

LOGIC: Long-form Outline Generation via Imitative and Critical Self-refinement

Hengwei Liu Yongliang Shen Zhe Zheng Haoyuan Ma

Xingyu Wu Yin Zhang[†] Weiming Lu[†]

Zhejiang University

{hengweiliu, syl, zhezhen, mahaoyuan, wuxingyu, yinzh, luwm}@zju.edu.cn

Abstract

Long-form outline generation for expository articles requires both comprehensive knowledge coverage and logical coherence, which is essential for creating detailed Wikipedia-like content. However, existing methods face critical limitations: outlines generated in the pre-writing stage often have low knowledge density and lack detail, while retrieval-augmented approaches struggle to maintain logical coherence across retrieved information. Additionally, unlike human writers who can iteratively improve through peer feedback and reference similar topics, current approaches lack effective mechanisms for systematic outline refinement. To address these challenges, we propose **LOGIC**, a Long-form Outline Generation system via Imitative and Critical self-refinement that mimics human writers' refinement process. **LOGIC** establishes a coherent planning framework and structured knowledge base, learns from similar topic outlines through imitation, and continuously improves through model-based critique. Experiments on FreshWiki and our dataset WikiOutline show that, compared to the best baseline, **LOGIC**'s long-form outlines are more organized (with increases of 22.85% and 21.65% respectively) and more logically coherent (with increases of 16.19% and 12.24% respectively). Human evaluation further validates **LOGIC**'s effectiveness in generating comprehensive and well-structured long-form outlines.

1 Introduction

Large language models (LLMs) have demonstrated remarkable capabilities in novel generation through planning-then-writing approaches (Yang et al., 2022; Huang et al., 2024; Yao et al., 2019). Such approaches first create a coherent plot outline, then decompose the generation into subtasks, enabling LLMs to produce richer and more engaging content

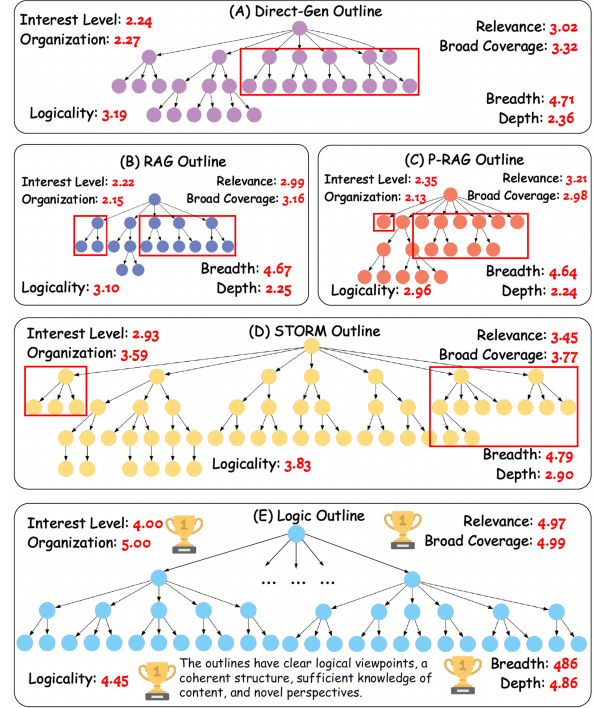


Figure 1: Comparison of the quality of outlines generated by different methods. The red boxes represent nodes that are too detailed and lack coherence. The outlines generated by **LOGIC** achieve better results in all metrics, and the level of detail of the outline of a section is equivalent to the outline of an article by other methods.

(Riedl and Young, 2010; Bai et al., 2024; Ma et al., 2024). When generating expository articles similar to Wikipedia entries (Fan and Gardent, 2022; Banerjee and Mitra, 2015) or reports (Liu et al., 2018; Ranade and Joshi, 2023), planning in the pre-writing stage becomes even more crucial (Rohman, 1965). Compared with the methods that bypass the pre-writing stage, the role-playing-based method STORM (Shao et al., 2024) has enhanced the quality of the generated articles by generating outlines in the pre-writing stage. However, maintaining the coherence and knowledge richness of the outlines

[†]Corresponding author.

still remains a significant challenge.

This challenge stems from three critical limitations: 1) *Knowledge Density Gap*: Outlines generated in the pre-writing stage often lack sufficient knowledge density and detail. 2) *Retrieval-Coherence Dilemma*: While retrieval-augmented generation (RAG) (Lewis et al., 2020) can introduce external knowledge to reduce the hallucination of LLMs (Xu et al., 2023; Rawte et al., 2023), it struggles to maintain logical coherence across retrieved information. 3) *Limited Refinement Capability*: Unlike human writers who can iteratively improve their outlines through peer feedback (Horbach and Halffman, 2018) and reference similar topics, current LLM approaches lack effective mechanisms for systematic outline refinement (Chamoun et al., 2024; Bae and Kim, 2024) and correcting factual contradictions (Lyu et al., 2024).

Current approaches attempt to address these challenges through different strategies. Direct generation (Figure 1 (A)) relies solely on LLMs’ internal knowledge but produces shallow content. Basic RAG approaches (Figure 1 (B, C)) enhance knowledge coverage but fail to improve coherence and topic coverage significantly. Even advanced methods like STORM (Shao et al., 2024), which employs multi-perspective information collection (Figure 1 (D)), and OmniThink (Xi et al., 2025), which expands knowledge boundaries through continuous reflection and exploration, still struggle with maintaining appropriate detail levels and logical consistency across the outline.

Inspired by human writers’ practices of studying similar topics and iterative refinement, we propose **LOGIC**, a novel long-form outline writing system. **LOGIC** operates through three innovative mechanisms: 1) establishing structural coherence through a planning list and knowledge framework, 2) learning from similar topic outlines through imitation, 3) implementing iterative refinement based on critical model feedback. This approach enables **LOGIC** to generate outlines with enhanced content richness, entity density, logical coherence, and engagement.

We evaluate **LOGIC** on both the FreshWiki dataset (Shao et al., 2024) and our newly proposed WikiOutline dataset (§4.1). Our experiments show that the long-form outlines have significantly improved in multiple key aspects, including entity recall, interest level, coherence and organization, relevance and focus, broad coverage, logicity, breadth, and depth. Human evaluation further validates our method’s effectiveness in addressing the

long-form outline generation challenge.

Our main contributions include:

- We propose a novel framework that integrates outline planning, knowledge structuring, and external knowledge in generating high-quality expository long-form outlines.
- We develop **LOGIC**, an innovative system that combines imitative learning with writer-like iterative refinement for outline generation.
- We demonstrate through automatic and human evaluation that **LOGIC** can effectively generate logically coherent and content-rich outlines, significantly advancing the state of the art in long-form outline generation.

2 Related Work

2.1 Retrieval-Augmented Generation (RAG)

Retrieval-augmented generation enables LLMs to generate factual responses based on knowledge introduced from external databases (Izacard and Grave, 2020; Qian et al., 2024a), improving the reasoning capabilities of the model (Izacard et al., 2023). RAG can not only achieve simple question-answering tasks (Gao et al., 2022) and improve real-time performance on intensive NLP tasks (Lewis et al., 2020), but also prevent Wikipedia-based chatbots from hallucinations (Semnani et al., 2023). GraphRAG (Edge et al., 2024) and StructRAG (Li et al., 2024) can extract entity relationships from unstructured documents and generate structured summaries, significantly enhancing the information integration ability in implicit reasoning tasks. Qian et al. (Qian et al., 2024b) proposed the long-term memory-enabled MemoRAG to solve the problem of difficulty in effectively retrieving unstructured external knowledge (Edge et al., 2024). Long-form article generation methods such as STORM (Shao et al., 2024) and Co-STORM (Jiang et al., 2024) proposed retrieval strategies from different perspectives, but there is no external knowledge of the model for precise and structured topic-based retrieval.

2.2 Automatic Outline Writing

With the rise of LLMs, the field of automatic machine writing has ushered in an unprecedented wave of development (Zhou et al., 2023; Wang et al., 2024c,b). The writing of automated expository articles (Balepur et al., 2023; Liang et al.,

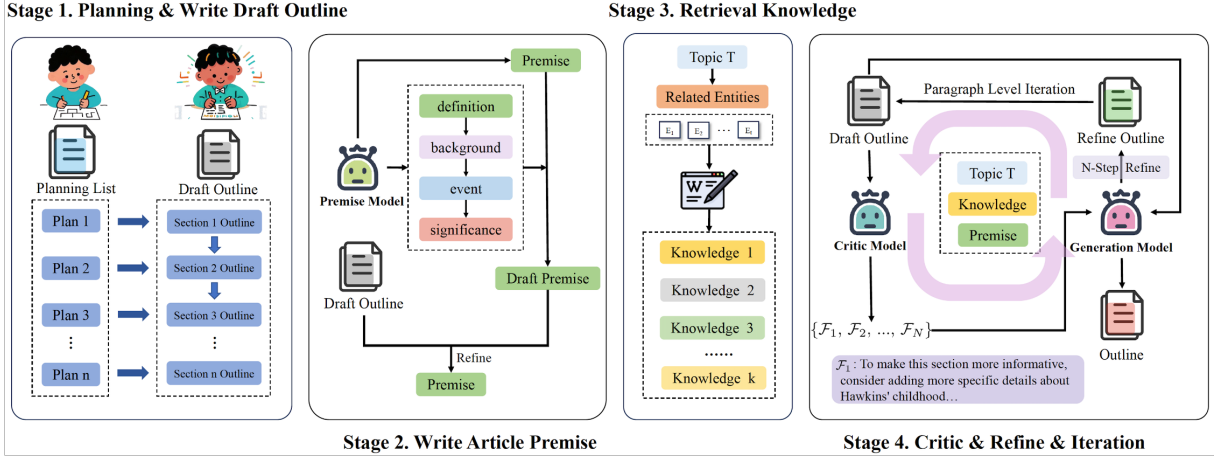


Figure 2: The overview of **LOGIC**. In stage 1 (§3.2), a planning list is generated according to the topic, and then an outline for each section is generated based on the planning list. In stage 2 (§3.3), an article premise is generated from four aspects, namely the definition, background, events, and significance related to the topic. In stage 3 (§3.4), the outline of the entities related to the topic is retrieved. In stage 4 (§3.5), the generation model continuously iteratively refines the draft outline according to the feedback from the critic model.

2024) is different from the generation of long-form articles such as novels, as it requires relying on external documents to ensure the factuality of the articles. Similarly, the generation of an expository article outline (Wang et al., 2024a), a question-and-answer article outline (Chen et al., 2024), or a blog outline (Lee et al., 2024), which can no longer depend on character relationships and plot to maintain continuity, differs from that of a novel outline. STORM (Shao et al., 2024) proposed a method for writing Wikipedia-like article outlines based on a role-playing approach. This method innovatively introduces the concept of role-playing and attempts to generate a reasonable outline by simulating the question-answer interaction between different roles. However, these methods often lack a process of continuous refinement based on the retrieved knowledge, resulting in the generated outlines being of insufficient quality. Different from previous works, we propose a long-form outline generation framework named **LOGIC** that is based on continuous refinement through feedback, enabling the iterative refinement of outlines to make them more content-rich, logical, and coherent according to the feedback from the critical model.

3 LOGIC

3.1 Overview

We propose **LOGIC** to automate long-form outline writing, the writing process starts with generating a comprehensive planning list that incorporates multiple detailed plans. Subsequently, a draft outline

is crafted based on these plans, and notably, each plan corresponds precisely to a distinct section outline. Next, a premise model is used, considering the definition, background, event, and significance related to the topic to generate the draft premise and refine the draft premise with the draft outline that serves as the fundamental basis for the article. Then, related entities will be identified, and relevant knowledge will be retrieved to enrich the content. Finally, a critic model examines the draft outline and provides feedback, while a generation model uses this feedback, along with the knowledge and premise, to refine the outline through an iterative process until it reaches a satisfactory state. The pseudocode of **LOGIC** can be found at Algorithm 1 and Appendix D. The input of **LOGIC** stage 4 can be found in Figure 5 and a study case of long-form outlines generated by **LOGIC** in Appendix F.

3.2 Generate Draft Outline

In order to achieve the output of long articles through LLMs, AgentWrite (Bai et al., 2024) proposed a divide-and-conquer framework that decomposes the writing task into multiple subtasks and finally connects the subtask outputs to obtain the final long-form output. In terms of long-form outline construction, we also used a divide-and-conquer strategy to split the complex long-form outline generation process into paragraph-level sub-steps based on the planning list.

Given an input topic $\mathcal{T}$, **LOGIC** generates a planning list $\mathcal{PL} = \{\mathcal{PL}_1, \mathcal{PL}_2 \dots \mathcal{PL}_n\}$ containing

plans for each section and uses these plans to generate a draft outline for the corresponding section. By generating the planning list in one fell swoop, it becomes feasible to comprehensively govern the entire document’s overall structure and logical orientation. This ensures the interconnection and coherence among all components, rendering each section’s outline closely intertwined within the overarching framework. Consequently, issues such as logical disruptions or deviations in the discussion direction can be effectively circumvented. When generating the draft outline of the n th section $\mathcal{O}_{Dn}$, the draft outlines of the previous $n-1$ sections $\{\mathcal{O}_{D1}, \dots, \mathcal{O}_{D(n-1)}\}$ and the plan of the n -th section $\mathcal{PL}_n$ are used as input. This serial generation method is more coherent and continuous than the outline generated in parallel.

Algorithm 1 LOGIC

Input: Topic $\mathcal{T}$

Output: Outline $\mathcal{O}$

```

1: PlanList  $\mathcal{PL} = \text{LLM}(\mathcal{T})$  with  $n$  sections;
2: for  $i \leftarrow 1$  to  $n$  do
3:    $\mathcal{O}_{Di} = \text{LLM}(\mathcal{T}, \mathcal{PL}_i, \{\mathcal{O}_{D1}, \dots, \mathcal{O}_{D(i-1)}\})$ 
4: end for
5: Draft Outline  $\mathcal{O}_D = \{\mathcal{O}_{D1}, \mathcal{O}_{D2} \dots \mathcal{O}_{Dn}\}$ 
6:  $\mathcal{P} = \text{PremiseGen}(\mathcal{T})$ 
7:  $\mathcal{K} = \text{KnowledgeRetrieve}(\mathcal{T})$ 
8: for  $i \leftarrow 1$  to  $n$  do
9:    $\mathcal{O}_i^0 = \text{LLM}(\mathcal{T}, \mathcal{P}, \mathcal{K}, \mathcal{O}_D, \{\mathcal{O}_1, \dots, \mathcal{O}_{i-1}\}, \mathcal{O}_i)$ 
10:  Feedback  $\mathcal{F} = \{\mathcal{F}_1, \mathcal{F}_2 \dots \mathcal{F}_m\} = \text{LLM}(\mathcal{T}, \mathcal{K}, \mathcal{O}_D, \{\mathcal{O}_1 \dots \mathcal{O}_{i-1}\}, \mathcal{O}_i^0)$  with  $m$  sections;
11:  for  $j \leftarrow 1$  to  $m$  do
12:     $\mathcal{O}_i^j = \text{LLM}(\mathcal{T}, \mathcal{K}, \mathcal{O}_i^{j-1}, \mathcal{F}_j)$ 
13:  end for
14:   $\mathcal{O}_i = \mathcal{O}_i^m$ 
15: end for
16: return  $\mathcal{O}$ 

```

3.3 Generate Modular Article Premise

The pre-trained language model (OpenAI, 2024; Yang et al., 2024; Dubey et al., 2024) itself contains rich parameter knowledge. When processing complex knowledge relationships, the model may fail to consider the association information between multiple dimensions, resulting in incomplete knowledge point structure or inaccurate relationships in the output. Constructing a long-form outline requires a high-quality premise to sort out knowledge points and their relationships accurately. We propose a

modular approach for article premise generation.

Begin by generating a draft premise $\mathcal{P}_D$ based on the topic $\mathcal{T}$. Then, four modules are generated progressively based on the topic: definition $\mathcal{D}$, background $\mathcal{B}$, events $\mathcal{E}$, significance, and impact $\mathcal{S}$. This step-by-step generation process allows each module to build upon the previous module, making the exploration of the topic more in-depth and comprehensive, and contributing to the overall coherence of the article premise. Next, modify the draft premise $\mathcal{P}_D$ based on these four modules so that the draft premise contains the content of these four models. Finally, modify and refine the draft premise $\mathcal{P}_D$ according to the draft outline to obtain the final article premise $\mathcal{P}$. The pseudocode of **PremiseGen** can be found at Algorithm 2 in Appendix D.

3.4 Retrieve Outlines of Related Entities

Unlike STORM (Shao et al., 2024) and Co-STORM (Jiang et al., 2024), which retrieve content related to the topic, **LOGIC** retrieves the outlines corresponding to the entities related to the topic. The following is the full implementation of the KnowledgeRetrieve in Algorithm 1. First, prompt LLMs to generate a list of entities $\mathcal{E} = \{\mathcal{E}_1, \mathcal{E}_2 \dots \mathcal{E}_t\}$ related to the topic. Then, outlines are extracted from Wikipedia API[†] based on the obtained entity lists, which can be used as knowledge input for LLMs. With the help of these outlines, LLMs can analyze their structures and contents, such as hierarchical and logical relationships. That is, the LLMs can imitate the outline of the relevant entity for writing, thus generating well-organized and information-rich outlines, effectively improving the quality and logic of the generated content.

From the perspective of the structural paradigm, STORM relies on unstructured historical dialogue context to retrieve topic-related content and thus lacks explicit organizational frameworks, while **LOGIC** retrieves structured entity-based outlines that are pre-formatted knowledge frameworks, such as hierarchical lists of key entities, attributes, and relationships; this structural design allows the generation model to directly leverage organized information, mimicking the way human writers create outlines. In terms of semantic focus, STORM retrieves contextual themes derived from dialogue history, whereas the "related entities" in **LOGIC** are defined as outlines of relevant entities; **LOGIC** an-

[†]<https://pypi.org/project/Wikipedia-API/>

chors the generation process in these entity-based outlines, enabling LLMs to mimic the structure of other entity outlines during writing, and this approach enhances the coherence and factual accuracy of the output by providing a structured semantic foundation, ensuring that the generated content remains tightly aligned with defined entities and their hierarchical relationships. The rationale for emphasizing entity-based retrieval lies in two aspects: on one hand, structured outlines provide a clear scaffold for the generation model, reducing the ambiguity of unstructured text and enabling more controlled content creation to improve model usability, and on the other hand, entities serve as unambiguous semantic anchors, allowing **LOGIC** to retrieve domain-specific knowledge with higher precision compared to **STORM**, which may suffer from contextual vagueness.

3.5 Refine Draft Outline based on Feedback from the Critical Model

The preliminary draft outline commonly shows relative coarseness and lacks the profundity and comprehensiveness crucial for creating high-quality long-form content. Refining the outline ensures that each section is elaborately structured and contains necessary elements to support the overarching topic $\mathcal{T}$. Distinct entities have their unique characteristics and specific details. When refining the outline, imitating the outlines of other entities allows us to draw on diverse perspectives and outstanding writing techniques, thus enhancing the richness and distinctiveness of our long-form outline.

We use the topic $\mathcal{T}$, the article premise $\mathcal{P}$, and the outline of relevant entities as the input for the generation model and the critical model. First, the generation model takes the outlines of relevant entities as the context, and then imitates these outlines to refine the initial draft outline, making the initial draft outline more complete and specific. Next, the critical model generates a series of feedbacks $\mathcal{F} = \{\mathcal{F}_1, \mathcal{F}_2, \dots, \mathcal{F}_m\}$, which are suggestions for refining the outline. We figured out the right number of times to refine the results (that’s the parameter m). For GPT-4o-mini, 5 refinement iterations work best. On the other hand, llama-3.1-8B and Qwen2-72B do their best when we use 3 iterations. The generation model will iteratively refine the outline based on this series of feedback, and during the iterative process, the outline of relevant entities will be fed into the LLMs as knowledge. The i in $\mathcal{O}_i^j$ in Algorithm 1 represents the outline of the i -th sec-

tion, and j represents the j -th self-refinement. This entire process is at the section level, and each input will include the outline of the previous section to ensure the coherence of generating the new section outline. Finally, we integrate the refined outline of each section into the final long-form outline $\mathcal{O}$.

3.6 Writing the Long-Form Article

With reference to the article premise $\mathcal{P}$ and long-form outline $\mathcal{O}$, the full long-form article can be written in parallel with sections. Since the content provided by the section outline and the article premise are relatively limited, we’ll search for relevant content according to the section outline to generate the paragraphs of the corresponding sections. Furthermore, we prompt the LLMs to expand each section immediately after its generation to obtain more substantial content. Finally, we connect all the expanded sections into a full long-form article.

4 Experiments

4.1 Datasets

Numerous studies (e.g., **STORM**, **OmniThink**) have widely used **FreshWiki** to evaluate outline generation tasks. In order to test the generalization ability of our method on Wikipedia articles from different fields, we create **WikiOutline** by referring to the **FreshWiki** construction method. The dataset contains Characters, Events, Films, Disasters, and Places five categories. The details can be found in Appendix A. We use **FreshWiki** and **WikiOutline** to verify the effectiveness of our method.

4.2 Automatic Metrics

Besides the heading soft recall and heading entity recall (**Fränti and Mariescu-Istodor, 2023**) for evaluating the outline quality, we also introduce the evaluation of the outline by modifying four metrics that **STORM** (**Shao et al., 2024**) uses to score the articles: (1) Interest Level, (2) Coherence and Organization, (3) Relevance and Focus, (4) Broad Coverage. Besides, we have introduced three new metrics: (5) Logicality, (6) Breadth, and (7) Depth to evaluate the quality of the outline. For aspects (1)-(7), we use **Prometheus** (**Kim et al., 2023**) to score the outline, and the specific scoring criteria can be found in Appendix C.

4.3 Baseline and Implementation

We select five representative baselines for comparison, including *Direct Gen*, *RAG*, *P-RAG*, *STORM*

		Heading Soft Recall	Heading Entity Recall	Rubric Grading						
				Interest	Organization	Relevance	Coverage	Logicity	Breadth	Depth
GPT-4o -mini	Direct Gen	87.24	33.67	2.24	2.27	3.02	3.32	3.19	4.71	2.36
	RAG	87.14	37.65	2.22	2.15	2.99	3.16	3.10	4.67	2.25
	P-RAG	83.42	37.30	2.35	2.13	2.98	3.21	2.96	4.64	2.24
	STORM	90.56	<u>35.51</u>	2.93	3.59	3.45	<u>3.77</u>	<u>3.83</u>	<u>4.79</u>	2.90
	OmniThink	<u>91.05</u>	33.36	<u>3.58</u>	<u>4.07</u>	<u>3.48</u>	3.54	3.78	4.49	<u>3.15</u>
	LOGIC	96.04	41.66	4.00	5.00	4.97	4.99	4.45	4.86	4.86
		+5.48%	11.73%	+36.52%	+22.85%	+42.82%	+32.36%	+16.19%	+1.46%	+54.29%
Qwen2 -72B	Direct Gen	<u>91.24</u>	36.18	1.87	1.74	2.07	1.88	1.50	2.60	1.96
	RAG	85.64	36.27	2.21	1.91	2.68	2.75	2.08	3.80	1.90
	P-RAG	83.49	35.20	2.26	1.99	2.80	2.97	2.29	3.96	1.99
	STORM	76.31	<u>39.47</u>	2.79	3.12	<u>3.14</u>	<u>3.14</u>	2.78	<u>3.96</u>	2.72
	OmniThink	62.56	32.71	<u>2.92</u>	<u>3.53</u>	2.69	3.07	<u>2.89</u>	3.93	<u>3.45</u>
	LOGIC	92.00	40.56	3.90	4.90	4.48	4.67	4.41	4.88	4.63
		+0.83%	+2.76%	+33.56%	+38.81%	+42.68%	+48.73%	+52.60%	+23.23%	+34.20%
Llama3.1 -8B	Direct Gen	85.46	32.72	1.44	1.33	1.52	1.34	1.29	2.02	1.59
	RAG	68.00	31.31	1.88	1.73	2.36	1.75	1.78	2.41	1.57
	P-RAG	65.58	30.63	1.73	1.61	2.16	1.72	1.64	2.39	1.54
	STORM	83.73	<u>35.83</u>	2.12	2.01	<u>2.64</u>	<u>2.23</u>	1.88	3.41	1.95
	OmniThink	93.64	32.98	<u>3.00</u>	<u>3.26</u>	2.44	1.98	<u>2.72</u>	<u>3.59</u>	<u>3.31</u>
	LOGIC	<u>86.92</u>	36.79	3.79	4.71	3.94	4.32	4.13	4.57	4.27
		-	+2.68%	+26.33%	+44.48%	+49.24%	+93.72%	+51.84%	+27.30%	+29.00%

Table 1: Results of automatic outline quality evaluation of FreshWiki dataset. The last row indicates the percentage improvement over the best baseline. The rubric grading uses a 1-5 scale.

		Heading Soft Recall	Heading Entity Recall	Rubric Grading						
				Interest	Organization	Relevance	Coverage	Logicity	Breadth	Depth
Direct Gen	95.75	68.20	2.18	2.57	3.15	3.35	3.51	4.71	2.46	
RAG	94.79	65.50	2.24	2.12	3.04	3.35	3.26	4.67	2.27	
P-RAG	92.49	66.83	2.09	2.12	2.98	3.53	3.46	4.69	2.25	
STORM	95.91	<u>68.89</u>	3.05	<u>4.11</u>	<u>3.68</u>	<u>3.72</u>	<u>3.92</u>	<u>4.62</u>	<u>3.32</u>	
OmniThink	<u>96.45</u>	67.83	<u>3.43</u>	3.81	3.58	3.38	3.84	4.33	3.07	
LOGIC	98.52	72.24	3.99	5.00	5.00	4.97	4.40	4.86	4.76	
	+2.15%	+5.44%	+16.32%	+21.65%	+35.87%	+33.60%	+12.24%	+5.19%	+43.37%	

Table 2: Results of automatic outline quality evaluation of WikiOutline dataset on GPT-4o-mini. The last row indicates the percentage improvement over the best baseline. The rubric grading uses a 1-5 scale.

(Shao et al., 2024), *OmniThink* (Xi et al., 2025). *Direct Gen* generates outlines by directly harnessing the inherent knowledge within the LLMs. *RAG* represents another baseline approach, generating outlines by relying on knowledge retrieved from topic-related sources. *P-RAG* generates outlines by incorporating knowledge retrieved from both topic-related and paragraph-list-related sources. *STORM* serves as a role-playing baseline specifically tailored for outline generation in multi-perspective dialogues. *OmniThink* is a baseline for generating outlines by expanding knowledge boundaries through continuous reflection and exploration. The implementations of *STORM* and *OmniThink* are consistent with their original papers, detailed im-

plementation can be found in the Appendix B.

5 Results and Analysis

5.1 Main Results

Table 1 presents the evaluation results on the FreshWiki dataset employing GPT-4o-mini, Qwen2-72B-Instruct-AWQ, and Llama-3.1-8B-Instruct as backbones. **LOGIC** attains the top score of 5 for the "Organization" rubric grading criteria by ensuring the organization and coherence of the long-form outline generated via the serial generation mode in the initial stage, along with the planning list. By simulating the creative instincts and mimetic writing capabilities inherent in human writing, **LOGIC**

		Heading Soft Recall	Heading Entity Recall
GPT-4o -mini	LOGIC	96.04	41.66
	w/o Feedback	94.41	38.50
	w/o F&P	93.10	35.55
	w/o F&P&K	89.60	33.64
Qwen2 -72B	LOGIC	92.00	40.56
	w/o Feedback	90.66	40.41
	w/o F&P	88.92	38.59
	w/o F&P&K	87.95	34.67
Llama3.1 -8B	LOGIC	86.92	36.79
	w/o Feedback	83.91	33.98
	w/o F&P	84.27	32.86
	w/o F&P&K	80.85	31.76

Table 3: The comparison between **LOGIC**, **LOGIC** w/o Feedback, **LOGIC** w/o F&P, and **LOGIC** w/o F&P&K, where **LOGIC** w/o F&P refers to **LOGIC** without feedback and premise, **LOGIC** w/o F&P&K refers to **LOGIC** without feedback, premise, and knowledge.

significantly enhances rubric grading criteria like depth and logicity while ensuring the breadth of the generated long-form outlines. In the seven main rubric grading criteria (Interest, Organization, Relevance, Coverage, Logicity, Breadth, and Depth), **LOGIC** with GPT-4o-mini as the backbone performed well in all aspects, especially in the three metrics of Organization, Relevance, and Coverage. Although the rubric grading criteria have dropped when Qwen2-72B-Instruct-AWQ and Llama3.1-8B-Instruct are used as the backbone, **LOGIC** still improves by more than 20% in all rubric grading criteria compared to the best baseline. When Llama3.1-8B-Instruct is used as the backbone, the two metrics of Coverage and Logicity increase by more than 50%.

Table 1 shows the heading soft recall and heading entity recall of the long-form outlines generated by GPT-4o-mini, Qwen2-72B-Instruct-AWQ, and Llama-3.1-8B-Instruct as the backbone. When using Qwen2-72B-Instruct-AWQ as the backbone, the heading soft recall of the long-form outlines generated by STORM is lower than that of other baselines. This is may because the long-form outlines generated by STORM do not conform to the expected format. When using Llama3.1-8B-Instruct as the backbone, OmniThink’s generated long outlines have higher heading soft recall due to many long sentences compared to other baselines and **LOGIC**. However, since outlines should not contain sentences, we believe that the outlines

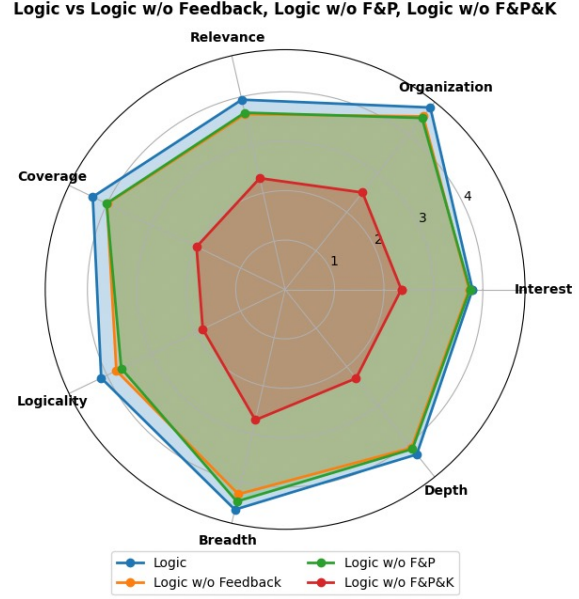


Figure 3: The comparison between **LOGIC**, **LOGIC** w/o Feedback, **LOGIC** w/o F&P, and **LOGIC** w/o F&P&K on Llama-3.1-8B-Instruct.

generated by OmniThink when using Llama3.1-8B-Instruct as the backbone are unreasonable. For long-form outlines generated by baselines such as RAG and P-RAG, the heading soft recall and entity recall are almost even lower than, those of outlines generated directly by LLMs. This is due to LLMs lacking the ability to extract topic-related structured information from retrieved content. **LOGIC** achieves high heading soft recall and heading entity recall as it learns to process complex knowledge relationships by generating modular article premise and achieves scores higher in the seven rubric grading criteria as it simulates the reflective process of writers. From Table 2, it can be clearly observed that in the long-form outline generation tasks across WikiOutline datasets, **LOGIC** demonstrates a more outstanding performance than STORM and other baseline models in heading soft recall, heading entity recall, and rubric grading criteria. Detailed experimental results on the five domains of characters, events, movies, disasters, and places in the WikiOutline dataset can be found in Appendix A.

5.2 Ablation Study

As described in §3, **LOGIC** guarantees the logicity of the generated long-form outlines by constructing article premises and, by retrieving the outlines of relevant entities as external knowledge, enables LLMs to mimic the writing of human writers. We conduct an ablation study on long-form outline cre-

		Prompts	Interest	Organization	Relevance	Coverage	Logicity	Breadth	Depth
GPT-4o-mini	Direct Gen	1	2.24	2.27	3.02	3.32	3.19	4.71	2.36
Llama3.1-8B	STORM	8	2.12	2.01	2.64	2.23	1.88	3.41	1.95
Llama3.1-8B	LOGIC	14	3.79	4.71	3.94	4.32	4.13	4.57	4.27

Table 4: Analyze the performance improvement of small models by increasing the number of prompts.

		Section counts	Outline lengths
GPT-4o	STORM	8.19	45.27
-mini	LOGIC	13.95	256.51
Qwen2	STORM	8.41	52.16
-72B	LOGIC	8.35	261.35
Llama3.1	STORM	11.09	36.16
-8B	LOGIC	9.83	202.22

Table 5: Comparison of section counts and outline lengths between STORM and **LOGIC**.

ation by comparing **LOGIC** with three variants on the FreshWiki dataset: (1) "**LOGIC** w/o Feedback", which omits feedback in the stage 4, (2) "**LOGIC** w/o F&P", which omits feedback in the stage 4 and removes the article premise as input, (3) "**LOGIC** w/o F&P&K", which omits feedback during the stage 4 while removing article premises and external knowledge as input. Table 3 shows the ablation results, indicating that the long-form outlines generated by the full **LOGIC** process have the highest recall, and Figure 3 also suggests that the full **LOGIC** process scores higher on all rubric grading criteria. Although the effects of feedback and premise seem negligible as shown in Figure 3, the two metrics, Heading Soft Recall and Heading Entity Recall, in Table 3 demonstrate the crucial roles of feedback and premise for **LOGIC**. Figure 3 shows that "**LOGIC** w/o F&P&K" performs much worse in all rubric grading criteria, demonstrating the importance of the outlines of relevant entities as external knowledge to **LOGIC**. It also shows that the main feature of **LOGIC** is its ability to effectively imitate human writing and self-refine critically, generating well-structured, detailed long-form outlines that align with human thinking habits.

5.3 Detailed Analysis

Outline Quality Assessment. Table 5 summarizes the count of sections and the length of outlines for **LOGIC** and STORM. Statistical analysis reveals no significant disparity in the number of sections between the two methods. Specifically, when GPT-4o-mini serves as the foundational model, **LOGIC**

generates outlines with a greater number of sections compared to STORM. Conversely, when Llama3.1-8B is the backbone model, STORM exhibits a higher section count. Notably, across all evaluated models, the outline lengths produced by **LOGIC** are significantly longer than those of STORM. This finding strongly indicates that **LOGIC** possesses a superior ability to generate long-form outlines that are not only more extensive but also contain finer details.

While **LOGIC** does generate longer outlines, keyword stuffing almost does not exist, and the content maintains semantic coherence and logical flow. Regarding the recall criteria, when calculating entity recall, we removed duplicate entities to ensure the criteria reflects genuine content diversity rather than mere repetition. This normalization step helps mitigate the effect of length on recall scores. Additionally, the effectiveness of **LOGIC** is supported by seven other rubric criteria, which collectively demonstrate that **LOGIC**'s improvements are not solely attributable to increased length but rather to the quality and comprehensiveness of the generated content.

Pipeline Granularity Analysis. When considering different models, the impact of pipeline granularity varies. Table 4 shows that for larger models like GPT-4o-mini, which can handle tasks relatively well in a single-prompt (Direct Gen) scenario, increasing pipeline granularity might not always yield proportional benefits. However, for smaller models such as Llama3.1-8B, a more fine-grained pipeline can potentially bridge the performance gap with larger models. This indicates that for smaller models, a more granular pipeline can enhance the quality of the generated output.

6 Human Evaluation

To better understand the strengths and weaknesses of **LOGIC**, we collaborated with five volunteers interested in long-form outline generation for a human evaluation. We randomly sample 10 topics from the FreshWiki dataset and evaluate the long-form outlines generated by our approach and

STORM. To comprehensively evaluate the quality of each long-form outline, we consider four key metrics: Logicality, Coherence, Knowledgeability, and Novelty, ensuring the outlines have clear logical viewpoints, a coherent structure, sufficient knowledge of content, and novel perspectives.

We calculated the intraclass correlation coefficient, and the value of ICC3k was 0.73, indicating a high degree of consistency among different participants. The p-value was 0.04, suggesting that the differences among different raters were statistically significant. For the improvement of our method compared to STORM in terms of the "Logicality" criteria, the Pearson correlation coefficient between the results of the human evaluation and the LLM evaluation was calculated as 0.71. This indicates that the evaluation results of the LLM and the human evaluation results are largely consistent, and there is a relatively obvious linear variation pattern between the two. Figure 4 in Appendix E shows the results of the human evaluation. The results indicate that **LOGIC** outperforms STORM across various dimensions. Detailed human evaluation results are provided in Appendix E.

Since the sophistication of content generated by LLMs reaches new heights, it is increasingly difficult for humans to discern the subtle differences between high-quality texts crafted by these models, human evaluation is not as obvious as automatic evaluation. However, from another perspective, it also proves the long-form outlines generated by our method are of higher quality, that is, more logical, coherent, novel, and knowledge-density-rich.

7 Conclusion

We propose **LOGIC**, a LLM-based long-form outline writing framework that emulates the reflection and imitation abilities of human writers. We curate the WikiOutline dataset and establish new evaluation criteria to more comprehensively study the generation of long-form outlines. Both automated and human evaluations indicate that **LOGIC** can generate high-quality long-form outlines that are coherent and logical. Moreover, our approach is model-agnostic and can achieve good results even on an 8-billion-parameter small-scale model. This flexibility and efficiency open up new possibilities for its application in resource-constrained environments. In the future, we will further expand the application of **LOGIC** in various fields and promote the automatic long-form outline writ-

ing technology to develop from theoretical research to wide-ranging practical applications. And, we plan to explore fine-tuning the model with more diverse and carefully curated datasets, so as to generate more interesting, user-friendly, and objective long-form outlines. Additionally, to narrow the performance gap of small-parameter models, we will investigate the application of reinforcement learning approaches to iteratively optimize small models, which will enable them to deliver high-quality long-form outlines while maintaining the advantages of low computational cost and fast inference speed.

Limitations

In this study, we explore the automatic generation of long-form outlines for expository articles, preparing for the writing of such articles. Although **LOGIC** has shown its advantages in both automatic and human evaluations, there are still some limitations, especially in terms of interestingness and objectivity. In terms of interestingness, since **LOGIC** mainly focuses on knowledge integration and logical structure construction, the generated long-form outlines often struggle to attract readers. Regarding objectivity, although **LOGIC** attempts to combine internal model knowledge with external retrieval to achieve logical coherence, ensuring absolute objectivity remains a challenge. In addition, the way the model integrates knowledge from multiple sources may not always result in a balanced presentation, leading to outlines that may not objectively represent all aspects of complex topics. In the future, we plan to explore fine-tuning the model with more diverse and carefully curated datasets, as well as incorporating new algorithms to more effectively analyze and present information in an engaging and unbiased manner, so as to generate more interesting, user-friendly, and objective long-form outlines.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (No. 62376245), the Key Research and Development Program of Zhejiang Province, China (No. 2024C01034), the Fundamental Research Funds for the Central Universities (226-2024-00170), National Key Research and Development Project of China (No. 2018AAA0101900), and MOE Engineering Research Center of Digital Library.

References

- Minwook Bae and Hyoungun Kim. 2024. [Collective critics for creative story generation](#). *arXiv preprint arXiv:2410.02428*.
- Yushi Bai, Jiajie Zhang, Xin Lv, Linzhi Zheng, Siqi Zhu, Lei Hou, Yuxiao Dong, Jie Tang, and Juanzi Li. 2024. Longwriter: Unleashing 10,000+ word generation from long context llms, 2024. URL <https://arxiv.org/abs/2408.07055>.
- Nishant Balepur, Jie Huang, and Kevin Chen-Chuan Chang. 2023. [Expository text generation: Imitate, retrieve, paraphrase](#). *arXiv preprint arXiv:2305.03276*.
- Siddhartha Banerjee and Prasenjit Mitra. 2015. [Wikikreator: Improving wikipedia stubs automatically](#). In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 867–877.
- Eric Chamoun, Michael Schlichktrull, and Andreas Vlachos. 2024. [Automated focused feedback generation for scientific writing assistance](#). *arXiv preprint arXiv:2405.20477*.
- Ziyang Chen, Xiaobin Wang, Yong Jiang, Jinzhi Liao, Pengjun Xie, Fei Huang, and Xiang Zhao. 2024. [An adaptive framework for generating systematic explanatory answer in online q&a platforms](#). *arXiv preprint arXiv:2410.17694*.
- Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, et al. 2024. [The llama 3 herd of models](#). *arXiv preprint arXiv:2407.21783*.
- Darren Edge, Ha Trinh, Newman Cheng, Joshua Bradley, Alex Chao, Apurva Mody, Steven Truitt, and Jonathan Larson. 2024. [From local to global: A graph rag approach to query-focused summarization](#). *arXiv preprint arXiv:2404.16130*.
- Angela Fan and Claire Gardent. 2022. [Generating full length wikipedia biographies: The impact of gender bias on the retrieval-based generation of women biographies](#). *arXiv preprint arXiv:2204.05879*.
- Pasi Fränti and Radu Marinescu-Istodor. 2023. [Soft precision and recall](#). *Pattern Recognit. Lett.*, 167:115–121.
- Luyu Gao, Xueguang Ma, Jimmy Lin, and Jamie Callan. 2022. [Precise zero-shot dense retrieval without relevance labels](#). *arXiv preprint arXiv:2212.10496*.
- SPJM (Serge) Horbach and W (Willem) Halfman. 2018. [The changing forms and expectations of peer review](#). *Research integrity and peer review*, 3:1–15.
- Lei Huang, Jiaming Guo, Guanhua He, Xishan Zhang, Rui Zhang, Shaohui Peng, Shaoli Liu, and Tianshi Chen. 2024. [Ex3: Automatic novel writing by extracting, excelsior and expanding](#). *arXiv preprint arXiv:2408.08506*.
- Gautier Izacard and Edouard Grave. 2020. [Leveraging passage retrieval with generative models for open domain question answering](#). *arXiv preprint arXiv:2007.01282*.
- Gautier Izacard, Patrick Lewis, Maria Lomeli, Lucas Hosseini, Fabio Petroni, Timo Schick, Jane Dwivedi-Yu, Armand Joulin, Sebastian Riedel, and Edouard Grave. 2023. [Atlas: Few-shot learning with retrieval augmented language models](#). *Journal of Machine Learning Research*, 24(251):1–43.
- Yucheng Jiang, Yijia Shao, Dekun Ma, Sina J Semnani, and Monica S Lam. 2024. [Into the unknown unknowns: Engaged human learning through participation in language model agent conversations](#). *arXiv preprint arXiv:2408.15232*.
- Omar Khattab, Arnav Singhvi, Paridhi Maheshwari, Zhiyuan Zhang, Keshav Santhanam, Saiful Haq, Ashutosh Sharma, Thomas T Joshi, Hanna Moazam, Heather Miller, et al. 2024. [Dspy: Compiling declarative language model calls into state-of-the-art pipelines](#). In *The Twelfth International Conference on Learning Representations*.
- Seungone Kim, Jamin Shin, Yejin Cho, Joel Jang, Shayne Longpre, Hwaran Lee, Sangdoo Yun, Seongjin Shin, Sungdong Kim, James Thorne, and Minjoon Seo. 2023. [Prometheus: Inducing fine-grained evaluation capability in language models](#). *Preprint*, arXiv:2310.08491.
- Yukyung Lee, Soonwon Ka, Bokyung Son, Pilsung Kang, and Jaewook Kang. 2024. [Navigating the path of writing: Outline-guided text generation with large language models](#). *arXiv preprint arXiv:2404.13919*.
- Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. 2020. [Retrieval-augmented generation for knowledge-intensive nlp tasks](#). *Advances in Neural Information Processing Systems*, 33:9459–9474.
- Zhuoqun Li, Xuanang Chen, Haiyang Yu, Hongyu Lin, Yaojie Lu, Qiaoyu Tang, Fei Huang, Xianpei Han, Le Sun, and Yongbin Li. 2024. [Struc-trag: Boosting knowledge intensive reasoning of llms via inference-time hybrid information structurization](#). *arXiv preprint arXiv:2410.08815*.
- Yi Liang, You Wu, Honglei Zhuang, Li Chen, Jiaming Shen, Yiling Jia, Zhen Qin, Sumit Sanghai, Xuanhui Wang, Carl Yang, et al. 2024. [Integrating planning into single-turn long-form text generation](#). *arXiv preprint arXiv:2410.06203*.
- Peter J Liu, Mohammad Saleh, Etienne Pot, Ben Goodrich, Ryan Sepassi, Lukasz Kaiser, and Noam Shazeer. 2018. [Generating wikipedia by summarizing long sequences](#). *arXiv preprint arXiv:1801.10198*.

- Zhiheng Lyu, Kevin Yang, Lingpeng Kong, and Daniel Klein. 2024. [Facttrack: Time-aware world state tracking in story outlines](#). *arXiv preprint arXiv:2407.16347*.
- Yan Ma, Yu Qiao, and Pengfei Liu. 2024. [Mops: Modular story premise synthesis for open-ended automatic story generation](#). *arXiv preprint arXiv:2406.05690*.
- Gpt OpenAI. 2024. [4o mini: Advancing cost-efficient intelligence, 2024](#). URL: <https://openai.com/index/gpt-4o-mini-advancing-cost-efficient-intelligence>.
- Hongjin Qian, Zheng Liu, Peitian Zhang, Kelong Mao, Yujia Zhou, Xu Chen, and Zhicheng Dou. 2024a. [Are long-lms a necessity for long-context tasks?](#) *arXiv preprint arXiv:2405.15318*.
- Hongjin Qian, Peitian Zhang, Zheng Liu, Kelong Mao, and Zhicheng Dou. 2024b. [Memorag: Moving towards next-gen rag via memory-inspired knowledge discovery](#). *arXiv preprint arXiv:2409.05591*.
- Priyanka Ranade and Anupam Joshi. 2023. [Fabula: Intelligence report generation using retrieval-augmented narrative construction](#). In *Proceedings of the International Conference on Advances in Social Networks Analysis and Mining*, pages 603–610.
- Vipula Rawte, Amit Sheth, and Amitava Das. 2023. [A survey of hallucination in large foundation models](#). *arXiv preprint arXiv:2309.05922*.
- Mark O Riedl and Robert Michael Young. 2010. [Narrative planning: Balancing plot and character](#). *Journal of Artificial Intelligence Research*, 39:217–268.
- D Gordon Rohman. 1965. Pre-writing: The stage of discovery in the writing process. *College Composition & Communication*, 16(2):106–112.
- Sina J Semnani, Violet Z Yao, Heidi C Zhang, and Monica S Lam. 2023. [Wikichat: Stopping the hallucination of large language model chatbots by few-shot grounding on wikipedia](#). *arXiv preprint arXiv:2305.14292*.
- Yijia Shao, Yucheng Jiang, Theodore A Kanell, Peter Xu, Omar Khattab, and Monica S Lam. 2024. [Assisting in writing wikipedia-like articles from scratch with large language models](#). *arXiv preprint arXiv:2402.14207*.
- Qianyu Wang, Jinwu Hu, Zhengping Li, Yufeng Wang, Yu Hu, Minghui Tan, et al. 2024a. [Generating long-form story using dynamic hierarchical outlining with memory-enhancement](#). *arXiv preprint arXiv:2412.13575*.
- Qiyao Wang, Shiwen Ni, Huaren Liu, Shule Lu, Guhong Chen, Xi Feng, Chi Wei, Qiang Qu, Hamid Alinejad-Rokny, Yuan Lin, et al. 2024b. [Autopatent: A multi-agent framework for automatic patent generation](#). *arXiv preprint arXiv:2412.09796*.
- Yidong Wang, Qi Guo, Wenjin Yao, Hongbo Zhang, Xin Zhang, Zhen Wu, Meishan Zhang, Xinyu Dai, Min Zhang, Qingsong Wen, et al. 2024c. [Autosurvey: Large language models can automatically write surveys](#). *arXiv preprint arXiv:2406.10252*.
- Zekun Xi, Wenbiao Yin, Jizhan Fang, Jialong Wu, Runnan Fang, Ningyu Zhang, Jiang Yong, Pengjun Xie, Fei Huang, and Huajun Chen. 2025. [Omnithink: Expanding knowledge boundaries in machine writing through thinking](#). *arXiv preprint arXiv:2501.09751*.
- Fangyuan Xu, Yixiao Song, Mohit Iyyer, and Eunsol Choi. 2023. [A critical evaluation of evaluations for long-form question answering](#). *arXiv preprint arXiv:2305.18201*.
- An Yang, Baosong Yang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Zhou, Chengpeng Li, Chengyuan Li, Dayiheng Liu, Fei Huang, Guanting Dong, Haoran Wei, Huan Lin, Jialong Tang, Jialin Wang, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Ma, Jin Xu, Jingren Zhou, Jinze Bai, Jinzheng He, Junyang Lin, Kai Dang, Keming Lu, Keqin Chen, Kexin Yang, Mei Li, Mingfeng Xue, Na Ni, Pei Zhang, Peng Wang, Ru Peng, Rui Men, Ruize Gao, Runji Lin, Shijie Wang, Shuai Bai, Sinan Tan, Tianhang Zhu, Tianhao Li, Tianyu Liu, Wenbin Ge, Xiaodong Deng, Xiaohuan Zhou, Xingzhang Ren, Xinyu Zhang, Xipin Wei, Xuancheng Ren, Yang Fan, Yang Yao, Yichang Zhang, Yu Wan, Yunfei Chu, Yuqiong Liu, Zeyu Cui, Zhenru Zhang, and Zhihao Fan. 2024. [Qwen2 technical report](#). *arXiv preprint arXiv:2407.10671*.
- Kevin Yang, Yuandong Tian, Nanyun Peng, and Dan Klein. 2022. [Re3: Generating longer stories with recursive reprompting and revision](#). *arXiv preprint arXiv:2210.06774*.
- Lili Yao, Nanyun Peng, Ralph Weischedel, Kevin Knight, Dongyan Zhao, and Rui Yan. 2019. [Plan-and-write: Towards better automatic storytelling](#). In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 7378–7385.
- Wangchunshu Zhou, Yuchen Eleanor Jiang, Peng Cui, Tiannan Wang, Zhenxin Xiao, Yifan Hou, Ryan Cotterell, and Mrinmaya Sachan. 2023. [Recurrentgpt: Interactive generation of \(arbitrarily\) long text](#). *arXiv preprint arXiv:2305.13304*.

A WikiOutline Dataset Details

As described in §4.1, we curate the WikiOutline dataset by collecting outlines of recent and high-quality English Wikipedia articles. We selected 80 pages from October 2023 to September 2024, and filtered out some low-quality articles to ensure that the vast majority of articles were evaluated as GA or above by ORES4. WikiOutline is a multi-domain dataset that contains outlines of Wikipedia articles in five categories: Characters, Events, Films, Disasters, and Places. People in characters categories hail from various countries, possess diverse ethnic and cultural backgrounds, are active in different epochs, and also vary in international influence. The events categories cover a wide range of events, including international conflicts, various violent incidents, sports events, political activities, and historical battles. The films categories cover various genres such as science fiction, action, animation, and variety shows, presenting a rich and diverse range of film and television entertainment content. The disasters categories mainly include events related to meteorological disasters (hurricanes, typhoons) and geological disasters (earthquakes) that occurred in different regions during the period from 2023 to 2024. The places categories cover different types of geographical entities, including little-known locations, countries, regions, sea areas, regions in specific historical periods, as well as specific engineering projects and religious buildings. Detailed experimental results on the five domains of characters, events, movies, disasters, and places in the WikiOutline dataset can be found in Table 6. The quantity of each category can be referred to Table 7. If the paper is accepted, we will make our dataset WikiOutline public.

Characters	10
Events	20
Films	20
Disasters	20
Places	10

Table 6: Statistics of the WikiOutline dataset used in our experiments.

B Implementation

Using the DSPy (Khatab et al., 2024) framework, we constructed **LOGIC** with zero-shot prompts and compared three different LLMs: GPT-4o-mini

(OpenAI, 2024), Qwen2-72B-Instruct-AWQ (the quantized version of Qwen2-72B-Instruct (Yang et al., 2024)), and Llama-3.1-8B-Instruct (Dubey et al., 2024). We implemented experiments on the FreshWiki (Shao et al., 2024) and WikiOutline datasets, with the temperature set to 1.0 and the top_p set to 0.9.

C Automatic Evaluation Details

The calculation methods of the two metrics, Soft Heading Recall and Heading Entity Recall, are exactly the same as those in STORM. Additionally, referring to the evaluation methods and metrics for articles in STORM, we also use the LLM evaluator Prometheus (Kim et al., 2023) to evaluate long-form outlines. The scoring criteria for long-form outline quality evaluation can be found in Table 9.

D Pseudocode of Logic

In §3, we introduced the specific process of **LOGIC**, a system for generating long-form outlines via iterative imitative and critical self-refinement. We implemented the zero-shot prompting **LOGIC** using the DSPy framework (Khatab et al., 2024). Listing 1, 2, 3, 4, 5, 6, 7, and 8 record the complete prompts for writing outline plans, writing draft outline, writing draft premise, refining draft premise, finding related entities, writing outline, writing feedback, and refining outline based on the feedback. **LOGIC** offers a general framework for creating detailed long-form outlines, with its core advantage being the reduced dependence on single-domain prompt engineering and thus having broader applicability and universality. The pseudocode of **PremiseGen** can be found at Algorithm 2.

Algorithm 2 PremiseGen

Input: Topic $\mathcal{T}$

Output: Premise $\mathcal{P}$

- 1: $\mathcal{D}$ for Definition, $\mathcal{B}$ for background, $\mathcal{E}$ for Events, $\mathcal{S}$ for Significance;
 - 2: Draft Premise $\mathcal{P}_D = \text{LLM}(\mathcal{T})$
 - 3: $\mathcal{D} = \text{LLM}(\mathcal{T})$
 - 4: $\mathcal{B} = \text{LLM}(\mathcal{T}, \mathcal{D})$
 - 5: $\mathcal{E} = \text{LLM}(\mathcal{T}, \mathcal{D}, \mathcal{B})$
 - 6: $\mathcal{S} = \text{LLM}(\mathcal{T}, \mathcal{D}, \mathcal{B}, \mathcal{E})$
 - 7: $\mathcal{P} = \text{LLM}(\mathcal{T}, \mathcal{P}_D, \mathcal{D}, \mathcal{B}, \mathcal{E}, \mathcal{S})$
 - 8: $\mathcal{P} = \text{LLM}(\mathcal{T}, \mathcal{P}, \mathcal{O}_D)$
 - 9: **return** $\mathcal{P}$
-

		Heading Soft Recall	Heading Entity Recall	Rubric Grading						
				Interest	Organization	Relevance	Coverage	Logicity	Breadth	Depth
Characters	Direct Gen	95.08	54.33	2.20	2.00	3.10	3.80	3.60	4.40	2.30
	RAG	92.59	46.67	2.20	2.30	3.10	3.80	3.70	4.20	2.40
	P-RAG	91.79	51.67	2.00	2.00	3.00	3.90	3.60	4.80	2.20
	STORM	95.24	57.67	2.60	2.50	3.70	3.80	3.90	4.40	3.40
	OmniThink	93.60	56.67	3.10	3.50	4.00	3.20	4.00	4.30	3.40
	LOGIC	97.54	68.92	4.00	5.00	5.00	5.00	4.33	4.89	4.89
Events	Direct Gen	96.69	64.17	2.40	3.50	3.10	3.55	3.75	4.70	2.75
	RAG	93.52	65.83	2.40	2.00	3.00	3.10	3.15	4.75	2.20
	P-RAG	91.88	67.50	2.30	2.60	3.00	3.30	3.90	4.75	2.45
	STORM	97.12	64.17	3.15	5.00	3.60	3.80	4.00	4.55	4.30
	OmniThink	96.88	65.00	3.77	4.70	3.35	3.45	3.85	4.30	3.20
	Logic	98.06	66.67	3.94	5.00	5.00	4.94	4.47	4.76	4.75
Films	Direct Gen	98.24	95.00	2.05	2.15	3.10	3.30	3.40	4.85	2.45
	RAG	94.65	95.00	2.10	2.00	3.00	3.20	2.85	4.85	2.05
	P-RAG	84.17	95.00	2.00	2.10	3.00	3.45	3.30	4.60	2.30
	STORM	98.29	95.00	3.05	4.10	3.75	3.45	4.00	4.65	3.10
	OmniThink	97.43	95.00	3.15	3.05	3.25	3.05	3.80	4.20	2.90
	LOGIC	98.37	96.67	4.00	5.00	5.00	5.00	4.35	5.00	4.94
Disasters	Direct Gen	93.80	72.50	2.05	2.00	3.25	3.40	3.60	4.80	2.50
	RAG	97.43	70.00	2.40	2.00	3.10	3.45	3.20	4.95	2.30
	P-RAG	98.97	70.00	2.25	2.00	3.10	3.50	3.40	5.00	2.30
	STORM	94.51	72.50	3.55	4.85	3.75	3.85	4.00	4.90	3.20
	OmniThink	98.30	70.00	3.75	4.10	3.70	3.50	3.95	4.65	3.05
	LOGIC	99.83	72.50	4.00	5.00	5.00	5.00	4.42	5.00	4.89
Places	Direct Gen	94.96	55.00	2.20	3.20	3.20	2.90	3.20	4.80	2.30
	RAG	95.76	50.00	2.10	2.30	3.00	3.20	3.40	4.60	2.40
	P-RAG	95.65	50.00	1.90	1.90	2.80	3.50	3.10	4.30	2.00
	STORM	94.41	55.00	2.90	4.10	3.60	3.70	3.70	4.60	2.60
	OmniThink	96.02	50.00	3.40	3.70	3.60	3.70	3.60	4.20	2.80
	LOGIC	98.81	56.43	4.00	5.00	5.00	4.90	4.44	4.67	4.33

Table 7: Results of automatic outline quality evaluation of WikiOutline dataset on GPT-4o-mini. The table shows the results for the five categories and the last row indicates the percentage improvement over the best baseline. The rubric grading uses a 1-5 scale.

E Human Evaluation Details

As discussed in §6, in order to compare the pros and cons of the outlines generated by **LOGIC** and **STORM**, we invited five volunteers to evaluate 10 randomly selected samples. We have designed scoring criteria that are more suitable for human evaluation. Volunteers are required to rate from four dimensions according to the scoring criteria. The specific scoring criteria are shown in Table 10. We summarized the scores given by the human evaluators and calculated the average value, the maximum value, and the minimum value. Compared with **STORM**, **LOGIC** has improved by 10%, 7%, 12%, and 13% respectively on the four metrics. The detailed results are shown in Table 8. Figure 4 shows the results of the human evaluation.

F Case Study

In Listing 9 - 23, we show an example of a long-form outline generated by **LOGIC**. It is generated using GPT-4o-mini as the backbone. We can see that compared with other methods, the outline generated by **LOGIC** is more coherent and logical and has richer content, that is, each section contains more information.

		Average	Maximum	Minimum
Logicity	STORM	3.82	4.40	3.20
	LOGIC	4.20	4.60	3.80
Coherence	STORM	3.84	4.20	3.40
	LOGIC	4.12	4.80	3.60
Knowledgeability	STORM	3.96	4.40	3.60
	LOGIC	4.42	5.00	3.80
Novelty	STORM	3.34	3.80	2.80
	LOGIC	3.76	4.00	3.60

Table 8: Detailed results of human evaluation.

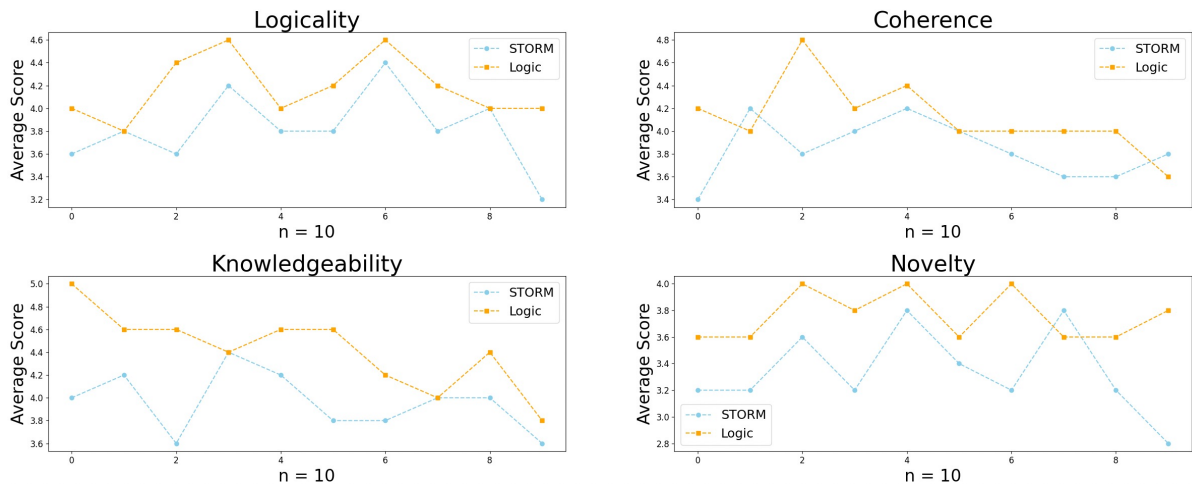


Figure 4: Comparison of **LOGIC** and **STORM** results under human evaluation.

Criteria Description	Interest Level: How engaging and thought-provoking is the outline?
Score 1 Description	Not engaging at all; no attempt to capture the reader's attention.
Score 2 Description	Fairly engaging with a basic narrative but lacking depth.
Score 3 Description	Moderately engaging with several interesting points.
Score 4 Description	Quite engaging with a well-structured narrative and noteworthy points that frequently capture and retain attention.
Score 5 Description	Exceptionally engaging throughout, with a compelling narrative that consistently stimulates interest.
Criteria Description	Coherence and Organization: Is the outline well-organized and logically structured?
Score 1 Description	Disorganized; lacks logical structure and coherence.
Score 2 Description	Fairly organized; a basic structure is present but not consistently followed.
Score 3 Description	Organized; a clear structure is mostly followed with some lapses in coherence.
Score 4 Description	Good organization; a clear structure with minor lapses in coherence.
Score 5 Description	Excellent organization; the outline is logically structured with seamless transitions and a clear argument.
Criteria Description	Relevance and Focus: Does the outline stay on topic and maintain a clear focus?
Score 1 Description	Off-topic; the content does not align with the headline or core subject.
Score 2 Description	Somewhat on topic but with several digressions; the core subject is evident but not consistently adhered to.
Score 3 Description	Generally on topic, despite a few unrelated details.
Score 4 Description	Mostly on topic and focused; the narrative has a consistent relevance to the core subject with infrequent digressions.
Score 5 Description	Exceptionally focused and entirely on topic; the outline is tightly centered on the subject, with every piece of information contributing to a comprehensive understanding of the topic.
Criteria Description	Broad Coverage: Does the outline provide an in-depth exploration of the topic and have good coverage?
Score 1 Description	Severely lacking; offers little to no coverage of the topic's primary aspects, resulting in a very narrow perspective.
Score 2 Description	Partial coverage; includes some of the topic's main aspects but misses others, resulting in an incomplete portrayal.
Score 3 Description	Acceptable breadth; covers most main aspects, though it may stray into minor unnecessary details or overlook some relevant points.
Score 4 Description	Good coverage; achieves broad coverage of the topic, hitting on all major points with minimal extraneous information.
Score 5 Description	Exemplary in breadth; delivers outstanding coverage, thoroughly detailing all crucial aspects of the topic without including irrelevant information.
Criteria Description	Logical Structure of the Outline: Is the logical sequence of ideas presented in the outline clear?
Score 1 Description	Completely disorganized. Ideas are listed randomly without any logic and cannot be connected.
Score 2 Description	There is a basic structure, but the logic often breaks down, transitions are unclear, and the organization is chaotic.
Score 3 Description	The structure is relatively clear. Most ideas are logical, but some connections are not well-defined.
Score 4 Description	The logical structure is excellent. Ideas are presented clearly and orderly, with smooth transitions.
Score 5 Description	The logical structure is perfect. Ideas progress step by step, and the argument is rigorous.
Criteria Description	Breadth of Coverage in the Outline: To what extent does the outline cover different aspects and dimensions of the topic?
Score 1 Description	The scope is extremely narrow, focusing on only one or two aspects and ignoring the most relevant dimensions.
Score 2 Description	The breadth is limited, covering a few aspects, missing important content, and presenting an incomplete view.
Score 3 Description	The breadth is moderate, covering most major aspects, but some content is briefly mentioned or not fully covered.
Score 4 Description	The breadth is good, covering many aspects, providing a comprehensive overview with only minor omissions.
Score 5 Description	The coverage is comprehensive, exploring every aspect and dimension of the topic without omission.
Criteria Description	Depth of Analysis in the Outline: How thoroughly does the outline analyze the details and implications of each aspect of the topic?
Score 1 Description	The analysis is only superficial, merely listing facts without exploring underlying causes and relationships.
Score 2 Description	The analysis is shallow, providing basic details but not delving into meanings, consequences, and potential connections.
Score 3 Description	The analysis has some depth, but some content can be further explored for subtleties.
Score 4 Description	In-depth analysis, comprehensively exploring details, implications, and connections of each aspect, with only a few areas for improvement.
Score 5 Description	The analysis is extremely in-depth, digging into every detail, implication, and subtle relationship without blind spots.

Table 9: Scoring rubrics on a 1-5 scale for the evaluator LLM.

Criteria Description	Logicity: How logical is the outline?
Score 1 Description	Completely illogical; ideas are disjointed and unorganized.
Score 2 Description	Some basic logical structure, but with significant flaws and jumps in reasoning.
Score 3 Description	Moderately logical with a clear flow of ideas, but some minor inconsistencies.
Score 4 Description	Quite logical with a well-structured sequence of ideas and sound reasoning.
Score 5 Description	Exceptionally logical, with a tight and flawless logical progression throughout.
Criteria Description	Coherence: How coherent is the outline?
Score 1 Description	No coherence; parts of the outline seem unrelated to each other.
Score 2 Description	Fairly coherent in basic aspects, but lacks smooth transitions between ideas.
Score 3 Description	Moderately coherent, with most ideas connected in a reasonable way.
Score 4 Description	Quite coherent, with seamless transitions and a unified overall theme.
Score 5 Description	Exceptionally coherent, where every part contributes to the whole in a harmonious way.
Criteria Description	Knowledgeability: How knowledgeable is the outline?
Score 1 Description	Lacks any real knowledge; contains only common - sense or incorrect information.
Score 2 Description	Some basic knowledge is presented, but it is shallow and not well-developed.
Score 3 Description	Moderately knowledgeable, with a reasonable amount of relevant and accurate information.
Score 4 Description	Quite knowledgeable, with extensive and well-researched information.
Score 5 Description	Exceptionally knowledgeable, presenting an abundance of diverse and in-depth knowledge.
Criteria Description	Novelty: How novel is the outline?
Score 1 Description	Entirely conventional; no new or interesting ideas.
Score 2 Description	Have a few minor new elements, but overall is quite traditional.
Score 3 Description	Moderately novel, with some fresh perspectives or approaches.
Score 4 Description	Quite novel, with a significant number of innovative and thought-provoking ideas.
Score 5 Description	Exceptionally novel, presenting groundbreaking and highly original concepts.

Table 10: Scoring rubrics on a 1-5 scale for human evaluation.

```

class WritePageOutlinePlan(dspy.Signature):
    """You want to write a detailed outline for a Wikipedia page, and start by writing a detailed outline
    plan for the Wikipedia page. You should write a detailed outline plan for each paragraph of the
    Wikipedia page based on the topic, making the Wikipedia page as comprehensive as possible.
    Here is the format of your writing:
    1. Please return it in JSON format, not markdown format, such as {"outline plan 1": "content", "
    outline plan 2": "content", ...}.
    2. Each plan should occupy a separate line, that is, there should be a line break after each plan.
    3. Please refer to the example, do not include other information.
    ##example
    {
        "outline plan 1": "Introduction: Provide an overview of Taylor Hawkins as an influential
        American musician and drummer for Foo Fighters. Highlight his musical achievements,
        collaboration with other artists, and mention his untimely death in 2022.",
        "outline plan 2": "Early Life and Education: Detail Hawkins' birth in 1972, his childhood in
        Texas and California, and his early interest in music. Discuss any significant events or
        experiences that led to his pursuit of a career in music.",
        "outline plan 3": "Musical Beginnings: Chronicle Hawkins' initial forays into music, including
        his first bands and gigs. Explore his early influences and the development of his musical
        style during these formative years.",
        "outline plan 4": "Joining Foo Fighters: Narrate the circumstances that led to Hawkins joining
        the Foo Fighters in 1997. Describe his audition, integration into the band, and the
        evolution of his role within the group.",
        "outline plan 5": "Contributions to Foo Fighters: Analyze Hawkins' contributions to the band's
        albums and live performances. Discuss key songs and albums where his influence is
        particularly notable.",
        "outline plan 6": "Side Projects and Collaborations: Document Hawkins' musical projects outside
        of Foo Fighters, including Taylor Hawkins and the Coattail Riders. Highlight collaborations
        with other artists and any solo work.",
        "outline plan 7": "Musical Style and Influences: Examine Hawkins' drumming technique, his
        musical influences, and how they shaped his style. Include commentary from music critics
        and fellow musicians.",
        "outline plan 8": "Personal Life: Share insights into Hawkins' life outside of music, including
        his family, hobbies, and other interests. Discuss his role in the music community and any
        philanthropic efforts.",
        "outline plan 9": "Death: Provide a detailed account of the circumstances surrounding Hawkins'
        death on March 25, 2022. Include the official statements, the music community's reaction,
        and the impact of his passing.",
        "outline plan 10": "Legacy and Tributes: Reflect on Hawkins' legacy in the music industry,
        including tributes from fellow artists, commemorative events, and his influence on future
        generations of musicians.",
        "outline plan 11": "Discography: Compile a comprehensive list of Hawkins' discography, including
        albums with Foo Fighters, his side projects, and notable guest appearances on other
        artists' recordings.",
        "outline plan 12": "Awards and Nominations: List the awards and nominations Hawkins received
        throughout his career, with a focus on his contributions to Foo Fighters as well as any
        individual accolades.",
        "outline plan 13": "References: Assemble a thorough list of references, including interviews,
        articles, and official statements, to provide verifiable sources for the information
        presented in the article.",
        "outline plan 14": "External Links: Curate a selection of external links to official websites,
        social media profiles, and other relevant online resources that offer additional
        information about Taylor Hawkins.",
        "outline plan 15": "See Also: Recommend related Wikipedia articles that provide context or
        additional information about topics mentioned in the article, such as Foo Fighters'
        discography, notable drummers, or music awards."
    }
    """
    topic = dspy.InputField(prefix="The topic you want to write: ", format=str)
    outline_plan = dspy.OutputField(prefix="Write a detailed outline plan for the Wikipedia page:\n")

```

Listing 1: Prompts used for writing outline plans in LOGIC.

```

class WritePageOutline(dspy.Signature):
    """Write a detailed outline for this section of the Wikipedia page, including as many entities as possible, based on the topic and the outline plan for the Wikipedia page. Here is the format of your writing:
    1. Use "## Title" to indicate the section title, do not use the topic as a title, the section title must differ from the topic. Use "### Title" to indicate subsection title. Note that each section title has at least three subsection titles (if there are subsection titles). Make sure there are no more than three levels of outlines.
    2. The outline levels must be progressive and no level can be skipped.
    3. Do not include other information.
    """

    topic = dspy.InputField(prefix="The topic you want to write: ", format=str)
    section_outline_plan = dspy.InputField(prefix="The outline plan for this section: \n", format=str)
    previous_sections_outlines = dspy.InputField(prefix="The previous sections' outlines: \n", format=str)
    section_outline = dspy.OutputField(prefix='Write a detailed outline for this section of the Wikipedia page(Use "## Title" to indicate the section title(only the first line), do not use the topic as a title, the section title must differ from the topic. Use "### Title" to indicate subsection title.): \n')

```

Listing 2: Prompts used for writing draft outline in LOGIC.

```

class PolishPagePremise(dspy.Signature):
    """Polish a premise for the Wikipedia page. Please polish the draft premise to include more information based on the given topic, definition, background, events, significance and impact. Here is the format of your writing:
    1. Please only output one paragraph.
    2. Do not include other information.
    """

    draft_premise = dspy.InputField(prefix="The draft premise: ", format=str)
    topic = dspy.InputField(prefix="The topic you want to write: ", format=str)
    definition = dspy.InputField(prefix="The definition of the topic: ", format=str)
    background = dspy.InputField(prefix="The background of the topic: ", format=str)
    event = dspy.InputField(prefix="The events of the topic: ", format=str)
    significance_and_impact = dspy.InputField(prefix="The significance and impact of the topic: ", format=str)
    premise = dspy.OutputField(prefix="Polish a premise for the Wikipedia page:\n")

```

Listing 3: Prompts used for writing draft premise in LOGIC.

```

class RefinePagePremise(dspy.Signature):
    """Refine a premise for the Wikipedia page. Please refine the draft premise to include more information based on the given topic and the outline for the Wikipedia page. The premise should cover the main information about the topic, including its definition, background, important events, significance and impact, key concepts and main ideas, etc. Here is the format of your writing:
    1. Please only output one paragraph.
    2. Do not include other information.
    """

    topic = dspy.InputField(prefix="The topic you want to write: ", format=str)
    premise = dspy.InputField(prefix="The draft premise for the Wikipedia page: ", format=str)
    outline = dspy.InputField(prefix="The outline for the Wikipedia page:\n", format=str)
    refine_premise_and_outline = dspy.OutputField(prefix="Refine a premise for the Wikipedia page:\n")

```

Listing 4: Prompts used for refining draft premise in LOGIC.

```

class FindRelatedEntity(dspy.Signature):
    """I'm writing a Wikipedia page for a topic mentioned below. Please identify and recommend some Wikipedia pages on closely related entities or topics. I'm looking for examples that provide insights into interesting aspects commonly associated with this topic, or examples that help me understand the typical content and structure included in Wikipedia pages for similar entities or topics. Please list 10 urls in separate lines. Each url should start with https://en.wikipedia.org/wiki/."""

    topic = dspy.InputField(prefix='Topic of interest:', format=str)
    related_entities = dspy.OutputField()

```

Listing 5: Prompts used for finding related entities in LOGIC.

```

class WritePageOutlineFromReo(dspy.Signature):
    """Improve a detailed outline for this section of the Wikipedia page. You already have a draft outline
    for this section of the Wikipedia page that covers the general information. Please improve it based
    on the premise for the Wikipedia page, the draft outline for the Wikipedia page, and the
    information learned from the outlines of related entities to make it more informative, but you
    should still use the draft outline for this section of the Wikipedia page as the main body.
    Here is the format of your writing:
    1. Use "# Title" to indicate the section title, do not use the topic as a title, the section title
       must differ from the topic. Use "## Title" to indicate subsection title, "### Title" to
       indicate subsubsection title, and so on. Note that each section title has at least three
       subsection titles (if there are subsection titles), each subsection title has at least two
       subsubsection titles (if there are subsubsection titles), each subsubsection title has at least
       two subsubsubsection titles (if there are subsubsubsection titles). Make sure there are no
       more than four levels of outlines.
    2. The outline levels must be progressive and no level can be skipped.
    3. Do not include other information.
    """

    topic = dspy.InputField(prefix="The topic you want to write: ", format=str)
    premise = dspy.InputField(prefix="The premise for the Wikipedia page: ", format=str)
    related_entities_outlines = dspy.InputField(prefix="Related entities outlines:\n", format=str)
    previous_sections_outlines = dspy.InputField(prefix="The previous sections' outlines:\n", format=str)
    draft_page_outline = dspy.InputField(prefix="The draft outline for the Wikipedia page:\n", format=str)
    draft_section_outline = dspy.InputField(prefix="The draft outline for this section of the Wikipedia page
    :\n", format=str)
    outline = dspy.OutputField(prefix='Improve a detailed outline for this section of the Wikipedia page(Use
    "# Title" to indicate the section title, do not use the topic as a title, the section title must
    differ from the topic. Use "## Title" to indicate subsection title, "### Title" to indicate
    subsubsection title, ...):\n')

```

Listing 6: Prompts used for writing outline in LOGIC.

```

class WritePageOutlineFeedback(dspy.Signature):
    """Provide three pieces of feedback to refine the outline for this section of the Wikipedia page. You
    already have a draft outline for this section of the Wikipedia page that covers the general
    information. Please provide three pieces of feedback to refine the outline based on the draft
    outline for the Wikipedia page, and the information learned from the outlines of related entities
    to make it more informative.
    Here is the format of your writing:
    1. Please return it in JSON format, not markdown format, such as {"feedback 1": "content", "feedback
       2": "content", "feedback 3": "content"}.
    2. Each feedback should occupy a separate line, that is, there should be a line break after each
       feedback.
    3. Please refer to the example, do not include other information.
    ##example
    {
        "feedback 1": "To make this section more informative, consider adding more specific details
            about Hawkins' childhood, family, and early life experiences, as these have been known to
            shape his musical style and career.",
        "feedback 2": "In the 'Discography and Touring' section, you may want to expand on Hawkins'
            contributions to Foo Fighters' discography, such as his work on specific albums, and
            notable live performances.",
        "feedback 3": "You could also consider adding more context to the 'Awards and Recognition'
            section by including specific accolades, such as the number of awards won, and notable
            mentions in reputable music publications."
    }
    """

    topic = dspy.InputField(prefix="The topic you want to write: ", format=str)
    related_entities_outlines = dspy.InputField(prefix="Related entities outlines:\n", format=str)
    previous_sections_outlines = dspy.InputField(prefix="The previous sections' outlines:\n", format=str)
    draft_page_outline = dspy.InputField(prefix="The draft outline for the Wikipedia page:\n", format=str)
    draft_section_outline = dspy.InputField(prefix="The draft outline for this section of the Wikipedia page
    :\n", format=str)
    feedback = dspy.OutputField(prefix='Provide three pieces of feedback to refine the outline for this
    section of the Wikipedia page:\n')

```

Listing 7: Prompts used for writing feedback in LOGIC.

```

class RefinePageOutlinebaseFeedback(dspy.Signature):
    """Refine a detailed outline for this section of the Wikipedia page based on the feedback. You already
    have a draft outline for this section of the Wikipedia page that covers the general information.
    Please improve it based on the feedback and the information learned from the outlines of related
    entities to make it more informative, but you should still use the draft outline for this section
    of the Wikipedia page as the main body.
    Here is the format of your writing:
    1. Use "# Title" to indicate section title, do not use topic as a section title. Use "## Title" to
    indicate subsection title, "### Title" to indicate subsubsection title, and so on. Note that
    each section title has at least three subsection titles (if there are subsection titles), each
    subsection title has at least two subsubsection titles (if there are subsubsection titles),
    each subsubsection title has at least two subsubsubsection titles (if there are
    subsubsubsection titles). Make sure there are no more than four levels of outlines.
    2. Do not include other information.
    """
    topic = dspy.InputField(prefix="The topic you want to write: ", format=str)
    related_entities_outlines = dspy.InputField(prefix="Related entities outlines:\n", format=str)
    feed_back = dspy.InputField(prefix="The feedback to refine the outline for this section of the Wikipedia
    page:\n", format=str)
    draft_section_outline = dspy.InputField(prefix="The draft outline for this section of the Wikipedia page
    :\n", format=str)
    refine_outline = dspy.OutputField(prefix='Refine a detailed outline for this section of the Wikipedia
    page based on the feedback:\n')

```

Listing 8: Prompts used for refining outline based on the feedback in **LOGIC**.

```

## The Journey of a Musical Innovator
### Overview of Taylor Hawkins
#### Brief Biography
#### Significance in Music
### Early Life and Musical Beginnings
#### Family Background
#### Growing Up in Texas and California
#### Formation of Musical Interests
### Musical Influences
#### Key Influences in Hawkins' Drumming Style
#### Impact of Influences on His Career
### Career with Foo Fighters
#### Joining the Band
#### Role in the Band
#### Achievements and Milestones
### Collaborations and Side Projects
#### Taylor Hawkins and the Coattail Riders
#### Collaborations with Other Artists
#### Solo Projects and Other Ventures
### Legacy and Impact
#### Influence on Future Generations of Drummers
#### Commemoration by Peers and Fans
#### Enduring Influence on Rock Music
### Untimely Death
#### Circumstances Surrounding His Passing
#### Reactions to His Death

```

Listing 9: Section 1 of the writing long-form outline with Taylor Hawkins as the topic.

```

## Foundations of a Musical Journey
### Birth and Family Background
#### Family Heritage
#### Early Exposure to Music
### Childhood in Texas and California
#### Moving to California
#### California Music Culture
### Initial Musical Interests and Experiences
#### First Encounters with Drumming
#### Formation of First Bands
#### Development of Musical Style

```

Listing 10: Section 2 of the writing long-form outline with Taylor Hawkins as the topic.

```

## Foundations of a Musical Career
### Early Influences
#### Family Musical Background
#### Key Drummers and Musicians
#### Musical Genres and Their Impact
### First Bands and Gigs
#### Formation of Initial Bands
#### Local Music Scene Engagement
#### Memorable Performances
### Development of Musical Style
#### Evolution of Drumming Techniques
#### Songwriting and Composition Skills
#### Artistic Growth Over Time

```

Listing 11: Section 3 of the writing long-form outline with Taylor Hawkins as the topic.

```

## Joining the Band
### Recruitment Process
#### Background Context
#### Initial Considerations
### The Audition Process
#### Preparing for the Audition
#### The Audition Experience
### Integration into the Group
#### Initial Rehearsals
#### Building Musical Chemistry
#### Touring and Public Reception
### Evolution of Role Within the Band
#### Expanding Responsibilities

```

Listing 12: Section 4 of the writing long-form outline with Taylor Hawkins as the topic.

```

## Impact on Foo Fighters' Success
### Contributions to Musical Catalog
#### Key Tracks Featuring Hawkins' Drumming
#### Songwriting Contributions
#### Influence on Albums
### Enhancing Live Performances
#### Stage Presence and Performance Style
#### Memorable Concerts and Tours
### Collaborations within Foo Fighters
#### Collaborative Songwriting Efforts
#### Instrumental Contributions Beyond Drumming
### Enduring Legacy in Foo Fighters
#### Influence on Band's Identity
#### Commemorations by Band Members

```

Listing 13: Section 5 of the writing long-form outline with Taylor Hawkins as the topic.

```

## Diverse Musical Pursuits of Taylor Hawkins
### Taylor Hawkins and the Coattail Riders
#### Formation of the Band
#### Discography and Major Releases
#### Tours and Live Performances
### Collaborations with Other Artists
#### Notable Collaborations
#### Influence on Other Musicians
### Solo Projects and Personal Endeavors
#### Solo Musical Releases
#### Personal Interests and Hobbies
#### Reflections on Artistic Growth

```

Listing 14: Section 6 of the writing long-form outline with Taylor Hawkins as the topic.

```

## Exploration of Musical Style and Influences
### Hawkins' Drumming Technique
#### Signature Techniques
#### Gear and Equipment
### Influences on Hawkins' Musical Style
#### Influential Drummers
#### Genre Influences
### Impact on Contemporary Drummers and Musicians
#### Influence on the Rock Genre
#### Broader Influence Across Genres
### Commentary from Music Critics and Fellow Musicians
#### Critical Reception of Hawkins' Work
#### Testimonials from Peers

```

Listing 15: Section 7 of the writing long-form outline with Taylor Hawkins as the topic.

```

## Personal Insights into Taylor Hawkins
### Family and Relationships
#### Relationships
#### Parenting
### Hobbies and Personal Interests
#### Passion for Sports
#### Other Artistic Pursuits
#### Travel and Exploration
### Philanthropic Efforts and Community Engagement
#### Charitable Organizations
#### Advocacy and Awareness
#### Legacy of Philanthropy
### Personal Values and Beliefs
#### Philosophical Views
#### Influence on Music
### Legacy Beyond Music
#### Cultural Impact
#### Commemorative Actions

```

Listing 16: Section 8 of the writing long-form outline with Taylor Hawkins as the topic.

```

## Reflecting on the Legacy of a Music Icon
### Circumstances Surrounding His Passing
#### Date and Location of Death
#### Discovery and Initial Reports
#### Official Statements from Authorities
### Reactions to His Death
#### Tributes from Fellow Musicians
#### Fan Reactions and Memorials
### Impact on the Music Community
#### Effect on Foo Fighters' Activities
#### Broader Industry Response
### Achievements and Milestones
#### Career Highlights
#### Personal Milestones
### Lasting Legacy
#### Commemorative Events and Honors
#### Influence on Future Generations of Musicians
#### Preservation of His Artistic Contributions
### Conclusion
#### Summary of His Lasting Influence on Music
#### Reflections on His Life, Career, and Contributions to Rock Music

```

Listing 17: Section 9 of the writing long-form outline with Taylor Hawkins as the topic.

```

## Commemoration and Influence in the Music Industry
### Tributes from Fellow Artists
#### Social Media Tributes
#### Public Statements and Interviews
#### Musical Tributes
### Commemorative Events
#### Memorial Concerts
#### Tributes at Awards Shows
#### Initiatives and Foundations
### Lasting Impact on Future Generations
#### Influence on Drummers
#### Impact on Rock Bands
#### Broader Musical Legacy

```

Listing 18: Section 10 of the writing long-form outline with Taylor Hawkins as the topic.

```

## Musical Contributions and Collaborations in Rock Music
### Contributions to Foo Fighters' Discography
#### Key Albums Featuring Hawkins
#### Notable Songs and Their Impact
#### Live Album Contributions
### Side Projects and Personal Endeavors
#### Other Musical Ventures
#### Solo Work
### Guest Appearances and Collaborations
#### Collaborations with Prominent Artists
#### Impact on Other Musicians
#### Music Festivals and Special Performances

```

Listing 19: Section 11 of the writing long-form outline with Taylor Hawkins as the topic.

```

## Recognitions in the Music Industry
### Overview of Awards and Nominations
#### Historical Context of Awards in Rock Music
#### Significance of Recognition in a Musician's Career
### Awards Won by Taylor Hawkins
#### Grammy Awards
#### Other Prestigious Awards
#### Industry Recognitions
### Nominations and Honors
#### Notable Nominations
#### Honors from Music Organizations
#### Lifetime Achievement Awards
### Cultural Significance of Awards and Recognitions
#### Reflection of Broader Trends in Music Industry
#### Influence on Future Musicians
### Impact of Awards on Legacy
#### Lasting Impression on the Music Industry
#### Reflection on Contributions

```

Listing 20: Section 12 of the writing long-form outline with Taylor Hawkins as the topic.

```

## Compendium of References for Taylor Hawkins
### Overview of Reference Types
#### Primary Sources
#### Secondary Sources
#### Tertiary Sources
### Notable Publications
#### Biographies
#### Journalistic Articles
### Academic Contributions
#### Research Articles
#### Theses and Dissertations
### Fan Contributions and Community Resources
#### Online Fan Sites and Forums
#### Community Memorials and Tributes
### Digital Resources and Databases
#### Music Streaming Platforms
#### Video Interviews and Documentaries
#### Archival Materials

```

Listing 21: Section 13 of the writing long-form outline with Taylor Hawkins as the topic.

```

## Digital Footprint and Commemoration
### Introduction
#### Overview of Taylor Hawkins' Impact on Music
#### Legacy and Influence on Future Generations
### Official Websites
#### Foo Fighters Official Site
#### Taylor Hawkins and the Coattail Riders Official Site
#### Memorial Pages and Tributes
### Social Media Profiles
#### Taylor Hawkins' Official Accounts
#### Foo Fighters' Social Media Accounts
#### Community Engagement and Fan Interaction
### Interviews and Articles
#### Major Interviews
#### Articles on His Contributions
#### Retrospectives by Music Publications
### Music and Media Platforms
#### Streaming Services
#### Documentary and Tribute Films
### Tributes and Memorials
#### Fan-Made Memorials
#### Commemorative Events
#### Charitable Initiatives in His Memory
#### Notable Artist Tributes

```

Listing 22: Section 14 of the writing long-form outline with Taylor Hawkins as the topic.

```

## Expanded Reading Suggestions
### Foo Fighters
#### Overview of the Band
#### Taylor Hawkins' Role and Contributions
#### Live Performances and Tours
#### Legacy and Recognition
### Notable Drummers in Rock Music
#### Techniques and Styles
#### Contributions to Bands
### Music Awards and Recognition
#### Overview of Major Music Awards
#### Awards Related to Rock Music
#### Taylor Hawkins' Achievements
### Collaborations in Music
#### Importance of Collaborations
#### Taylor Hawkins' Collaborations
#### Broader Collaborative Trends in Rock

```

Listing 23: Section 15 of the writing long-form outline with Taylor Hawkins as the topic.

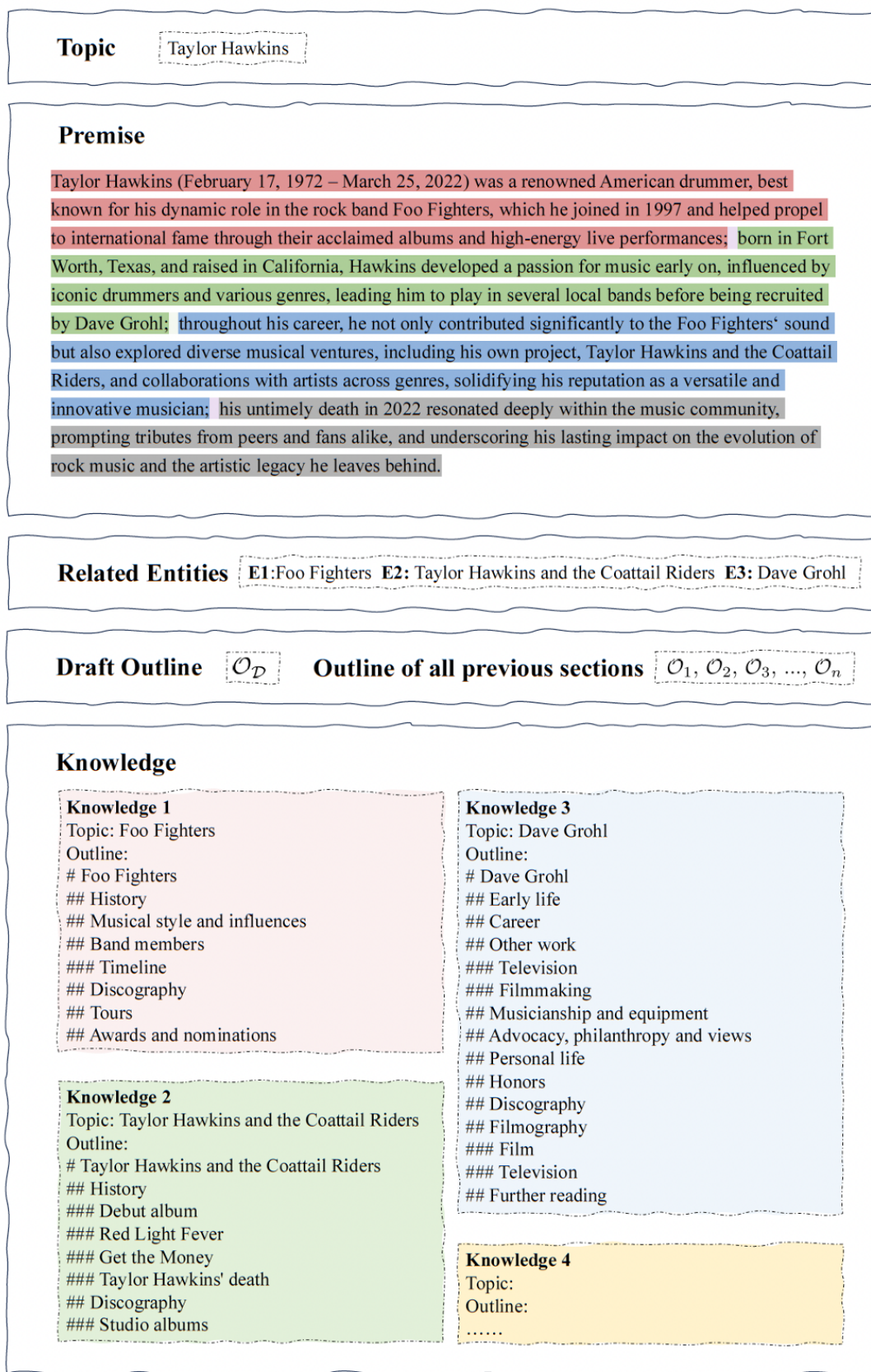


Figure 5: A case of LOGIC stage 4.