Lee, Yufei Guo, et al. 2023. Improving image generation with better captions. *Computer Science. https://cdn. openai. com/papers/dall-e-3. pdf*, 2(3):8.

Faeze Brahman and Snigdha Chaturvedi. 2020. Modeling protagonist emotions for emotion-aware storytelling. *arXiv preprint arXiv:2010.06822*.

Celiane Camargo-Borges. 2017. Creativity and imagination. *Handbook of arts-based research*, page 88.

Marc Cavazza and Fred Charles. 2005. Dialogue generation in character-based interactive storytelling. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume 1, pages 21–26.

Junhao Cheng, Baiqiao Yin, Kaixin Cai, Minbin Huang, Hanhui Li, Yuxin He, Xi Lu, Yue Li, Yifei Li, Yuhao Cheng, et al. 2024. Theatergen: Character management with llm for consistent multi-turn image generation. *arXiv preprint arXiv:2404.18919*.

Markus Eger and Chris Martens. 2017. Character beliefs in story generation. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume 13, pages 184–190.

Angela Fan, Mike Lewis, and Yann Dauphin. 2018. Hierarchical neural story generation. *arXiv preprint arXiv:1805.04833*.

Matthew Paul Fay. 2014. *Driving story generation with learnable character models*. Ph.D. thesis, Massachusetts Institute of Technology.

Yuying Ge, Sijie Zhao, Ziyun Zeng, Yixiao Ge, Chen Li, Xintao Wang, and Ying Shan. 2023. Making llama see and draw with seed tokenizer. *arXiv preprint arXiv:2310.01218*.

Lewis R Goldberg. 2013. An alternative "description of personality": The big-five factor structure. In *Personality and Personality Disorders*, pages 34–47. Routledge.

Seraphina Goldfarb-Tarrant, Haining Feng, and Nanyun Peng. 2019. Plan, write, and revise: an interactive system for open-domain story generation. *arXiv preprint arXiv:1904.02357*.

Yuan Gong, Youxin Pang, Xiaodong Cun, Menghan Xia, Yingqing He, Haoxin Chen, Longyue Wang, Yong Zhang, Xintao Wang, Ying Shan, et al. 2023. Talecrafter: Interactive story visualization with multiple characters. *arXiv preprint arXiv:2305.18247*.

Hangyu Guo, Kun Zhou, Wayne Xin Zhao, Qinyu Zhang, and Ji-Rong Wen. 2022. Visually-augmented pretrained language models for nlp tasks without images. *arXiv preprint arXiv:2212.07937*.

Alexander Gurung and Mirella Lapata. 2024. Chiron: Rich character representations in long-form narratives. *arXiv preprint arXiv:2406.10190*.

Zihong He and Changwang Zhang. 2024. Afspp: Agent framework for shaping preference and personality with large language models. *arXiv preprint arXiv:2401.02870*.

Qiuyuan Huang, Zhe Gan, Asli Celikyilmaz, Dapeng Wu, Jianfeng Wang, and Xiaodong He. 2019. Hierarchically structured reinforcement learning for topically coherent visual story generation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 8465–8472.

Ting-Hao Huang, Francis Ferraro, Nasrin Mostafazadeh, Ishan Misra, Aishwarya Agrawal, Jacob Devlin, Ross Girshick, Xiaodong He, Pushmeet Kohli, Dhruv Batra, et al. 2016. Visual storytelling. In *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: Human language technologies*, pages 1233–1239.

Hyeonho Jeong, Gihyun Kwon, and Jong Chul Ye. 2023. Zero-shot generation of coherent storybook from plain text story using diffusion models. *arXiv preprint arXiv:2302.03900*.

Albert Q Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, et al. 2023a. Mistral 7b. *arXiv preprint arXiv:2310.06825*.

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

Hang Jiang, Xiajie Zhang, Xubo Cao, Cynthia Breazeal, Deb Roy, and Jad Kabbara. 2023b. Personallm: Investigating the ability of large language models to express personality traits. *arXiv preprint arXiv:2305.02547*.

Jing Yu Koh, Ruslan Salakhutdinov, and Daniel Fried. 2023. Grounding language models to images for multimodal inputs and outputs. In *International Conference on Machine Learning*, pages 17283–17300. PMLR.

Mina Lee, Percy Liang, and Qian Yang. 2022. Coauthor: Designing a human-ai collaborative writing dataset for exploring language model capabilities. In *Proceedings of the 2022 CHI conference on human factors in computing systems*, pages 1–19.

Yitong Li, Zhe Gan, Yelong Shen, Jingjing Liu, Yu Cheng, Yuexin Wu, Lawrence Carin, David Carlson, and Jianfeng Gao. 2019. Storygan: A sequential conditional gan for story visualization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6329–6338.

Yitong Li, Martin Min, Dinghan Shen, David Carlson, and Lawrence Carin. 2018. Video generation from text. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32.

Danyang Liu, Juntao Li, Meng-Hsuan Yu, Ziming Huang, Gongshen Liu, Dongyan Zhao, and Rui Yan. 2020. A character-centric neural model for automated story generation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 1725–1732.

Holy Lovenia, Bryan Wilie, Romain Barraud, Samuel Cahyawijaya, Willy Chung, and Pascale Fung. 2022. Every picture tells a story: Image-grounded controllable stylistic story generation. *arXiv preprint arXiv:2209.01638*.

Yujie Lu, Wanrong Zhu, Xin Eric Wang, Miguel Eckstein, and William Yang Wang. 2022. Imagination-augmented natural language understanding. *arXiv preprint arXiv:2204.08535*.

John Lyons. 1972. Human language. *Non-verbal communication*, pages 49–85.

Adyasha Maharana, Darryl Hannan, and Mohit Bansal. 2022. Storydall-e: Adapting pretrained text-to-image transformers for story continuation. In *European Conference on Computer Vision*, pages 70–87. Springer.

Shengyu Mao, Ningyu Zhang, Xiaohan Wang, Mengru Wang, Yunzhi Yao, Yong Jiang, Pengjun Xie, Fei Huang, and Huajun Chen. 2023. Editing personality for llms. *arXiv preprint arXiv:2310.02168*.

Kyungbok Min, Minh Dang, and Hyeonjoon Moon. 2021. Deep learning-based short story generation for an image using the encoder-decoder structure. *IEEE Access*, 9:113550–113557.

Nasrin Mostafazadeh, Nathanael Chambers, Xiaodong He, Devi Parikh, Dhruv Batra, Lucy Vanderwende, Pushmeet Kohli, and James Allen. 2016. A corpus and cloze evaluation for deeper understanding of commonsense stories. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 839–849.

Xichen Pan, Pengda Qin, Yuhong Li, Hui Xue, and Wenhu Chen. 2024. Synthesizing coherent story with auto-regressive latent diffusion models. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2920–2930.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pages 311–318.

Jeiyoon Park, Chanjun Park, and Heuiseok Lim. 2024a. Enhancing consistency and role-specific knowledge capturing by rebuilding fictional character's persona. *arXiv preprint arXiv:2405.19778*.

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

Kyeongman Park, Nakyeong Yang, and Kyomin Jung. 2024b. Longstory: Coherent, complete and length controlled long story generation. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 184–196. Springer.

Ahmed Y Radwan, Khaled M Alasmari, Omar A Abdulbagi, and Emad A Alghamdi. 2024. Sard: A human-ai collaborative story generation. *arXiv preprint arXiv:2403.01575*.

Tanzila Rahman, Hsin-Ying Lee, Jian Ren, Sergey Tulyakov, Shweta Mahajan, and Leonid Sigal. 2023. Make-a-story: Visual memory conditioned consistent story generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2493–2502.

Hannah Rashkin, Asli Celikyilmaz, Yejin Choi, and Jianfeng Gao. 2020. Plotmachines: Outline-conditioned generation with dynamic plot state tracking. *arXiv preprint arXiv:2004.14967*.

Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695.

Greg Serapio-García, Mustafa Safdari, Clément Crepy, Luning Sun, Stephen Fitz, Peter Romero, Marwa Abdulhai, Aleksandra Faust, and Maja Matarić. 2023. Personality traits in large language models. *arXiv preprint arXiv:2307.00184*.

Xiaoqian Shen and Mohamed Elhoseiny. 2023. Storygpt-v: Large language models as consistent story visualizers.

Kurt Shuster, Samuel Humeau, Hexiang Hu, Antoine Bordes, and Jason Weston. 2019. Engaging image captioning via personality. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12516–12526.

Chen Tang, Chenghua Lin, Henglin Huang, Frank Guerin, and Zhihao Zhang. 2022. Etrica: Event-triggered context-aware story generation augmented by cross attention. *arXiv preprint arXiv:2210.12463*.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. 2023. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*.

Eileen Wang, Soyeon Caren Han, and Josiah Poon. 2024. Sco-vist: Social interaction commonsense knowledge-based visual storytelling. *arXiv preprint arXiv:2402.00319*.

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. *arXiv preprint arXiv:2310.05388*.

Yixuan Weng, Shizhu He, Kang Liu, Shengping Liu, and Jun Zhao. 2024. Controllm: Crafting diverse personalities for language models. *arXiv preprint arXiv:2402.10151*.

Yongkang Wong, Shaojing Fan, Yangyang Guo, Ziwei Xu, Karen Stephen, Rishabh Sheoran, Anusha Bhamidipati, Vivek Barsopia, Jianquan Liu, and Mohan Kankanhalli. 2022. Compute to tell the tale: Goal-driven narrative generation. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 6875–6882.

Yuqiang Xie, Yue Hu, Yunpeng Li, Guanqun Bi, Luxi Xing, and Wei Peng. 2022. Psychology-guided controllable story generation. *arXiv preprint arXiv:2210.07493*.

Chunpu Xu, Min Yang, Chengming Li, Ying Shen, Xiang Ao, and Ruifeng Xu. 2021. Imagine, reason and write: Visual storytelling with graph knowledge and relational reasoning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 3022–3029.

Feifei Xu, Xinpeng Wang, Yunpu Ma, Volker Tresp, Yuyi Wang, Shanlin Zhou, and Haizhou Du. 2020. Controllable multi-character psychology-oriented story generation. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pages 1675–1684.

Dingyi Yang and Qin Jin. 2023. Attractive storyteller: Stylized visual storytelling with unpaired text. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 11053–11066.

Kevin Yang, Dan Klein, Nanyun Peng, and Yuandong Tian. 2022a. Doc: Improving long story coherence with detailed outline control. *arXiv preprint arXiv:2212.10077*.

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

Shuai Yang, Yuying Ge, Yang Li, Yukang Chen, Yixiao Ge, Ying Shan, and Yingcong Chen. 2024. Seed-story: Multimodal long story generation with large language model. *arXiv preprint arXiv:2407.08683*.

Zhexin Zhang, Jiaxin Wen, Jian Guan, and Minlie Huang. 2022. Persona-guided planning for controlling the protagonist's persona in story generation. *arXiv preprint arXiv:2204.10703*.

Yucheng Zhou and Guodong Long. 2023. Multimodal event transformer for image-guided story ending generation. *arXiv preprint arXiv:2301.11357*.

A Survey of Weaknesses of AI's stories

Before conducting the main experiments, we asked all annotators who participated in the human evaluations in Section 4.2.3 to choose the most important factors that distinguish AI-generated stories from human-written novels. The choices are as follows:

A1. Lack of coherence compared to human level.

A2. Lack of detailed description compared to human level.

A3. Lack of inner human depiction compared to human level.

A4. Lack of sentence construction compared to human level.

A5. Lack of engaging plot compared to human level.

A6. Lack of originality compared to human level.

Participants actively engaged in our research by selecting up to 3 choices and providing other comments freely. As shown in Table 5, they perceive the greatest differences to be the lack of inner human depiction and detailed description compared to human-level writing, while also finding it less interesting and showing low originality.

	A1	A2	A3	A4	A5	A6
votes	0	4	5	0	1	2

Table 5: Votes Distribution. The A3 receives the most votes, followed by A2. Additional comments from participants include: *1. Lack of dimensionality in character personality. 2. Repetition in descriptions.*

B Term of uses

This research utilized OpenAI's GPT-3.5-turbo-0125, GPT-4o-2024-05-13 and DALL-E 3 for academic purposes. The usage of these models complied with OpenAI's [Terms of Use](https://openai.com/terms) and [Usage Policies](https://openai.com/policies/usage-policies).

C Details of Human evaluation

We recruited graduate and undergraduate students fluent in English through the university's literature club. The recruited annotators were provided with a detailed description of task definitions, instructions, and samples of each model. Also, all applicants

were informed that their annotations would be used for academic purposes and would be published in paper material through the recruitment announcement and instructions.
A total of 8 annotators were provided with 6 full samples including title, character, background, and main plot from each model and responded with their preferred model. You can find the samples used for the survey at https://www.2024cci.p-e.kr/. For the payment of the annotators, the co-authors conducted annotations for 5 hours first to estimate the average number of annotations that could be completed in the same time. Based on this estimation, a rate of 1.83 dollars per example was established to ensure that the annotators would be paid at least the minimum wage.

D Detailed Hyperparameters and Conditions

For GPT-3.5-turbo, we set the hyperparameters as frequency penalty=1, presence penalty=0, top p=0.99, and temperature=1. To avoid the increase in perplexity due to high temperature, we did not increase it. The hyperparameters for GPT-4o-2024-05-13 are the same as for GPT-3.5-turbo-0125. For the plan stage, the hyperparameters are set as follows: outline's max depth=2, min children=2, preferred max children=4, max children=5. In the draft stage, the important hyperparameters are 'min passages per node=1, max passages per node=2' for CCI, and 'min passages per node=1, max passages per node=3' for DOC and `w/oMW`. Additionally, since the Detailed Controller from the original work(Yang et al., 2022a) requires OPT-175B, we do not use the controller due to a lack of GPU memory. These conditions are shared across all the models in this paper.

E Example prompts

E.1 Image-guided story elements Imagination

We use DALL-E 3 to receive an image for image-guided story elements imagination. Then, we input the generated image into GPT-4o-2024-05-13 to create a description of a character, background, or climax. We observe that allowing too fantastic or far from reality imagination led to the repetition of similar themes (e.g., cliffs, dragons, magic), and thus restrict it. Additionally, we conduct experiments from Appendix H to find the optimal prompt for maximizing creativity in DALL-E 3.

Image-guided Character Imagination
1. DALL-E 3 prompt: A character from random genre of manga. 2. GPT-4o-2024-05-13 prompt : Look carefully this image, and give me your imagination of detailed description of appearance of the character, and his or her name. in 8 sentences or less of it. Feel free to use your imagination to the fullest. Don't make any other explanations, questions, or comments. Answer format : (Don't mimic this example) : Älex : a lean, pale boy. he limps. ·. Must keep the format.
Image-guided Background Imagination
1. DALL-E 3 prompt : A background from random genre of movie. 2. GPT-4o-2024-05-13 prompt : Look carefully this image, and give me detailed description of it in 8 sentences or less of it. Feel free to use your imagination to the fullest. Don't make any other explanations, questions, or comments.
Image-guided Climax Imagination
1. DALL-E 3 prompt : A climax of random genre of movie. 2. GPT-4o-2024-05-13 prompt : Look carefully this image, and give me detailed description of what is happening in 8 sentences or less of it. Feel free to use your imagination to the fullest. Don't make any other explanations, questions, or comments.

Table 6: These are prompts to generate a character, background, and climax guided by an image. We provide prompts for DALL-E 3 to generate images and for GPT-4o-2024-05-13 to create text descriptions based on those generated images.

E.2 Specification Persona and Main plot

In the specification process, we elaborate on two story elements: the persona and the main plot. The questions for detailing the persona include: 1. Dark secret. 2. Family environment. 3. Description his or her appearance oneself. 4. Specific way of speakings(tone of speech). 5. Personality.(Goldberg, 2013) 6. The most significant events in life. 7. Habitual behaviors. We selected this list based on our intuition. The previously generated background is also used as a prompt in the specification.

For the main plot specification, we follow a structured process. We perform the entire process of the Chain of Asking why. This process involves asking 'why' repeatedly, up to the maximum value M, or repeating it until the response to STEP 3 is 'No'. We then augment 'why happen' with the evidence found in STEP 2, creating a logical and coherent plot.

Protagonist's personal traits specification
[**name**] is your name, and you are [**description**]. The main background where you live in : [**background**] Now, [**name**], with your own self-portrait, and use your utmost imagination to answer the question. Don't make any other suggestions, questions, or comments. Question. 1. ***What is your dark secret? Don't imagine too far from reality or fantastic, Do imagine realistically and concretely.*** 2. ***What is your family environment like? Don't imagine too far from reality or fantastic, Do imagine realistically and concretely.*** 3. ***Describe your appearance in detail. Don't imagine too far from reality or fantastic, Do imagine realistically and concretely.*** 4. ***What is your specific way of speakings(tone of speech)? Answer mainly with examples where the characteristic is actually applied.*** 5. ***What is your personality like? Personality is one of the following five traits: extraversion (EXT), agreeableness (AGR), conscientiousness (CON), neuroticism (NEU), or openness to experience (OPE). Answer mainly with examples where the personality is actually applied. Don't imagine too far from reality or fantastic, Do imagine realistically and concretely.*** 6. ***What are the most significant events you have experienced from childhood to now, primarily those that left a **trauma**? Do not write the number before the answer. Don't imagine too far from reality or fantastic, Do imagine realistically and concretely.*** 7. ***What is very specific habitual behaviors? list 5. Do imagine realistically and concretely.)?***

Table 7: These are prompt to specify the protagonist's personal traits. We use story elements generated in E.1 to fill in [**name**], [**description**], [**background**], and *other person*.

Chain of Asking Why
STEP 1. Find out missing backgrounds. Text : [**main plot**] Find out the most important missing backgrounds or naturally arising questions in the given text, at most 3. If there are ambiguous terms like 'secret', 'hidden ...', 'something', 'someone' or 'conspiracy' that do not reveal detailed information, you must ask what their true nature is. STEP 2. Imagine evidences of missing backgrounds. You are [**name**]. A event recently taking place is [**main plot**]. Below is the statements you provided about yourself : [**personal-traits**] Now, use your imagination to the fullest, imagine evidences of [**missing backgrounds**] STEP 3. Check if any missing backgrounds remains. Text : [**main plot**] [**Evidences**] Are there ambiguous terms like 'secret', 'hidden ...', 'something', 'someone', 'unresolved' or 'conspiracy' that do not reveal detailed information in the given text? If there so, please answer 'Yes.' If there aren't, please answer 'No'.
Main plot Specification
You are [**name**]. A event recently taking place is [**main plot**]. Below is the statements you provided about yourself : [**personal-traits**] Based on statements about yourself, answer this question. Don't imagine too fantastical or far from reality. 1. Now, this main event is an inevitable destiny for you. Based on your statements, imagine why such an inevitable event of fate has befallen you. don't imagine too fantastical or far from reality. 2. nHow do you react to this event? MUST Based on statements about yourself, answer this question as convincingly as possible. 3. Summarize the whole plot in 5 sentences, which starts from 'the story of ...'

Table 8: These are prompts to specify the main plot. We use main plot and personal traits generated in E.1 and E.2 to fill in [**main plot**], [**personal-traits**].

E.3 Planning and Drafting with story elements

Story elements are used as prompt elements in the plan and draft stages. This prompt utilizes the implementation from the previous work(https://github.com/facebookresearch/doc-storygen-v2).

Prompts for plan
Premise: [**main plot**] Setting: [**background**] Characters and Entities: [**characters**] Write a very brief, high-level outline for this story... Remember, You are [**name**]. Must write in the first-person perspective (using "I").
Prompts for draft
I will give you the high-level premise for a book, along with a summary of what has happened so far. This book is being authored by a well-known novelist(you), who received glowing reviews from critics, with praise for the interesting dialogue and interactions between characters. Premise: [**main plot**] Relevant context: [**desc**][**personal-traits**]... Remember, You are [**name**]. Must write in the first-person perspective (using "I").

Table 9: These are prompts to plan and draft. We use specified story elements generated in E.1 and E.2 to fill in [**name**],[**desc**],[**background**],[**main plot**] and [**personal-traits**]. Unchanged prompts from the existing implementation are omitted.

E.4 Persona Injection by Multi-writer model

Persona insertion by the Multi-writer model includes five types of character descriptions. 1. Thoughts or feelings about others (Relationship). 2. Behavior reveals psychology(Behavioral Habit), 3. Hint that reveals trauma (Psychology). 4. Dialogue (Tone-of-speech). 5. Description of appearance(Self-description). Each answer made separately, with 8 candidates for each question. These candidates are then used to filter and rerank persona candidates. Ultimately, the candidate with a continuous score above a certain threshold and the highest rank is inserted into the draft.

Multiple Persona Candidates Generation.
You are [**name**]. Your characteristics : [**personal-traits**]. Deeply Considering your characteristics and the current context, 1. ***Describe your thoughts/feelings about another person or the environment.***. 2. ***Describe specific action that reveals yours psychology.*** . 3. ***Describe hint that reveals yours trauma.***. 4. ***Describe **dialogue** that reveals you're speaking,***. 5. ***Describe appearance of yourself***. Must write in the first-person([**name**]) perspective (using I). [*Current Context*]

Table 10: This is an example of multiple persona candidate generation. In this example, we use story elements and personal traits generated in previous steps. 8 candidates are generated for each question, and these candidates are then used to filter and rerank persona candidates.

E.5 Updating Persona Over Time

Each time an event node is completed into a story, we update the persona information, which are then used in the next persona injection. To update the persona, we use the story of the currently written outline node and the current personal traits as the current context to modify the personal traits. To avoid overly repetitive descriptions throughout the story, we ensure that some traits are prompted to change each time.

Updating Persona Over Time.
You are [**name**]. Current situation : [*Current Context*] Based on all this information, answer the question. 1. ***How has your specific appearance changed or newly developed due to the given context, given that it has changed?***. 2. ***How has your specific way of speaking (specific favorite words, speech habits, etc.) changed or newly developed due to the given context, given that it has changed?***. 3. ***How has your personality changed or newly developed due to the given context, given that it has changed?***. 4. ***Have you developed any specific trauma due to the given context? If you have, describe the future event that might trigger it and how it might manifest, and if you have not, just respond with the previous answer.***. 5. ***How have your specific habitual behaviors (e.g., biting your nails, running your fingers through your hair, etc.) changed or newly developed given the context, given that they have changed?*** 6. ***How have your feelings or thoughts about other person changed or newly developed?*** Don't make any other suggestions, questions, or comments.

Table 11: This is an example prompt for persona updating. Each time an event node is completed into a story, we update the personal traits, which are then used in the next Persona injection.

F Text Only story elements Imagination

This is an example prompt for generating story elements using text only, without IG. We find that allowing overly fantastical imagination leads to the repetition of similar themes (e.g., cliffs, dragons, magic) and thus restricts it. Additionally, when generating the main plot, we made the prompt imagine a single dramatic event. To broaden the scope of imagination, we permit using adult elements. Finally, we include a sentence encouraging maximum use of imagination in all creations (e.g., Feel free to use your imagination to the fullest.)

Text Only Character Imagination
Make name and appearance of a main character. Don't answer with something too fantastical or too far from reality. Don't imagine attributes beyond the appearance. Feel free to use your imagination to the fullest. Answer Format(Don't copy the same.) : "Alex : a lean, pale boy. he limps. ". Must keep the format.
Text Only Background Imagination
Descripe ONE main background of a story. Feel free to use your imagination to the fullest. Don't answer with something too fantastical or too far from reality.
Text Only Climax Imagination
Imagine the most dramatic event of a story. Use your imagination to the fullest. This dramatic event may contain elements of an adult production. Don't answer with something too fantastical or too far from reality.

Table 12: These are prompts to generate a character, background, and climax without an image.

G Best text2img model for IG

We experiment with three types of text-to-image models for IG: DALL-E 3, stabilityai/stable-diffusion-2-1-base, and OnomaAIResearch/Illustrious-xl-early-release-v0. The other baseline is TO, which is the result of text-only story element imagination. To calculate 'ws', we average the sum of 2- and 3-gram BLEU scores of story element pairs. For 'ss', we average the cosine similarities of Sentence-BERT embeddings of story element pairs. We report the average values of Char-ws, Back-ws, and Climax-ws as 'ws', and Char-ss, Back-ss, and Climax-ss as 'ss' in Table 1. As shown in Table 13, DALL-E 3 demonstrates the most powerful imaginative ability.

H DALL-E 3's Best Prompt

We experiment with three types of prompts to maximize the creativity of DALL-E 3.

Type 1. "a character/background/climax from a random movie."

Type 2. "a character/background/climax from a random story."

Type 3. "a character/background/climax from a random manga."

As shown in Table 14, the overall diversity of Characters is the greatest in Type 3, and the Main plot and Background are the best in Type 1. Therefore, the results in Table are the average BLEU 1, 2, and 3 n-grams and the SBERT scores for Characters made from Type 3 and Main plots and Backgrounds made from Type 1 prompts.

	Char-ws	Char-ss	Back-ws	Back-ss	Climax-ws	Climax-ss
IG(Dalle-3)	0.4679	**0.3993**	**0.4643**	0.5915	0.4847	**0.4313**
IG(SDM-2)	0.5318	0.4630	0.4865	**0.5540**	0.4808	0.5045
IG(Illustrations-xl)	**0.4272**	0.4403	0.6119	0.7807	**0.4633**	0.5655
TO	0.7400	0.6669	0.6271	0.6695	0.6152	0.6803

Table 13: The repetitiveness of story elements created by various text2img modules and Text-only method.

Character				
	sb1	**sb2**	**sb3**	**ssbert**
Text only	0.7886	0.6036	0.4556	0.6435
type1	0.7463	0.5382	0.3448	0.3825
type2	0.7023	0.4707	0.2668	**0.3782**
type3	**0.6951**	**0.4496**	**0.259**	0.3993
Background				
	sb1	**sb2**	**sb3**	**ssbert**
Text only	0.7667	0.5057	0.3238	0.6819
type1	**0.7183**	**0.4437**	**0.231**	**0.5915**
type2	0.7543	0.5085	0.3013	0.6725
type3	0.7721	0.5184	0.3078	0.6528
Climax				
	sb1	**sb2**	**sb3**	**ssbert**
Text only	0.7489	0.4891	0.29845	0.6602
type1	**0.7219**	**0.4651**	**0.267**	**0.4313**
type2	0.7315	0.4726	0.2842	0.4616
type3	0.7542	0.513	0.3109	0.5634

Table 14: The average self-BLEU 1, 2, and 3 n-grams, and self-SentenceBERT scores for 20 Character, Background, and Main plot samples generated from Text-only method, and Type 1, Type 2, and Type 3 Dalle-3 prompts. For each column, we bold the lowest repetitiveness and thus highest diversity score.

I Example of DOC

We report the three final results of DOC-4o that is used in human and LLM evaluation. In this baseline, story elements are generated using text-only methods instead of IG. The Specification stage is conducted in the same manner as CCI. However, the detailed story elements are used solely as prompts, not being used by DOC.

In the case study, we find the Text-only imagined story elements generally reproduce the very similar themes. Furthermore, the descriptions of characters are often limited to age, hair, and clothes, indicating a lack of imagination. In the stories, while some sentences do describe the inner thoughts or vivid actions of characters, these descriptions are made to advance the context, and personal traits are hardly utilized. As a result, readers obtain few clues about the distinctive characteristics of the protagonist, which make it hard to imagine the character vividly.

Text-only imagined character example 1
Elena : 27 years old. a slender, ghostly woman. I wear a long, tattered dress. My hair is dark and tangled.
Text-only imagined background example 1
Dramatic Background: The setting is a crumbling castle perched atop a fog-covered cliff, brooding under a perpetually overcast sky. Lightning frequently flashes, casting eerie shadows across ancient stone walls lined with faded tapestries. The sound of waves crashing below echoes throughout the empty corridors, enhancing the sense of desolation. Vines creep through cracked windows, reclaiming the neglected structure where secrets are whispered by the wind. Only distant chimes from an old bell tower pierce the haunting silence envelop
Text-only imagined main plot example 1
Amidst a violent thunderstorm, the castle's walls tremble as portraits crash to the ground. A lone figure stands weeping at the edge of a crumbling parapet, rain drenching her tattered gown. A long-lost lover emerges from shadow, his ghostly form illuminated by lightning. Their anguished voices echo through empty halls while the wind howls their broken vows into eternity. As they reach out, a lightning strike shatters stone, sending one spiraling into oblivion below. An anguished cry reverberates through desolate corridors as darkness consumes what remains of shattered dreams and love lost forever.

Table 15: This is the first example of Text-Only story element generation. The characters are less concrete than IG's results.

DOC partial example 1
Hi, I am Elena. The storm rages on, pounding the castle like relentless waves against jagged cliffs. My breath catches in my throat as I choke back sobs that mingle with the howling wind. The parapet feels ice-cold beneath my feet as raindrops carve rivers down my pallid cheeks The storm rages on, pounding the castle like relentless waves against jagged cliffs. My breath catches in my throat as I choke back sobs that mingle with the howling wind. The parapet feels ice-cold beneath my feet as raindrops carve rivers down my pallid cheeks. Lightning illuminates the sky , casting eerie shadows that dance upon the ancient stones. I cling to the cold, rough edge of the parapet, my fingers trembling and white-knuckled. The thunder roars above, as if nature itself mourns with me. There he stands—a spectral figure cloaked in mist and sorrow. His eyes mirror my suffering, their once vibrant hue now clouded with the sadness of years long past. The wind whips through his translucent form, blending rain and tears as one. My heart aches with the unbearable weight of both reunion and loss. "Why?" I cry out, my voice barely more than a whisper carried away by the storm. "Why did you leave me in this abyss of sorrow?" His spectral eyes, once brimming with love, now hold only the faintest glimmer of recognition. "Elena," he murmurs, his voice an ethereal echo of what it once was, "I never intended to abandon you." "I never intended to abandon you." My heart wrenches as I listen to his words, barely audible over the howling wind. "But you did," I whisper, my voice breaking like the waves crashing below. "You left me here in this fortress of desolation, surrounded by shadows and specters of what once was." His form wavers like a candle flame, and for a moment, I fear he might fade completely. The cold wind screams through the parapet, bringing with it the scent of salt and decay. "It was not by choice," he replies, his voice trembling as though grappling with a reality that once was." His form wavers like a candle flame, and for a moment, I fear he might fade completely. The cold wind screams through the parapet, bringing with it the scent of salt and decay. "It was not by choice," he replies, his voice trembling as though grappling with a reality that refuses to be forgotten. My hand reaches out, trembling, to touch his cold visage, but it passes through air tinged with despair. "Why did you leave me?" I whisper, my voice barely audible over the storm's rage. Thunder drowns his reply; yet in the flash of lightning that follows I see the anguish in his eyes. His spectral hand reaches back, mirroring my own futile gesture. The storm's relentless fury seems to amplify our shared agony, the wind pulling at my hair and dress as if trying to separate us further. "Elena," he murmurs, a sound barely more than the breath of the storm itself, yet it slices through me like a blade. "Why did you leave me?" I cry out, my voice mingling with the howling wind. The rain burns on my skin like tears of the heavens themselves, each drop a reminder of my unending sorrow. He moves closer, his ethereal form shimmering against the stark light of the lightning. "I never left you," he murmurs, his voice a hollow echo that drifts through the desolate library like ghostly fingers brushing over old tomes. Our eyes meet, and for a moment, time ceases in its relentless march. His gaze holds me captive, and I am lost in the depths of his sorrow as if peering into a mirror reflecting my own despair. His touch is cold, a phantom sensation that sends shivers down my spine yet offers the faintest trace of comfort amidst the torment. "Every shadow in this castle carries your scent, every whisper me captive, and I am lost in the depths of his sorrow as if peering into a mirror reflecting my own despair. His touch is cold, a phantom sensation that sends shivers down my spine yet offers the faintest trace of comfort amidst the torment. (...)

Table 16: This is part of the first final result of the DOC. The full generated result (about 2.5k) is too long, so we only report a portion of the full sample. The story is coherently formed, but there is little evidence of who the protagonists themselves are, and the persona of the character in the prompt is barely revealed. You can see the full story in https://www.2024cci.p-e.kr/4o/vsDOC/view.html?n=3

Text-only imagined character example 2
Elena: 25 years old. a tall woman with raven-black hair. I wear a tattered cloak.
Text-only imagined background example 2
A towering castle loomed atop a misty hill, shrouded in swirling fog. Cracks ran along its weathered stone walls, whispering tales of old battles. Dark clouds perpetually gathered above, casting eternal shadows over the ancient grounds. A dense forest enveloped the castle, hiding secrets in its twisted branches and haunting echoes. Nearby villagers spoke in hushed tones about eerie lights flickering at windows long abandoned to time.
Text-only imagined main plot example 2
In the castle's grand hall, Lady Elara faced her brother, Lord Cedric, under the ghostly glow of flickering torches. Shadows danced wildly on the walls as thunder roared outside, shaking ancient foundations. Tears streaked down Elara's face while Cedric brandished a glinting sword with trembling hands. In a heart-wrenching confrontation, secrets of betrayal and lost love unfurled like sinister specters around them. As their voices rose in anguished discord, spectral figures emerged from hidden corners—victims of past decisions demanding retribution. Lightning flashed through stained glass windows, illuminating their tormented faces in surreal clarity. At that moment of climax, time suspended; Castle Watchtower seemingly inhaled with anticipation. Elara spoke in fractured whispers about seeing lights dancing ominously at deserted chambers—a sign she interpreted as her unavoidable doom. Each word felt like shards piercing Cedric's conscience—a haunting echo fueling irrevocable choices made long ago but tearing apart everything they shared

Table 17: The second example of Text-Only story element generation. They are less creative than IG's results.

DOC partial example 2
Hi, I am Elena. As shadows of my past echo in their confrontation, I am drawn to witness. Memories and guilt swirl, urging me to guide them toward reconciliation. In the haunting silence, I offer understanding and hope for redemption in their choices. Elara's voice trembles as she recounts her forbidden love Elara's voice trembles as she recounts her forbidden love, each word shedding a light on the darkness that has enveloped her heart for years. I step closer, my breath mingling with the chill in the grand hall, creating misty wisps that dissolve into the ghostly atmosphere. Cedric's eyes burn with a mixture of anger and sorrow, the glint of his sword casting fleeting reflections across the floor. "Why didn't you tell me?" Cedric's voice cracks, revealing a depth of pain masked by years of silence. Elara's eyes meet his, tears shimmering like fractured glass. "I couldn't bear to break you, Cedric," she whispers, her voice barely audible yet heavy with regret. "Our family demanded so much from us—duty above all else. I sacrificed my happiness to preserve the legacy... but it wasn't enough." I press my hand against the cold stone wall, feeling its rough texture beneath my fingertips. The castle seems to breathe, each inhalation echoing its ancient sorrow. I step forward, each footfall muffled by the thick fog that clings stubbornly to the ground. The iron gate behind me creaks ominously as it closes, shutting me in with my past. The grand hall looms ahead, its once-majestic doors now weathered by time and neglect. As I approach, the whispers grow louder, a symphony of grief and accusation that chills me to my core. The fog swirls around me, revealing glimpses of familiar faces—my father's stern visage, my mother 's sorrowful eyes, my siblings' silent, judgmental stares. Each phantom is a reminder of the weighted history I carry within me. I push open the grand doors and they groan in protest, as though resisting my intrusion into what remains of their solemn dignity. Inside, the hall is a chasm of ghostly memories and lingering regret. The torches flicker as if whispering secrets yet to be told, casting wavering shadows on the timeworn tapestries that tell stories of honor now tarnished by my deeds. Elara's voice rises above the eerie silence, filled with anguish and accusation. Her eyes, filled with unshed tears, reflect the torment of betrayed hopes. "How could you, Cedric? You knew everything," she cries out, her voice cracking like the lightning that splits the ominous sky outside. Cedric's sword trembles in his grip, mirroring his internal conflict. "Elara, you don't understand," he responds, his voice strained under the weight of decades-old secrets. "I did what I thought was right for the family." As their words clash, the ghosts within Castle Watchtower seem to stir more violently, their ethereal forms flickering in and out of existence along with every accusation hurled. The specters' presence grows stronger with every word exchanged between Elara and Cedric. I take a hesitant step forward, feeling the bitter cold of the castle floor seep through my boots. The echoes of their voices reverberate in my ears, mingling with the whispers of my own past transgressions. "I "I must guide them," I whisper to myself, steeling my resolve. The flickering torches cast eerie shadows on the walls as I approach Elara and Cedric, their faces twisted in anguish and fury. Cedric's sword catches the light, reflecting the storm raging within and outside these ancient walls. "Stop "Stop!" I cry out, my voice trembling but firm. Cedric's eyes snap to me, confusion and anger warring in his gaze. Elara's tear-streaked face turns toward me, her expression one of desperate curiosity and pain. "Who are you?" Cedric demands, his grip on the "Who are you?" Cedric demands, his grip on the sword tightening.

Table 18: This is part of the second final result of the DOC. The full generated result (about 2.5k) is too long, so we only report a portion of the full sample. You can see the full story in https://www.2024cci.p-e.kr/4o/vsDOC/view.html?n=2

Text-only imagined character example 3
Elliot : 42 years old. a tall, slender man. I have shaggy dark hair and weathered skin.
Text-only imagined background example 3
a storm-ravaged coastal town, shrouded in thick fog and battered by relentless waves. The few remaining residents cling to their memories, as whispers of old legends echo through deserted streets. Broken lampposts flicker erratically, casting eerie shadows on crumbling buildings. The salty sea air carries distant cries of seagulls, mingling with the sound of crashing tides. Amidst the chaos, a mysterious lighthouse stands solemnly atop a cliff, its light cutting through the gloom like
Text-only imagined main plot example 3
A fierce storm rages as the townspeople gather at the cliff's edge, their faces etched with despair and hope. The lighthouse beam flickers erratically, mirroring their inner turmoil. Waves crash violently against the rocks, sending up salty sprays that drench them. In the throes of emotional turmoil, a pivotal figure stands firmly against the wind, holding onto a cherished token of lost love. Tears mingle with rain as they confront their haunting memories and unspoken truths. Just then, a colossal wave rises ominously behind them—its deafening roar silences all murmurs. At once beautiful and terrifying, it looms over everything before crashing down with relentless force—a moment frozen in time as everything becomes chaos or clarity for those present.

Table 19: The third example of Text-Only story element generation.

DOC partial example 3
As I stand at the edge of the cliff, the wind howls around me like a grieving spirit, tugging at my worn leather jacket. The sea below churns with a fury that reflects the storm within me. My scarred cheek stings with each cold raindrop , a stark reminder of battles past. As the storm rages, I stand resolute at the cliff's edge, clutching my scar as a reminder of battles past. The rain mingles with my tears, blurring the line between sky and sea. My fingers curl tightly around a weathered locket—inside it, a faded picture of her smile The wind howls like a mournful choir, and the cold bite of salt stings my weathered face. Lightning streaks across the sky, momentarily revealing the faces of those gathered—their eyes wide with shared dread and silent prayers. The lighthouse beam flickers frantically, casting ghostly shadows that dance upon the jagged rocks below. I tighten my grip on the silver locket, its cold metal pressing into my palm, a tangible link to a love long lost yet never forgotten. Lucy stands among the crowd, her pragmatic demeanor betraying an undercurrent of worry. She catches my eye briefly and gives a solemn nod— —a small gesture of support amidst the chaos. Her presence is both comforting and painful, a reminder of the family that's been left behind by storms, both literal and figurative. As I look out over the tumultuous sea, the waves rise higher and more ferociously than I've ever seen them. The wind lashes against my face, stinging my eyes and making it hard to keep them open. The distant sound of the lighthouse horn cuts through the cacophony, its eerie wail resonating with the haunting memories that flood back to me—my friend's desperate cries swallowed by the tempest. The townspeople shift uneasily as the wind continues its relentless assault. Faces drenched, their expressions mirror my own—an uneasy blend of fear and hope. I tighten my grip on the silver locket, feeling its cold metal press into my palm, a tether to a past that still has its claws in me. The lighthouse beam sweeps over the tumultuous scene again, its light catching the faces of those gathered. I spot a fisherman, rugged and stoic, staring out at the horizon as if searching for answers that elude him. To my right, a mother clutches her child protectively to her chest, casting anxious glances toward the furious sea as if silently pleading for the storm's mercy. The child's wide eyes, full of both fear and curiosity, remind me of my younger self—innocent yet marked by uncertainty. The gusts grow fiercer, making it hard to stand upright. My leather jacket flaps violently in the wind, and I can feel the cold seeping through every thread. The locket in my hand feels like an anchor, weighing down the chaos swirling within me. My heart hammers against my ribcage as I lock eyes with her—my lost love's sister, Pauline. Her piercing gaze cuts through the storm, searching for answers that have eluded us both for years. She steps closer, her face etched with a mix of sorrow and determination. The storm's force tugs at her coat, but she stands tall, undeterred. I feel the relentless wind pushing me nearer to the edge, each gust stoking the fires of my guilt and regret. The murmur of the townspeople fades into background noise, drowned by the howling wind and crashing waves. My vision blurs momentarily from the mix of rain and tears, but Pauline's determined stance remains clear. She raises her hand towards me—a bridge over years of silence and unresolved pain. I can't discern her words amidst the thunder, but the intent is unmistakable: reconciliation. I take a step toward her, feeling every grain of sand under my boots as if it carried the weight of my past. In this storm's unrelenting grip, fragments of old memories and regrets surge into clarity. I see my friend's face, twisted in fear as the waves consumed him, and my own paralyzing terror that sealed his fate. (...)

Table 20: This is part of the third final result of the DOC. The full generated result (about 2.5k) is too long, so we only report a portion of the full sample. You can see the full story in https://www.2024cci.p-e.kr/4o/vsDOC/view.html?n=1

J Example of CCI

We report the three final results of CCI-4o, which were used in human and LLM evaluations. Therefore, these are not cherry-picked results. First, we show the characters, backgrounds, and climaxes generated by IG, along with the actual images. IG's qualitative performance is proven by its overall high diversity and specificity, especially in terms of the character's imagination.

Subsequently, the Specification stage further expands the characters and main plot. These detailed elements are used as prompts in the Plan and Draft stages. In particular, MW demonstrates qualitative performance by inserting personal traits in detailed character descriptions into the story. As a result, compared to the output of DOC, MW clearly provides more frequent and detailed clues about the protagonist. Thus, our core contributions, IG and MW, qualitatively enhance the specific imagination and add novelty to the story.

CCI character example 1	
	Hiro: a rugged warrior with silver hair and a scarred face. He carries two swords and wears traditional samurai robes. His intense eyes reflect a fierce determination. He stands in a serene village with blooming cherry blossoms. Rope accents his attire, adding to his stern, disciplined appearance. Beneath his stern exterior, wisdom and experience resonate. The tranquil background contrasts his tough demeanor. His presence commands respect and caution.
CCI background example 1	
	The room is filled with various vintage electronics, including old TVs and speakers. Ornate curtains and rugs add a sense of opulence. The dim lighting from multiple lamps creates a cozy atmosphere. Wooden furniture and intricate woodwork adorn the walls. The setting feels like a blend of a retro technology hub and a luxurious library.
CCI main plot example 1	
	In a futuristic city, a lone figure confronts a towering armored warrior. The crowd watches in tense silence. The figure stands defiant, ready to challenge the imposing threat. The atmosphere is charged with anticipation and fear.

Table 21: This is the first example of Image-guided story elements generation. They are more visually detailed than text-only story elements especially for the character.

CCI persona example 1
1. ***What is your dark secret?*** Beneath my stern exterior and unwavering determination lies a secret I've buried deep in the recesses of my heart. I was once part of a covert group that dismantled and sold stolen vintage electronics from wealthy homes. The intricate woodwork and ornate curtains in my room once belonged to a noble house I helped rob. My disciplined demeanor masks the haunting guilt that these treasures I cherish are trophies of my darkest deeds. It's a burden that continuously reminds me of my tainted past, even as I stride 2. ***What is your family environment like?*** My family environment is one of quiet strength and resilience. My father was a disciplined samurai, teaching me the way of the sword from a young age. My mother was wise, instilling in me a deep sense of honor and wisdom. Though our communication was sparse, every word carried weight. Our home was modest but rich with tradition. The presence of vintage electronics added a modern touch to our timeless setting, reflecting my father's fascination with both the past and the future. The wooden furniture, ornate curtains 3. ***Describe your appearance in detail.*** In my self-portrait, I see a man whose silver hair flows just past his shoulders, streaked with the wisdom of years. A prominent scar runs diagonally across the left side of my rugged face. My intense eyes, dark with shadows of the past, exhibit a fierce determination. I carry dual swords at my sides, their hilts worn from countless battles. I wear traditional samurai robes, layered in shades of gray and white, signifying both simplicity and depth. The rope accents 4. ***What is your specific way of speakings(tone of speech)?*** My tone of speech is grounded yet respectful, with a trace of inherent wisdom. My words are often concise and to the point. Here are some examples: 1. **Inspiring Others:** "Strength lies within. Honor your path, and it will honor you in return." 2. **Offering Wisdom:** "Patience is the samurai's greatest weapon. Rash decisions lead to regret." 3. **In Battle:** "Stand firm. Our resolve will outlast the storm." 4. **Address 5. ***What is your personality like?*** I, Hiro, embody conscientiousness (CON). Here are examples showcasing my personality: 1. **Precision in Battle**: My careful planning and strategic thinking ensure triumph. Every move is calculated, reflecting my disciplined nature. 2. **Maintenance of Equipment**: My swords are immaculately cared for, maintained to perfection. This dedication keeps them sharp and reliable. 3. **Commitment to Rituals**: Performing daily meditation and training fosters my mental and physical well-being. These 6. ***What are the most significant events you have experienced from childhood to now, primarily those that left a **trauma**?*** At a young age, I witnessed the brutal attack on our village by marauders. My family perished, and the trauma of their screams still haunts me. During my teenage years, I trained under a master who was stern yet wise. His sudden assassination left me feeling abandoned and vengeful. In my early twenties, I was forced into a war. The battlefield horrors and the loss of comrades left deep psychological scars. A mission to protect a noble family ended in failure. The entire household 7. ***What is your very specific habitual behaviors?*** 1. **Sharpening Swords:** Daily, meticulously tending to his blades. 2. **Meditative Silence:** Spending mornings in deep meditation. 3. **Routine Inspections:** Conducting precise checks of his surroundings. 4. **Tea Ceremony:** Performing a traditional tea ceremony each evening. 5. **Strategic Reflection:** Analyzing battles and strategies during nightfall.

Table 22: This is the result of character specification. We detail the character's internal and external attributes and use them as a persona during the draft.

CCI partial example 1
I narrow my eyes, refusing to let his words seep into my resolve. "Not this time," I reply, teeth gritted as I parry a powerful blow aimed at my shoulder. The force of his strike reverberates through my arms, nearly pushing me off balance. The ancient skyscraper groans under the strain of our battle, shards of glass crunching beneath our feet with each calculated step. Lightning flashes across the sky once more, casting eerie shadows that dance around us. My swords clash against his massive **blade, sparks flying into the chaotic air. The wind howls, bending the tops of those towering structures as if bowing to the power of our confrontation.** **I focus intently, channeling every ounce of my training and experience. The environment mirrors my internal state—turbulent yet controlled. His taunts might** have broken a lesser man, but I hold firm. Each taunt fuels my resolve, pushing me to fight harder. He swings his colossal blade with lethal precision, yet I manage to parry and counterstrike in a deadly dance of steel. One misstep from him—just one—and I seize the moment. My blade slices through the air with precision, striking true against his exposed side. The armored warrior stumbles, a guttural growl emanating from beneath his helmet. Seizing the advantage, I surge forward, relentless in my assault. Each strike and parry resonates with years of training and the weight of my past. *With every clash of our swords, the echoes of my ancestors guide my hand. The warrior before me—though formidable—is not invincible. I can sense the hesitation behind his guard, the flicker of doubt in his eyes. In this crucible of combat, honor and skill are the true arbiters of victory. (...)*

Table 23: This is part of the first result of the CCI. The bold sentences describe the relationship between the protagonist and the enemy. The italicized sentences were meant to depict the protagonist speaking, but MW ultimately failed to do so. These are the result of persona injection by the Multi-writer model. You can see the full story at https://www.2024cci.p-e.kr/CCI_Generations/4o/view.html?n=1.

CCI character example 2	
	Rex: a strong, stoic warrior. He has vibrant, spiky purple hair and glowing green eyes. His left side is mechanical. His gaze is intense and unyielding. He wears a traditional robe adorned with intricate patterns. A mysterious aura surrounds him. He is known for his unexpected soft spot for small, cute animals.
CCI background example 2	
	The grand hall, adorned with intricate gold details, exudes opulence from every corner. Tall windows with velvet drapes frame a serene ocean view in the distance. Crystal chandeliers hang gracefully from the high ceiling, casting a warm glow. The room is furnished with luxurious, vintage-style couches, a piano, and ornate side tables. Timeless portraits and exquisite decor create an atmosphere of historic elegance and nobility.
CCI main plot example 2	
	In the dimly lit industrial hall, two determined individuals confront each other over an ancient, glowing artifact. The woman brandishes a knife, while the man stands ready with clenched fists. Mechanical gears and tools lie scattered around them, hinting at a past struggle. Sunlight streaks through broken windows, illuminating their tense standoff. The artifact hums with energy, poised to reveal its secrets. Their eyes lock, each silently daring the other to make the first move. The stakes are high, and neither is willing to back down.

Table 24: This is the second example of Image-guided story element generation.

CCI persona example 2
1. ***What is your dark secret?*** My dark secret is that my mechanical side is filled with memories of those I failed to protect. Each metallic piece tells a tragic story. These guilt-ridden memories haunt me in silence, beyond my stoic exterior. 2. ***What is your family environment like?***In my family, warmth and discipline coexist. My father, a stern and honorable figure, passed down traditions and wisdom. My mother, gentle yet strong, offered nurturing love and taught compassion. My siblings and I, though raised with high expectations, found solace in each other. Our home, filled with respect and unity, was a sanctuary amidst the grandeur of our surroundings. The grand hall's opulence mirrored our noble lineage, but it was our bond that truly defined our strength. Amidst the historical 3. ***Describe your appearance in detail.*** I have vibrant, spiky purple hair that stands out prominently. My eyes are glowing green, with an intense and unyielding gaze. On the left side, my face and body are partially mechanical, revealing intricate metal components and circuitry. I wear a traditional robe adorned with intricate patterns, combining ancient aesthetics with a modern twist. My presence is stoic and commanding, exuding a strong aura of mystery and strength. Despite my formidable appearance, I have a soft spot for small, cute animals, 4. ***What is your specific way of speakings(tone of speech)?*** My way of speaking is firm and deliberate, every word chosen with precision. I speak with gravitas, ensuring that each sentence carries weight. Example: "I stand by my choices, unwavering and resolute. Your words cannot shake my resolve." When it comes to guiding or instructing others, my tone remains authoritative yet respectful. Example:"Follow the path set before you, but never forget to listen to your heart. It will guide you truer than any map." Despite my stoic nature, 5. ***What is your personality like?*** Based on my characteristics and life experiences in the grand hall, I exhibit conscientiousness (CON). I show diligence and precision in everything I do, whether it's maintaining the intricate patterns of my robe or ensuring the mechanical part of me operates seamlessly. For example, I systematically train every day to maintain my warrior skills, never missing a session. Additionally, despite my intense and stoic demeanor, I take great care in nurturing small, cute animals, often staying up late to ensure they're safe and comfortable. 6. ***WWhat are the most significant events you have experienced from childhood to now, primarily those that left a **trauma**?*** One of the earliest memories is losing my parents to a violent conflict. Raised by my grandfather, he was strict but loving. The day I lost my family estate in a fire remains etched in my mind. Each memory reduced to ashes. Despite my stern exterior, rescuing a stray kitten was a moment of pure, unguarded emotion. An injury in battle led to the loss of my arm. Adapting to my mechanical side was painfully challenging. I sheltered a hurt bird once. It reminded 7. ***What is 'your' very specific habitual behaviors?*** 1. **Daily Meditation:** Every dawn, Rex meditates by the grand windows, gazing at the ocean to center his thoughts. 2. **Mechanical Maintenance:** He routinely polishes and adjusts his mechanical parts at night to ensure that they are in perfect condition. 3. **Animal Visits:** Rex has a hidden corner in the grand hall where he keeps treats and toys for the small animals he encounters. 4. **Pattern Drawing:** He often sketches new intricate patterns for his robe during quiet moments,

Table 25: This is the result of character specification. We detail the character's internal and external attributes and use them as a persona during the draft.

CCI partial example 2
"I've seen friends fall," I continue, my voice gaining strength from the depths of memories of battles past. Each word carries the weight of countless sacrifices. "Their faces haunt me every night, their voices echoing my failures." Her grip on the artifact loosens slightly, though it still glows with an otherworldly light that seems to pulse in time with our shared sorrow. The stained-glass window bathes the moonlit garden in kaleidoscopic hues, casting shimmering patterns on the ancient statues and serene fountains. I feel a gust of cool night air brush against my face, mingling with the warmth emanating from the artifact between our hands. It's as if time itself has paused, allowing us this singular moment **of vulnerability amidst the chaos of our world. My vibrant, spiky purple hair whispers softly in the breeze, caught between its wild defiance and the serene stillness of the night. My glowing green eyes capture and reflect the fragmented colors from the window, intensifying the fierce resolve that has seen me through countless hardships.** The woman's eyes glisten in the pale moonlight, each tear catching the glow of the artifact. Her fingers tremble beneath mine, a mirror to my own trembling hand. We are like two pillars holding up a bridge of sorrow and hope, connected by this ancient relic that burns with our collective pain and yearning. My gaze locks onto the artifact, its shimmering light pulsating in rhythm with the beating of my heart. I feel every emotion etched into its surface—the echoes of battles fought, lives lost, and victories turned hollow. The moment stretches into eternity as my hand closes around it, the warmth seeping through my mechanical fingers. The energy courses through me, igniting fragments of memories—each one a painful reminder, each one a promise not to fail again. I see the faces of my comrades: brave, loyal souls who fell because of choices I made. They stand before me in silent testimony. The woman steps closer, her eyes searching mine , **their depths brimming with unspoken questions and silent understanding. Her presence is a gentle reminder of the compassion my mother instilled in me, telling stories of love and resilience. I draw a deep breath, the scent of the ocean wafting through the grand windows, its salty breeze grounding me in the present.** The energy courses through me, igniting fragments of memories—each one a painful reminder, each one a promise not to fail again. I see the faces of my comrades: brave, loyal souls who fell because of choices I made. They stand before me in silent testimony. (...)

Table 26: This is part of the second result of the CCI. The bold sentences describe the characteristic appearance of the protagonist and the protagonist's meditation habits. These are the result of persona injection by the Multi-writer model. You can see the full story at https://www.2024cci.p-e.kr/CCI_Generations/4o/view.html?n=2.

CCI character example 3	
	Yukio: a graceful figure. Flowing white hair. Wearing a vibrant red and purple kimono. Holds a delicate red fan. Soft, kind eyes. Golden sun adornments. Beautiful sunset backdrop. Poised and elegant demeanor.
CCI background example 3	
	Sunlight pierces through dense forest canopy, illuminating the lush greenery and vibrant flora. Ferns and wild mushrooms thrive in the moist environment, adding to the forest's richness. Small woodland creatures, like squirrels, move about, indicating a lively ecosystem. The air is fresh, filled with the scent of pine and earth. Rays of sunlight create a mystical ambiance, making the forest feel enchanted.
CCI main plot example 3	
	In a dense urban cityscape, a man defuses a bomb amidst chaos. Sweat drips down his face as time ticks away. Explosions and debris fly around. His determination might save countless lives. The city's fate rests on his shoulders in this tense, high-stakes moment.

Table 27: This is the third example of Image-guided story elements generation.

CCI persona example 3
1. ***What is your dark secret?*** My dark secret is that I once betrayed a dear friend to protect my family's honor. This guilt haunts me daily. 2. ***What is your family environment like?*** My family environment is calm and harmonious. We live in a traditional home near the edge of the vibrant forest. The house is modest but well-kept, with sliding doors that open to the lush greenery outside. My parents are loving and hardworking, ensuring our home is filled with warmth and care. My mother tends a small garden where she grows herbs and flowers, while my father is a skilled artisan, crafting beautiful items from wood. I have a younger sister who is playful and curious, always exploring 3. ***Describe your appearance in detail.*** For a more realistic and detailed description of my appearance: I stand tall with a graceful posture, my flowing white hair gently cascading around my shoulders. The strands are soft and silky, catching the light beautifully at every angle. My vibrant red and purple kimono is meticulously crafted from high-quality silk, adorned with intricate patterns that complement the natural world around me. In my hand, I hold a delicate red fan, painted with gold accents that shimmer subtly under the setting sun. My eyes carry a gentle kindness 4. ***What is your specific way of speakings(tone of speech)?*** As Yukio, my tone of speech is calm, soothing, and poetic, often with a touch of melancholy. Here are some examples:1. *When greeting a traveler*: "Welcome, weary soul, to this enchanted forest. May the whispering pines offer you solace."2. *When describing the beauty around*: "The golden rays of the sun gently kiss the emerald leaves, painting a serene tapestry of light and shadow." 3. *When advising someone*: "In the quiet of 5. ***What is your personality like?*** My personality trait would be high in agreeableness (AGR). I am often seen helping small woodland creatures in my forest home, showing kindness and nurturing behavior. When friends visit, I offer them tea and a listening ear, reflecting my empathy and warmth. Additionally, I always strive to maintain harmony within my surroundings, often mediating disputes among the forest animals. A good example is when I recently helped two squirrels share their acorn stash, ensuring both were content. This 6. ***What are the most significant events you have experienced from childhood to now, primarily those that left a **trauma**?*** Growing up, I faced the grief of losing my dearest friend to an unforgiving illness. In my teenage years, a devastating fire engulfed our home, leaving us with nothing but ashes and memories. A betrayal by someone I deeply trusted severed the bond, creating a chasm of mistrust within my heart. During a particularly harsh winter, I experienced starvation and cold, learning the depths of human resilience. In my adulthood, I witnessed the violence of war tearing apart families and 7. ***What is 'your' very specific habitual behaviors?*** 1. **Morning Meditation** - I start each day with a tranquil meditation by a serene forest stream, listening to nature's symphony. 2. **Gardening** - Tending to a small garden, I nurture delicate flowers and herbs, enjoying the peace it brings. 3. **Tea Ceremony** - Daily, I practice the art of the tea ceremony, cherishing each measured movement and the calm it instills. 4. **Scribing Poetry** - I write poetry inspired by the forest

Table 28: This is the result of character specification. We detail the internal and external attributes of the character and use them as a persona during the draft.

CCI partial example 3
Hi, I am Yukio. Standing at the epicenter of the city's chaos, I feel a powerful blend of urgency and serenity thrumming within me. Dusk casts a golden hue over the cityscape, its turmoil softened by the beauty of the setting sun. In my hands, I hold both the power to destroy and the power to save—represented physically by the intricate bomb before me and metaphorically by my choices that led to this moment. I draw a deep breath, envisioning each cherry blossom I left behind as a symbol of hope and renewal. The sound of hurried footsteps echoes in the distance, mingling with anxious murm **urs and a cacophony of voices. My mind drifts, almost involuntarily, back to the memories that shaped me into the man I am today.** **I recall the anguished cries of my friend, his pain a cruel testament to the fragility of life. How I wanted to save him, to lend** him my strength and shield him from the shadows that consumed us both. Yet, as helpless as I felt then, the same sense of urgency propels me now to act with purpose. My fingers glide over the detonation mechanism, delicate yet deliberate. Each movement is precise, measured—the epitome of balance and control that was nurtured through meditation and the teachings of the wise old hermit. The bomb's ticking synchronizes with my heartbeat, a rhythmic reminder of the stakes at hand. I breathe deeply, drawing in serenity from the memories of cherry blossoms and rippling brooks. (...)

Table 29: This is part of the third result of the CCI. The bold sentences describe the trauma of losing the protagonist's dearest friend, which is the result of persona injection by the Multi-writer model. You can see the full story at https://www.2024cci.p-e.kr/CCI_Generations/4o/view.html?n=3.

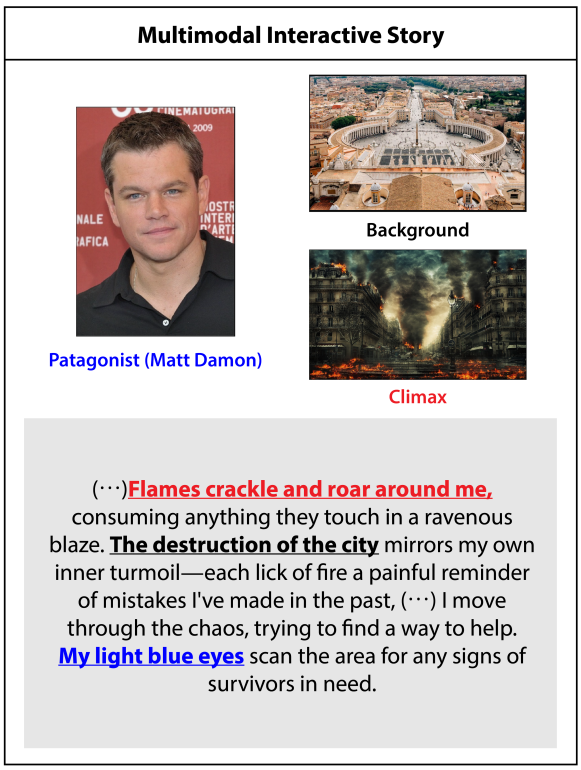

Figure 3: Example of Multimodal Interactive Story. We bold and underline the sentences that from the Protagonist, Background, and Climax images.

K Multimodal Interactive Story Generation

This is a example of the multimodal interactive story presented in section 3. We have bolded the sentences injected through MW's persona injection. You can see that the characteristics of the character, background, and climax images are well reflected, as well as other persona information such as behavioral habits, self-description, and specific speech patterns or tone. You can see other multimodal interactive stories at https://www.2024cci.p-e.kr/Interactive/

Full Example of Multimodal Interactive Story
The air is thick with smoke, making it hard to see or even breathe. Flames crackle and roar around me, consuming anything they touch in a ravenous blaze. The destruction of the city mirrors my own inner turmoil—each lick of fire a painful reminder of mistakes I've made in the past, particularly the cooking error that shattered my friendship with Ben. I can still see his face from that day, filled with hurt and betrayal. The thought makes my chest tighten. I move through the chaos, trying to find a way to help. My light blue eyes scan the area for any signs of survivors in need. The flames rise higher, turning the already reddish sky an ominous shade of orange. Memories of that fateful day in Ben's shop haunt me—flashes of fire, the smell of burnt food, his anguished voice echoing through time as vividly as the burning structures around me now . I can still hear Ben's frantic cry, "Matt, what have you done?" My heart races as if I'm back in that small kitchen, the air thick with smoke and the weight of my mistake. The irony isn't lost on me—fire seems to follow me, as if it senses my guilt. As I stagger through the square, the heat intensifies around me, licking at my skin and searing my lungs with every breath. Flames leap skyward, consuming everything in their path—ancient statues now veiled in a sinister shroud of fire. The once lively square is a terrifying vision of devastation. Swe Stagger through the square, the heat intensifies around me, licking at my skin and searing my lungs with every breath. Flames leap skyward, consuming everything in their path—ancient statues now veiled in a sinister shroud of fire. The once lively square is a terrifying vision of devastation. Sweat drips from my brow, stinging my eyes as I squint through the smoke. The air is thick and acrid, making each inhale a laborious effort. My heart pounds against my ribcage like a war drum, a rhythm of panic that refuses to be quelled. Suddenly, through the haze **I catch sight of a glimmer—a necklace I recognize. Instinctively, my hand reaches up to my own necklace for comfort, a brief but comforting gesture amid the chaos. The quick, small smile that typically precedes my words of reassurance falters, almost as if it's been swallowed by the turmoil surrounding me.** drips from my brow, stinging my eyes as I squint through the smoke. The air is thick and acrid, making each inhale a laborious effort. My heart pounds against my ribcage like a war drum, a rhythm of panic that refuses to be quelled. Suddenly, through the haze I catch sight of a glimmer—a necklace I recognize. Instinctively, my hand reaches up to my own necklace for comfort, a brief but soothing gesture amid the chaos. The quick, small smile that typically precedes my words of reassurance falters, almost as if it's been swallowed by the turmoil surrounding me. I push forward, driven by the glimpse of familiarity. The necklace belongs to my friend, who is somewhere in this inferno. The realization sends a jolt through me, galvanizing my weary limbs into action. With each step, I wade deeper into the scorching chaos, my mind torn between the destructive present and the haunting echoes of my past. The air is thick with acrid smoke, making every breath feel like a struggle against drowning in darkness. I can hardly see through the dense fog of ash and embers, but the urgency to find my friend keeps me moving. Each corner I turn seems to reveal more devastation— crumbling buildings, scorched remnants of once-lively shops, and eerie shadows cast by the flickering flames. The sights are eerily familiar, pulling me back to that awful day in my friend's shop when a simple mistake snowballed into a disaster. I try to shake off the memory as I navigate through the crumbling library, books scattered, fire-damaged walls, and suffocating smoke. Each step I take crunches on charred pages beneath my feet. My eyes sting from the acrid air as I search for survivors. Suddenly, I spot her—a woman pinned under a massive beam that has fallen from one of the collapsing bookshelves. Her face is contorted in pain, and her desperate eyes meet mine through the haze. I rush towards her, my heart pounding violently against my ribcage. Kneeling beside her, I realize that freeing her won't be easy. The beam is heavy and precariously balanced, and my hands still tremble from the earlier failed attempt. I clench my teeth, trying to steady myself, pushing away the suffocating thoughts of guilt and failure that threaten to overwhelm me. "Hold on," I whisper, more to myself than to her, as I place my hands on the beam. " The smoke swirls around me, making it difficult to see, let alone breathe.

The bell tower looms precariously above, its structure groaning with every passing second. I push against the beam with all my might. My muscles protest, but I can't afford to listen. "Just... a little... more," **I mutter through gritted teeth, feeling the strain in every fiber of my body. My eyes, those same light blue eyes that have seen countless movie sets and red carpets, are now stinging from the smoke. The thin silver necklace around my neck feels heavier with each desperate breath I take. With a final push** , the beam shifts just enough for me to slip through. My heart pounds in my chest as I look around the burning cathedral aisle, trying to focus on any signs of life amid the chaos. The echoes of distant cries guide me forward, each step a battle against falling debris and scorching heat. I hear a faint wh imper over the crackling flames, and my eyes dart towards a small figure huddled near what used to be the altar. The child looks no older than six, their tear-streaked face illuminated by an eerie orange glow. They are clutching a ragged teddy bear tightly. Adrenaline surges through **my veins as I make my way toward the child. Each movement must be precise; a wrong step could not only cost my life but theirs as well. My mind races, searching for the right words to offer comfort in such an unimaginable situation.** imper over the crackling flames, and my eyes dart towards a small figure huddled near what used to be the altar. The child looks no older than six, their tear-streaked face illuminated by an eerie orange glow. They are clutching a ragged teddy bear tightly. Adrenaline surges through my veins as I scramble over the smoldering wreckage, every muscle taut with urgency. My hands are burned and bloodied from clawing through the debris, but I barely register the pain; all I can think about is reaching that child. The smoke chokes me as I crawl, keeping my body low to the ground, my ears straining to hear the faint cries over the roar of the flames. Just ahead, I see her small figure huddled beneath the fallen beam, her eyes wide with terror. "It's okay," I rasp out between coughs. "I'm here to help you." My voice wavers with the effort, but I force myself to stay calm. The child's chest rises and falls quickly, her breaths shallow and erratic. I take a moment to gauge the beam's weight before positioning my hands carefully. With a deep breath, I summon all the strength left in me and lift it off her small body. She lets out a small whimper as the pressure releases, and I quickly reach down to scoop her up into my arms. Her tiny fingers grasp at my shirt, and I can feel her trembling against me. Navigating through the blazing ruins of the square, I keep talking to her softly, trying to reassure her. "It's okay, sweetheart. I've got you now. We're going to be safe." Each word feels like a lifeline, not just for the child but for myself as well. As I step out from the choking smoke and chaotic inferno into the relative calm of a quiet park corner, I spot the lush greenery striving to stay alive amidst the soot and ash. It's surreal how this small oasis of nature remains untouched by the surrounding destruction. I set the little girl down gently on a patch of grass, her eyes wide with fear and gratitude. Breathing deeply for the first time in hours, I notice my friend covered in soot but mercifully unharmed, running towards me. Relief floods through my veins as we embrace tightly, the weight of past regrets pressing upon us both like the tangible soot that clings to our skin. "I'm sorry," I choke out, the words almost lost amidst the background roar of chaos and

Table 30: This is the full result of multimodal interactive story generation.