Boundary Matters: Leveraging Structured Text Plots for Long Text Outline Generation

Yuanchi Ma [1], **Jiamou Liu**[2], **Hui He**[1], **Libo Zhang**[2], **Haoyuan Li**[2], **Zhendong Niu** [1*]
[1]Beijing Institute of Technology, Beijing, China
[2]University of Auckland, Auckland, New Zealand
{yma,zniu}@bit.edu.cn

Abstract

Outline generation aims to uncover the internal content structure of a document by identifying potential chapter connections and generating corresponding summaries. A robust outline generation model strives for coherence between and within plots. However, existing methods perform well on short- and medium-length texts and struggle with generating readable outlines for very long texts (e.g., fictional literary works). The primary challenge lies in their inability to accurately segment plots within long texts. To address this issue, we propose a novel unsupervised guidance framework, LeStrTP, to guide large language model (LLM) outline generation. This framework ensures that each structured plot encapsulates complete causality by accurately identifying plot boundaries. Specifically, the LeStrTP constructs chapter-level graph from long texts and learns their embeddings. Subsequently, through Markov chain modeling chapter dependence, a unique search operator is designed to achieve plot segmentation. To facilitate research on this task, we introduce a new annotated benchmark dataset, NovOutlineSet. Experimental results demonstrate that structured plots not only enhance the coherence and integrity of generated outlines but also significantly improve their quality.

1 Introduction

Well-crafted stories typically consist of multiple semantically coherent chapters, each of which revolves around a particular theme. An good outline concisely captures the structural content of a story, offering clear guidance for navigation and significantly easing the cognitive load required to comprehend the entire text. Furthermore, it reveals the underlying thematic structure of the text (Shi et al., 2024; Jha et al., 2024; Song et al., 2024).

Prior research on outline generation (Zhang et al., 2019; Zhou et al., 2015a) has predominantly focused on short-to-medium texts (50-1,000 tokens) such as news articles and public notices, employing paragraph-level segmentation strategies to assist rapid content structure comprehension. While effective in domains like socio-economic analysis (Ma et al., 2023), these methods face critical challenges when processing fictional narratives exceeding 10,000 tokens (e.g., Greek mythology cycles, or novels like *Fights Break Sphere* and *The Count of Monte Cristo*). The extended narrative scope introduces multi-layered semantic dependencies and nonlinear thematic progression (Hertling and Paulheim, 2020), complicating traditional outline extraction. Although fictional works primarily serve entertainment purposes, their complex narrative architectures often encode subcultural paradigms and historical zeitgeist—exemplified by *Camel Xiangzi*'s critique of social stratification and *Fights Break Sphere*'s exploration of agency versus determinism. Automated outline generation (OG) for such texts provides dual benefits: enabling efficient plot navigation for general readers through hierarchical event graphs, and creating structured corpora for computational humanities research across history, media psychology, and cultural studies (Chu et al., 2021).

When dealing with long texts in the realm of fiction, the focus shifts from basic relationships like birthplace or spouse to the occurrence of specific plot events. Including alliances, enemies, clan members, betrayals, and vendetta (Hua et al., 2016a). Upon further analysis of the OG challenge, we observe that it involves two structured prediction tasks: 1) identifying chapter features and plot boundaries, and 2) generating chapter summaries. These two tasks correspond to predicting the hierarchical relationships between chapters and summarizing individual chapters. While the second task can be well-handled by existing LLM, particularly for shorter and medium-length texts. However, LLMs often exhibit inaccuracies and increased hallucinations when applied to longer texts

*Corresponding author

Findings of the Association for Computational Linguistics: EMNLP 2025, pages 49–63
November 4-9, 2025

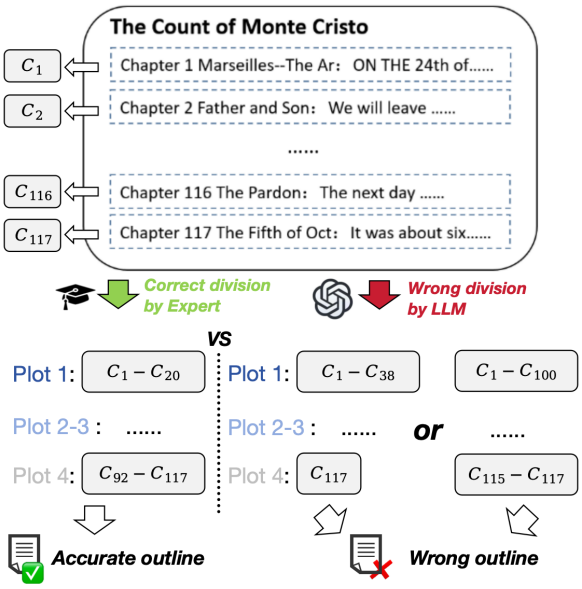


Figure 1: A demonstration of the process of generating long text outlines.

(Shen et al., 2020). For example, when the LLM is asked to summarize a work of more than one million words, it often ignores some important plots or produces a wrong division of plots, resulting in readers unable to understand the full text. As shown in Fig. 1. Another popular solution approach - fine-tuning - does not solve this problem very well either (Zhang et al., 2024; Goyal et al., 2024). The main reason lies in the catastrophic forgetting that may occur when fine-tuning the model, and also leads to excessive computational costs. Therefore, we explore whether an ideal OG framework could lightweight extend the strengths of LLM to long texts. And the key challenge of this framework is to address the tasks 1).

In this paper, we propose a novel end-to-end architecture, **Le**veraging **Str**uctured **T**ext **P**lots for outlines (LeStrTP), to address this challenge. The core idea is to enhance LLM outputs by enriching them with plot boundary information, which is first determined by a neural network and then used to guide the LLM in generating a detailed outline. We find that graph more effectively captures relationships between entities within chapters, thereby better reflecting chapter characteristics. Therefore, we begin by constructing the individual chapters into graphs. Our method first generates entity nodes using a chapter-level graph generation module and constructs an adjacency matrix between nodes based on syntactic dependency relationships. For every node feature vectors, we select entity word vectors and expand the feature set to include the TF-IDF matrix of entities while incorporating chapter numbers to represent contextual coherence. We then apply an enhanced graph neural network utilizing GAT (Velickovic et al., 2017) to learn from the chapter graph. Additionally, we design a search operator that uses properties of Markov chains to determine plot boundaries based on the potential distance between chapter embeddings. Finally, the theme and summary of each plot are generated by the LLM, which are then combined to form the overall outline.

To facilitate research, we have constructed a new Chinese benchmark dataset, named NovOutlineSet. The entire dataset contains ultra-long texts of 28 different topics. The average number of chapters is 580.8, and the maximum number of chapters in a single text is 2271. Experimental results demonstrate that our proposed method significantly outperforms all baselines in plot segmentation accuracy, with an average improvement of 45.2%. We also performed a detailed diversity and fluency analysis of the generated outlines, which showed a minimum performance improvement of 12.5% and 37.7% respectively. This demonstrates that our model can better understand the structure of the learned content. The contributions of our work are summarized as follows:

- We propose LeStrTP, a novel framework for long text outline generation, the first unsupervised framework to implement long text outline generation based on chapter-level graph.
- To support our study, we constructed a new annotated public dataset, NovOutlineSet, based on long Chinese fictional texts for the OG task. This is the first Chinese-language dataset containing multiple types of long texts, which can further promote broader research in the field of generation.
- Comprehensive evaluation of the outline generation results was conducted. The results of various indicators verify the effectiveness of our method.

2 Related Works

2.1 Topic Segmentation and Outline Generation

Early studies predominantly employed unsupervised methods (Choi, 2000; Glavas et al., 2016)

for topic segmentation. However, with the development of large-scale thematic structure corpora, supervised methods have become the dominant approach, including sequential labeling models (Badjatiya et al., 2018; Lukasik et al., 2020). In contrast, limited research has addressed topic segmentation, utilizing sequential labeling models inspired by English methodologies (Wang et al., 2016; Xing et al., 2020) or local classification models (Jiang et al., 2021) to predict topic boundaries.

Most existing OG methods (Zhang et al., 2019; Sun et al., 2022) are designed for short- and medium-length texts and predominantly focus on English content. Whether solving problems from a traditional model (Kitaev et al., 2020; Dai et al., 2019) perspective or from developing a "new generation" approach (Shen et al., 2020; Fu et al., 2019; Wiseman et al., 2018) to solving problems. Whereas our OG task generates a sequence of plot-level titles with fine-grained semantics and contextual coherence.

2.2 Storyline Generation

Storyline generation focuses on modeling event progression and temporal dynamics within narratives. This can provide ideological guidance for the OG task. Huang and Huang (2013a) established a foundational taxonomy by distinguishing local event granularity from global thematic evolution. Early approaches leveraged Bayesian networks for probabilistic storyline inference (Hua et al., 2016b; Zhou et al., 2015b), while Lin et al. (2012) introduced graph-based optimization for social media narrative extraction. Huang and Huang (2013b) advanced this through hierarchical Dirichlet processes (HDP) for theme evolution modeling. Recent work has expanded into multi-modal narrative analysis, with Yang et al. (2024a), Sui et al. (2023), and Yang et al. (2024b) pioneering structured frameworks for video storyline generation, achieving state-of-the-art performance on ActivityNet Captions. To summarise, the storyline generation task refers to marking the occurrence of each plot node in a story or text based on the timeline, and ultimately forming the occurrence axis of a storyline sequence. The OG task, on the other hand, requires generating a readable composite text to describe the overall narrative direction. Consequently, most existing methods for these related tasks are not directly applicable to the OG task.

3 Method

The long text OG task is defined as follows: Given a long text $B = \{c_1, c_2, ..., c_n\}, n \in \mathbb{N}$, where c_i denotes a chapter, we formulate the task as identifying plot boundaries $I_{ij} \in \{0, 1\}$ between chapter pairs (c_i, c_j). The segmentation results are entered into the LLM to obtain an accurate text outline.

In the transformation of the plot, there must be changes in scenes or the evolution of the relationship between characters and things accompanying it (Zhang et al., 2019). This variation corresponds to the frequency, or affiliation, of characters or objects appearing in each chapter. Therefore, by constructing chapter graph to explore relationship and achieve the separation of chapter features in the inner product space, the plot boundaries can be well distinguished. The detailed framework is illustrated in Fig. 2.

3.1 Data Preprocessing

For narrative texts of millions of words, we first need to segment the chapter content by matching the chapter names. Then, entity recognition and syntactic relation dependency analysis of chapter texts will be conducted through LTP (Che et al., 2021) to provide support for the subsequent construction of chapter diagrams. Step 1 in Fig. 2 presents this process very well.

3.2 Chapter-Level Graph Construction

For each chapter, it is essential to construct graph to encapsulate its content. This involves selecting chapter nodes and generating both feature vectors and adjacency matrices for nodes. We extract noun entities $X = \{x_1, x_2, ..., x_n\}, x_i \in \mathbb{R}^{n \times d}$ as chapter's nodes, encompassing crucial plot elements such as main characters, locations, and items within the chapter. d represents the node dimension. Syntactic dependency information provides the relationships $E = \{e_{ij}\}$ between entity nodes i and j, enabling the chapter's graph topology to be represented by an adjacency matrix A, where $A_{ij} = 1$ if $e_{ij} \in E$; otherwise, $A_{ij} = 0$.

We introduce a novel approach for constructing chapter node features, which incorporates three critical attributes: node entity name, chapter number, and entity node TF-IDF (T values). 1) For obtaining node entity name vectors En, we employ the bert-wmm (Cui et al., 2021), a state-of-the-art model for Chinese word vector representation. 2) To compute the T values for entity nodes, we first

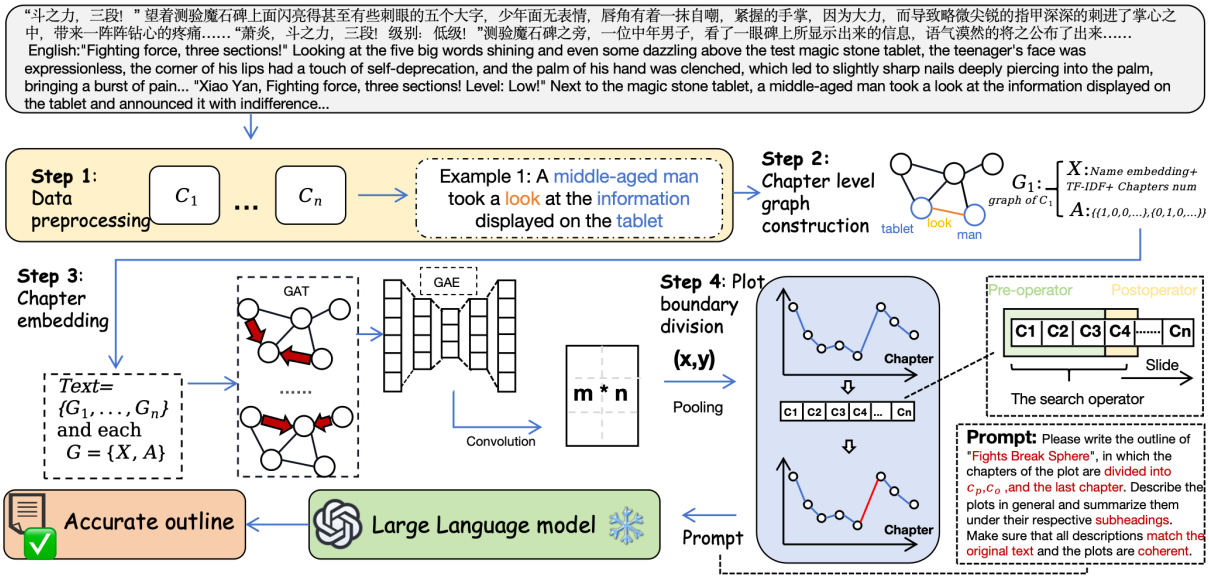


Figure 2: Our proposed LeStrTP flowchart. The proposed framework implements a four-stage pipeline: 1) Data preprocessing, 2) construction of chapter-level graphs, 3) graph embedding by GAT-based autoencoder, and 4) plot boundary identification. The LLM of the output layer freezes the parameters. In addition, the principle of the search operator is drawn in detail on the right side. Prompt is also detailed.

identify the top-10 T values within the chapter and then assign these values to the corresponding chapter nodes. Each entity node is matched with its respective T value. If an entity node corresponds to one of the top-10 T values, the value is added; otherwise, the matrix dimension value is set to 0. This process generates a 10-dimensional T value matrix Tn, where each row represents an entity node. 3) Considering that most fictional long texts follow sequential progression, we emphasize the inclusion of chapter counts. The chapter number is concatenated as a set of features, donate as Cn. Finally, the vectors of each node are combined from the corresponding representations in these three parts to give the node feature $x_i = \{En_i, Tn_i, Cn_i\}, i \in [1, n]$ of the graph.

3.3 Chapter Embedding

To learn chapter features, we propose an unsupervised learning framework based on a graph autoencoder with GAT (Velickovic et al., 2017; Ma et al., 2025) layers. This model extracts chapter features by jointly learning node attributes and the adjacency matrix. The hidden representation of the current node v_i is computed as follows:

$$Z_i^l = \sigma(\sum_{j \in N_i} \alpha_{ij} W Z_i^{l-1}) \quad (1)$$

α_{ij} denotes the attention coefficient, representing the relative importance of the neighboring node v_j to the target node v_i, while σ is a nonlinear activation function. W is a learnable parameter. Z_i^l denotes the output representation of node v_i, and N_i represents the set of neighboring nodes of v_i. The attention coefficient α_{ij} is computed through a single-layer feedforward neural network, which takes the concatenated attribute vectors $\vec{x_i}$ and $\vec{x_j}$ as input, with a trainable weight vector $\vec{a} \in \mathbb{R}^{n \times d}$.

$$\alpha_{ij} = \frac{exp(a(W\vec{x_i}, W\vec{x_j}))}{\sum_{k \in N_i} exp(a(W\vec{x_i}, W\vec{x_j}))} \quad (2)$$

To reconstruct the graph structure as part of the decoder, we employ the $sigmoid$ function to map values from $(-\infty, +\infty)$ to a probability space. The reconstruction error is minimized by quantifying the difference between the adjacency matrix A_i and its reconstructed counterpart A_i'.

$$L_r = \sum_{i=1}^{n} loss(A_i, A_i'), A_i' = sigmoid(Z^T Z) \quad (3)$$

The learned deep embeddings are subsequently reduced to low-dimensional data in the feature space using a pooling layer, resulting in the chapter embeddings $\hat{Z}$.

3.4 Plot Boundary Division

The plot boundary prediction is formulated as follows: Given a long Chinese text B, it is divided into

successive chapters $\{c_1, c_2, \ldots, c_n\}$. For each pair of chapters (c_i, c_j) where $j = i + 1$, a boundary label $I_{ij} \in \{0, 1\}$ is defined to indicate whether a plot boundary exists between the chapters. If $I_{ij} = 0$, chapters c_i and c_j are within the same plot boundary, and plot boundary prediction continues. If $I_{ij} = 1$, chapter c_i marks the plot boundary.

The properties of Markov chains can address the challenges of ensuring contextual consistency and identifying paragraph boundaries. Traditionally, Markov chain dependence assumes that the current chapter c_i is influenced solely by the preceding chapter c_{i-1}. However, in the context of very long fictional texts, a complete plot often spans multiple chapters (Chu et al., 2021). To account for this, we consider both the adjacency dependencies and the long-distance influences of preceding chapters. Accordingly, we propose a search operator that leverages chapter features to accurately identify plot boundaries.

The Search Operator consists of two components: 1) the pre-operator and 2) the postoperator, each with different stride, designed to capture short-range and long-range contextual information, respectively, as illustrated in the right side of Fig. 2.

Pre-operator: Set the stride to t. The embedding $\hat{Z}_i$ of the current chapter c_i and $\{\hat{Z}_{i-t} : \hat{Z}_{i-1}\}$ of the preceding t chapters c_{i-t} are fed into the pre-operator, which computes the mean Euclidean distance EU_1 between each chapter. Compared with the Manhattan distance or Chebyshev distance, the Euclidean distance is directly defined by the inner product, without the need for additional transformation and with low computational complexity $(O(n))$. It is more suitable for measuring the differences between scalars such as chapter features.

$$EU_1 = \frac{1}{t} \sum_{n=1}^{t} ||\hat{Z}_i - \hat{Z}_{i-t}||_2 \tag{4}$$

Considering the fluctuations between plots, the pre-operator also tracks the Euclidean distance EU_{max} corresponding to the maximum variance, which is recorded as the maximum fluctuation value.

$$EU_{max} = MAX(||\hat{Z}_i - \hat{Z}_{i-t}||_2) \tag{5}$$

A threshold $s_t = \beta \cdot EU_1$ is defined, where $\beta > 0$ represents the equilibrium parameter, allowing adaptation to varying text styles and plot fluctuations.

Postoperator: The stride is set to 1, meaning that only the Euclidean distance $EU_2 = ||z_i - z_j||_2$ between the current chapter c_i and the subsequent chapter c_j is computed.

If $s_t < EU_2 < EU_{max}$, $I_{ij} = 1$, chapter c_i is identified as a plot boundary. If $EU_2 < s_t$, $I_{ij} = 0$, the chapter is regarded as part of the same episode. The operator module continues searching for plot boundaries by iterating through subsequent chapters. Otherwise, c_i is regarded as an important plot boundary, a major plot twist.

An important consideration is the inherent variability in text length across different literary styles. Our framework addresses this through hyperparameter β, which dynamically adjusts the scale parameter s_t to ensure optimal model adaptation. The empirical validation of this adaptive mechanism is presented in Fig 5, demonstrating consistent performance across diverse stylistic domains.

To optimize computational efficiency and account for the continuity of long-text episodes, a safe distance d is introduced. Upon detecting a plot boundary at chapter c_i, the operator skips the next d chapters and resumes the plot boundary search from chapter c_{i+d}. This approach is based on the intuitive structure of long stories: following a plot twist, several successive chapters typically belong to the same plot, eliminating the need for evaluating every subsequent chapter.

The pipeline concludes by feeding chapter segmentation data and structural directives into our default LLM implementation (freeze DeepSeek-V3) for automated outline generation.

4 Experiments

4.1 Experiments Setup

Dataset. Among all the public dataset, we have not found a suitable dataset of tens of millions of Chinese characters. So we constructed a Chinese dataset NovOutlineSet comprising 28 long texts to facilitate outline generation. Among them are one biography, three world classics, two martial arts novels, nine fairy tales, and thirteen fantasy novels. Three domain experts with experience in relevant works were invited to contribute. Using insights from forum discussions and encyclopedia, precise boundaries were determined. The average text size is 2234.63KB, comprising an average of 580.8 chapters. And the maximum number of chapters in a single text is 2271. And the size of it is 23.1MB. Table 1 provides an overview of the statistical information of the dataset. Examples of relevant content are provided in Appendix A.

Type	β value	Quantity	Average of plot
Classics	$\beta = 0.3$	3	3.67
Biography	$\beta = 0.3$	1	4.00
Wuxia	$\beta = 0.4$	2	4.00
Fantastic	$\beta = 0.6$	13	5.31
Immortal Heroes	$\beta = 0.7$	9	5.67

Table 1: NovOutlineSet Information Overview

Links to more detailed datasets can be found at *https://anonymous.4open.science/r/y-8E7D/*. All experiments are perform on four RTX 3090s. The LLM layer adopts the method of calling api and freezing parameters for experiments.It is worth noting that the texts selected in our dataset all have clear plot boundaries to prevent confusion in the narrative structure of extremely long texts. However, in reality, there are some novels that do not have clear boundaries. We will discuss this in the context of potential limitations.

Baseline. We not only use the LLMs method as the baseline, but also select appropriate neural network-based deep learning methods. LLMs baselines are as follows: 1) GPT-3.5 (Brown et al., 2020) and 2) GPT-4.0 (OpenAI, 2023): two kinds of GPT models, 4.0 exceeding 3.5 in both size and performance. 3) Llama-3.0 (Touvron et al., 2023): Excellent open source LLM. 4) Deepseek-V3 (Dai et al., 2024) and 5) Qwen2 (Bai et al., 2023): Two of the best-performing LLMs in Chinese. 6) BooookScore (Chang et al., 2024) and 7) Readagent (Lee et al., 2024): Two of the latest long text summarization techniques based on LLM obtain the text outline by iteratively reading the text and concatenating it.

Among the few DL methods, we have selected two representative ones: 8) CPTs (Jiang et al., 2024): The outline is generated through the constructed Chinese paragraph-level subject structure corpus. 9) HiStGen (Zhang et al., 2019): The OG task is formulated as a hierarchical and structured prediction problem, which is predicted and generated according to semantic perception and Markov chain paragraph dependence mechanism.

Evaluation Metric. Follow Sun et al. (2022) and Zhang et al. (2019), our main experimental design fully reflects the accuracy and fluency of the generated outline from four aspects: prediction boundary accuracy, diversity, human-machine recognition and syntactic fluency. 1) Accuracy(ACC): If all the plot boundaries of the text are predicted accurately, it is marked as a positive sample, otherwise it is negative. 2) Perplexity (PPL) (Li et al., 2016): Perplexity measures the fluency of generated text. 3) Average phrasal word count, Text diversity: Two indicators are used to measure plot richness. The first represents the complexity of the narrative text, while the second indicates the completeness of the plot content. Following the Ehara (2023), the calculation formula is: *Text diversity = (proportion of adverbs and conjunctions in a clause + average phrasal word count)* $\times$ *0.5*. 4) Sentence Coherence (Cui et al., 2021): Current metrics do not assess sentence-level fluency. We use the next sentence prediction (NSP) from the pre-trained BERT model as a measure of consistency between each sentence and its next sentence. We report the average NSP score for all consecutive sentence pairs in the generated text. 5) Adversarial Success (Sun et al., 2022): It is defined as the model's ability to successfully deceive trained evaluators into believing that machine-generated text is human-authored. A higher adversarial success rate indicates better text quality.

4.2 Main Result

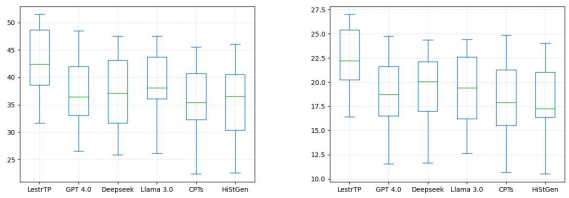
We evaluated boundary accuracy using our constructed dataset, as shown in Table 2. Our method demonstrates superior accuracy, with improvements of 35.6% and 28.6% compared to state-of-the-art LLM methods, Llama 3.0 and Deepseek-V3, respectively. We assessed the fluency, comprehensibility, and adversarial success of the generated outlines. In long text approximation generation, LLM-based models generally outperform deep learning strategies in fluency and adversarial success but exhibit lower comprehensibility. This indicates the advantages of LLMs in long text generation, though hallucination issues detract from text comprehensibility. The precise division of plot boundaries effectively improves fluency and readability. They are up at least 10.8% and 0.1% percent, respectively.

We conducted text diversity and readability experiments, text diversity is up at least 2.5%, as shown in Fig. 3(b). Our method produces more words per clause, resulting in richer paragraphs and a more diverse outline. This aligns with our goal: to include key details in the outline, enabling readers to better grasp the full text's development.

To summarise, our approach achieved superior performance across all metrics, producing more fluid, coherent, and varied text. The consistent

Type of method	Baseline	ACC(%) ↑	PPL ↓	Adversarial Success ↑	Sentence Coherence (NSP) ↑
LLM method	*GPT 3.5*	17.8	34.3	0.029	0.852
	GPT 4.0	28.5	31.3	0.101	0.871
	Llama 3.0	35.7	30.2	0.097	0.872
	Deepseek-V3	42.8	28.6	0.102	0.878
	Qwen2	30.2	27.3	0.038	0.856
	BooookScore	22.1	31.0	0.030	0.860
	Readagent	10.8	27.0	0.025	0.859
DL method	*CPTs*	14.2	39.9	0.029	0.832
	HiStGen	17.8	34.3	0.016	0.819
Ours	***LeStrTP (Default)***	**71.4**	**17.8**	**0.104**	**0.879**
	GAT's effectiveness	64.2	24.3	0.096	0.876
	Dimensionality reduction	60.7	30.0	0.078	0.875
	LLM performance	71.4	20.8	0.101	0.872

Table 2: Comparison of results with other baseline experiments and the results of the ablation experiment



(a) Average phrasal word count (b) Text diversity

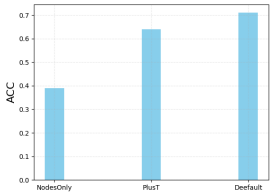
Figure 3: All long text generates outline readability data graph.

gains highlight the importance of integrating the text generation capabilities of LLMs with the predictive strengths of deep learning frameworks. It is worth mentioning that the calculation of the standard deviation is meaningless in our experimental design because this model is essentially for accurately predicting the plot boundaries.

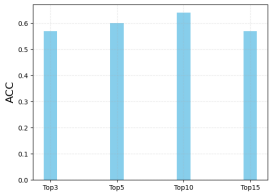
Ablation Experiments. Followed Ma et al. (2026), we conducted ablation experiments on several key modules of the model to evaluate its stability and the contributions of individual components. The modifications include: 1) Effectiveness of GAT: Replacing the GAT layer to assess its impact. We use the normal GNN layer instead of GAT. 2) Dimensionality reduction: Evaluating the dimensionality reduction strategy to determine its effect on algorithm prediction outcomes. 3) LLM performance: Replacing the default Deepseek-V3 used for outline generation with GPT 4.0 to analyze its impact on the model. As shown in lower part of Table 2, modifications in 1) and 2) resulted in the poorest performance, highlighting the critical role of the model in learning chapter data features.

In order to find the best form of chapter vector representation, we break down the three parts of chapter vector, which are 1) NodesOnly embedded with node name only, 2) PlusT with T value added on this basis, and 3) add the Default setting of chapter sequence number default. In addition, we also delve into the optimal vector dimension required by T value. As can be seen from Fig. 4, T value and chapter number are crucial for long linear narrative texts, which contain more plot information. This information has often been overlooked in previous studies. Effective learning of these can significantly improve the accuracy of plot boundary prediction.

Parameter Sensitivity Analysis. We conducted a parameter sensitivity analysis, focusing first on an in-depth examination of the retrieval operator, which is a critical component for predicting plot boundaries. The goal of this experiment is to determine the optimal parameters, including the stride for the preceding and following operators and the β in the threshold s_t calculation. The results of our experiment in the category of fantasy fiction are shown in Fig. 5, which clearly demonstrate that the stride and β have a significant impact on model performance. The best stride is $\lfloor n \times 0.025 \rfloor$. This highlights that, under classical Markov properties, only adjacent chapter dependencies are considered. To address cross-chapter correlations in long plots, our method incorporates embedded information from the first t chapters, significantly enhancing the sensitivity to remote information. Moreover, a critical observation is the significant length variation across literary styles, necessitating style-specific β parameterization. The complete mapping between stylistic categories and their corresponding β val-



(a) The influence of chapter vector composition on ACC



(b) T-value analysis

Figure 4: Chapter vector composition analysis

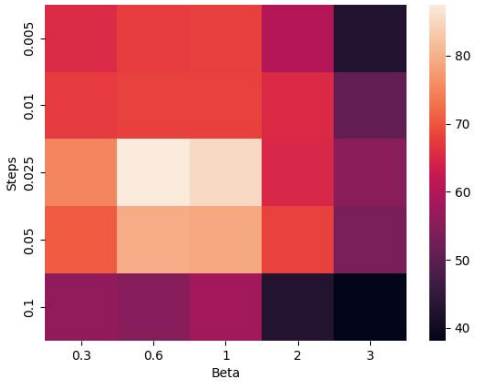


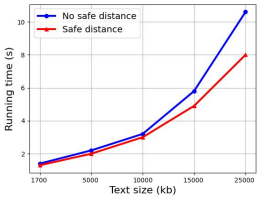
Figure 5: Parameter sensitivity analysis of search operators. The experimental results show the influence of β and strides on the prediction accuracy ACC. Steps represents the proportion of the number of strides to the total number of chapters.

ues is systematically documented in Table 4 (Appendix A), providing a comprehensive reference for model adaptation.

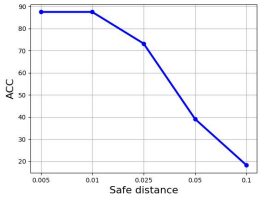
Dynamic hyperparameters are common problems in natural language processing. Take the Transformer as an example: position encoding needs to adapt to different sequence lengths. Therefore, we analyzed the impact of the safety distance d on computation time, as shown in Fig. 6(a). The results indicate that the safety distance strategy effectively reduces redundant calculations without significantly affecting detection accuracy, making it suitable for large-scale long text processing. Then we experiment to find the most suitable d. As shown in Fig. 6(b), the optimal safe distance d should be $\lfloor n \times 0.01 \rfloor$.

4.3 Case Study

To visually demonstrate the difference between LeStrTP and other Baselines, *Fights Break Sphere* was chosen to demonstrate the effect of plot segmentation and summarization. This book is one



(a) The effect of safe distance



(b) The effect on ACC

Figure 6: Parameter sensitivity analysis of safety distance. The number in safe distance represents the proportion of the number of steps to the total number of chapters.

Method	Plot segmentation	Content summary
LeStrTP	plot 1:[1,343]	The fall and rise of genius
	plot 2:[344,704]	The Battle of Cashing in the Black Corner
	plot 3 : [705,1200]	The Fire Collection and the Spirit Conspiracy
	plot 4 : [1201,1631]	The Final battle and the battle to break the sky
GPT 4.0	plot 1: [1,40]	The loser strikes back, reignites the fight
	plot 2: [41,240]	College practice, showing the edge
	plot 3: [240,600]	Ancient ruins, looking for opportunities
	plot 4: [600,1050]	the family rises again
	plot 5: [1051,1631]	Strife within the clan, a battle of destiny

Table 3: Generate outline and plot division results

of the most influential works of Chinese online literature. There is almost no ambiguity in the division of the plot boundaries of this book, and no more expert filtering is needed. According to the correct division, the correct plot boundary for the *Fights Break Sphere* is [343,705,1200]. The results are shown in Table 3. LeStrTP correctly captures plot boundaries and generates an accurate outline, but GPT 4.0 has a serious hallucination problem. In-depth analysis reveals that imprecise narrative segmentation significantly compromises GPT's outline generation capability. While a subset of plot summaries demonstrates acceptable accuracy, the majority exhibit substantial deviations from ground truth narrative structures. In addition, we provide not only several detailed outline examples, but also a text of visualizations that demonstrate the effectiveness of our chapter-level approach. The corresponding experimental data are shown in Fig.9 to Fig.12 in appendix A.

5 Conclusion

In this paper, we propose a method based on plot segmentation to guide LLM in generating better outlines for long texts. First, the chapter-level graph effectively captures chapter feature information. Based on the chapter embeddings learned by the GAT, we divide plot boundaries use im-

proved Markov chain properties. Finally, LLM generates a complete outline based on precise plot boundaries. Comparative experiments demonstrate that our method not only enhances the accuracy of outline generation but also improves fluency and diversity. This approach addresses a gap in Chinese language research, and aids readers in quickly understanding the content of ultra-long texts, facilitates scholars across various fields in conducting relevant humanistic research. The future research direction is to further explore how to solve the effects of hyperparameters, so as to reduce expert intervention.

6 Limitations

To the best of our knowledge, the primary limitation of this model resides in the selection of hyperparameter search step size and β. Suboptimal parameter choices may lead to degraded prediction accuracy. We are exploring an automatic parameter selection method based on data field moisture principles, moving beyond expert experience dependence. Furthermore, the text partitioning process exhibits instability due to its reliance on specific keyword detection for chapter segmentation—an approach that proves ineffective for specially named long documents. Consequently, chapter division at the word embedding level emerges as a promising research direction. Moreover, some novel novels may even complete the entire story in the same setting. Therefore, the plot boundaries of such novels are not easy to determine or are ambiguous. Future research may need to determine the plot boundaries from more dimensions.

Another potential limitation that needs to be explained is that perhaps adding a better manually generated outline would be beneficial for enhancing the diversity of the dataset. However, for this task, the cost of marking high-quality long text transitions is extremely high (requiring domain experts over 10 hours per text). In the future, we will consider this issue when we have a higher budget and more time.

References

Pinkesh Badjatiya, Litton J. Kurisinkel, Manish Gupta, and Vasudeva Varma. 2018. Attention-based neural text segmentation. In *ECIR*, volume 10772 of *Lecture Notes in Computer Science*, pages 180–193. Springer.

Jinze Bai, Shuai Bai, Yunfei Chu, Zeyu Cui, Kai Dang, Xiaodong Deng, Yang Fan, Wenbin Ge, Yu Han, Fei Huang, Binyuan Hui, Luo Ji, Mei Li, Junyang Lin, Runji Lin, Dayiheng Liu, Gao Liu, Chengqiang Lu, Keming Lu, Jianxin Ma, Rui Men, Xingzhang Ren, Xuancheng Ren, Chuanqi Tan, Sinan Tan, Jianhong Tu, Peng Wang, Shijie Wang, Wei Wang, Shengguang Wu, Benfeng Xu, Jin Xu, An Yang, Hao Yang, Jian Yang, Shusheng Yang, Yang Yao, Bowen Yu, Hongyi Yuan, Zheng Yuan, Jianwei Zhang, Xingxuan Zhang, Yichang Zhang, Zhenru Zhang, Chang Zhou, Jingren Zhou, Xiaohuan Zhou, and Tianhang Zhu. 2023. Qwen technical report. *CoRR*, abs/2309.16609.

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners. In *NeurIPS*.

Yapei Chang, Kyle Lo, Tanya Goyal, and Mohit Iyyer. 2024. Booookscore: A systematic exploration of book-length summarization in the era of llms. In *ICLR*. OpenReview.net.

Wanxiang Che, Yunlong Feng, Libo Qin, and Ting Liu. 2021. N-LTP: an open-source neural language technology platform for chinese. In *EMNLP (Demos)*, pages 42–49. Association for Computational Linguistics.

Freddy Y. Y. Choi. 2000. Advances in domain independent linear text segmentation. In *ANLP*, pages 26–33. ACL.

Cuong Xuan Chu, Simon Razniewski, and Gerhard Weikum. 2021. Knowfi: Knowledge extraction from long fictional texts. In *AKBC*.

Yiming Cui, Wanxiang Che, Ting Liu, Bing Qin, and Ziqing Yang. 2021. Pre-training with whole word masking for chinese BERT. *IEEE ACM Trans. Audio Speech Lang. Process.*, 29:3504–3514.

Damai Dai, Chengqi Deng, Chenggang Zhao, R. X. Xu, Huazuo Gao, Deli Chen, Jiashi Li, Wangding Zeng, Xingkai Yu, Y. Wu, Zhenda Xie, Y. K. Li, Panpan Huang, Fuli Luo, Chong Ruan, Zhifang Sui, and Wenfeng Liang. 2024. Deepseekmoe: Towards ultimate expert specialization in mixture-of-experts language models. In *ACL (1)*, pages 1280–1297. Association for Computational Linguistics.

Zihang Dai, Zhilin Yang, Yiming Yang, Jaime G. Carbonell, Quoc Viet Le, and Ruslan Salakhutdinov. 2019. Transformer-xl: Attentive language models beyond a fixed-length context. In *ACL (1)*, pages 2978–2988. Association for Computational Linguistics.

Yo Ehara. 2023. A novel interpretation of classical readability metrics: Revisiting the language model underpinning the flesch- kincaid index. In *ICCE*. Asia-Pacific Society for Computers in Education.

Yao Fu, Yansong Feng, and John P. Cunningham. 2019. Paraphrase generation with latent bag of words. In *NeurIPS*, pages 13623–13634.

Goran Glavas, Federico Nanni, and Simone Paolo Ponzetto. 2016. Unsupervised text segmentation using semantic relatedness graphs. In **SEM@ACL*. The *SEM 2016 Organizing Committee.

Sachin Goyal, Christina Baek, J. Zico Kolter, and Aditi Raghunathan. 2024. Context-parametric inversion: Why instruction finetuning may not actually improve context reliance. *CoRR*, abs/2410.10796.

Sven Hertling and Heiko Paulheim. 2020. Dbkwik: extracting and integrating knowledge from thousands of wikis. *Knowl. Inf. Syst.*, 62(6):2169–2190.

Ting Hua, Xuchao Zhang, Wei Wang, Chang-Tien Lu, and Naren Ramakrishnan. 2016a. Automatical storyline generation with help from twitter. In *CIKM*, pages 2383–2388. ACM.

Ting Hua, Xuchao Zhang, Wei Wang, Chang-Tien Lu, and Naren Ramakrishnan. 2016b. Automatical storyline generation with help from twitter. In *CIKM*, pages 2383–2388. ACM.

Lifu Huang and Lian'en Huang. 2013a. Optimized event storyline generation based on mixture-event-aspect model. In *EMNLP*, pages 726–735. ACL.

Lifu Huang and Lian'en Huang. 2013b. Optimized event storyline generation based on mixture-event-aspect model. In *EMNLP*, pages 726–735. ACL.

Prince Jha, Raghav Jain, Konika Mandal, Aman Chadha, Sriparna Saha, and Pushpak Bhattacharyya. 2024. Memeguard: An LLM and vlm-based framework for advancing content moderation via meme intervention. In *ACL (1)*, pages 8084–8104. Association for Computational Linguistics.

Feng Jiang, Yaxin Fan, Xiaomin Chu, Peifeng Li, Qiaoming Zhu, and Fang Kong. 2021. Hierarchical macro discourse parsing based on topic segmentation. In *AAAI*, pages 13152–13160. AAAI Press.

Feng Jiang, Weihao Liu, Xiaomin Chu, Peifeng Li, Qiaoming Zhu, and Haizhou Li. 2024. Advancing topic segmentation and outline generation in chinese texts: The paragraph-level topic representation, corpus, and benchmark. In *LREC/COLING*, pages 495–506. ELRA and ICCL.

Nikita Kitaev, Lukasz Kaiser, and Anselm Levskaya. 2020. Reformer: The efficient transformer. In *ICLR*. OpenReview.net.

Kuang-Huei Lee, Xinyun Chen, Hiroki Furuta, John F. Canny, and Ian Fischer. 2024. A human-inspired reading agent with gist memory of very long contexts. In *ICML*. OpenReview.net.

Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao, and Bill Dolan. 2016. A diversity-promoting objective function for neural conversation models. In *HLT-NAACL*, pages 110–119. The Association for Computational Linguistics.

Chen Lin, Chun Lin, Jingxuan Li, Dingding Wang, Yang Chen, and Tao Li. 2012. Generating event storylines from microblogs. In *CIKM*, pages 175–184. ACM.

Michal Lukasik, Boris Dadachev, Kishore Papineni, and Gonçalo Simões. 2020. Text segmentation by cross segment attention. In *EMNLP (1)*, pages 4707–4716. Association for Computational Linguistics.

Yuanchi Ma, Hui He, Zhongxiang Lei, Kaize Shi, Xueping Peng, and Zhendong Niu. 2025. Deepedd: Nonparametric deep graph clustering through effective distance and density. *Neurocomputing*, 655:131359.

Yuanchi Ma, Hui He, and Zhendong Niu. 2023. BDC: using BERT and deep clustering to improve chinese proper noun recognition. In *SEKE*, pages 57–62. KSI Research Inc.

Yuanchi Ma, Hui He, Gang Zhang, and Zhendong Niu. 2026. Multimodal nonparametric clustering via monte carlo method and fusion embedding generated by variational autoencoder. *Information Fusion*, 126:103612.

OpenAI. 2023. GPT-4 technical report. *CoRR*, abs/2303.08774.

Tianxiao Shen, Victor Quach, Regina Barzilay, and Tommi S. Jaakkola. 2020. Blank language models. In *EMNLP (1)*, pages 5186–5198. Association for Computational Linguistics.

Kaize Shi, Xueyao Sun, Qing Li, and Guandong Xu. 2024. Compressing long context for enhancing RAG with amr-based concept distillation. *CoRR*, abs/2405.03085.

Juntong Song, Xingguang Wang, Juno Zhu, Yuanhao Wu, Xuxin Cheng, Randy Zhong, and Cheng Niu. 2024. RAG-HAT: A hallucination-aware tuning pipeline for LLM in retrieval-augmented generation. In *EMNLP (Industry Track)*, pages 1548–1558. Association for Computational Linguistics.

Peiqi Sui, Kelvin K. Wong, Xiaohui Yu, John Volpi, and Stephen T. C. Wong. 2023. Storyline-centric detection of aphasia and dysarthria in stroke patient transcripts. In *ClinicalNLP@ACL*, pages 422–432. Association for Computational Linguistics.

Xiaofei Sun, Zijun Sun, Yuxian Meng, Jiwei Li, and Chun Fan. 2022. Summarize, outline, and elaborate: Long-text generation via hierarchical supervision from extractive summaries. In *COLING*, pages

6392–6402. International Committee on Computational Linguistics.

Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, Dan Bikel, Lukas Blecher, Cristian Canton-Ferrer, Moya Chen, Guillem Cucurull, David Esiobu, Jude Fernandes, Jeremy Fu, Wenyin Fu, Brian Fuller, Cynthia Gao, Vedanuj Goswami, Naman Goyal, Anthony Hartshorn, Saghar Hosseini, Rui Hou, Hakan Inan, Marcin Kardas, Viktor Kerkez, Madian Khabsa, Isabel Kloumann, Artem Korenev, Punit Singh Koura, Marie-Anne Lachaux, Thibaut Lavril, Jenya Lee, Diana Liskovich, Yinghai Lu, Yuning Mao, Xavier Martinet, Todor Mihaylov, Pushkar Mishra, Igor Molybog, Yixin Nie, Andrew Poulton, Jeremy Reizenstein, Rashi Rungta, Kalyan Saladi, Alan Schelten, Ruan Silva, Eric Michael Smith, Ranjan Subramanian, Xiaoqing Ellen Tan, Binh Tang, Ross Taylor, Adina Williams, Jian Xiang Kuan, Puxin Xu, Zheng Yan, Iliyan Zarov, Yuchen Zhang, Angela Fan, Melanie Kambadur, Sharan Narang, Aurélien Rodriguez, Robert Stojnic, Sergey Edunov, and Thomas Scialom. 2023. Llama 2: Open foundation and fine-tuned chat models. *CoRR*, abs/2307.09288.

Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2017. Graph attention networks. *CoRR*, abs/1710.10903.

Liang Wang, Sujian Li, Xinyan Xiao, and Yajuan Lyu. 2016. Topic segmentation of web documents with automatic cue phrase identification and BLSTM-CNN. In *NLPCC/ICCPOL*, volume 10102 of *Lecture Notes in Computer Science*, pages 177–188. Springer.

Sam Wiseman, Stuart M. Shieber, and Alexander M. Rush. 2018. Learning neural templates for text generation. In *EMNLP*, pages 3174–3187. Association for Computational Linguistics.

Linzi Xing, Brad Hackinen, Giuseppe Carenini, and Francesco Trebbi. 2020. Improving context modeling in neural topic segmentation. In *AACL/IJCNLP*, pages 626–636. Association for Computational Linguistics.

Dingyi Yang, Chunru Zhan, Ziheng Wang, Biao Wang, Tiezheng Ge, Bo Zheng, and Qin Jin. 2024a. Synchronized video storytelling: Generating video narrations with structured storyline. In *ACL (1)*, pages 9479–9493. Association for Computational Linguistics.

Dingyi Yang, Chunru Zhan, Ziheng Wang, Biao Wang, Tiezheng Ge, Bo Zheng, and Qin Jin. 2024b. Synchronized video storytelling: Generating video narrations with structured storyline. *CoRR*, abs/2405.14040.

Biao Zhang, Zhongtao Liu, Colin Cherry, and Orhan Firat. 2024. When scaling meets LLM finetuning: The effect of data, model and finetuning method. In *ICLR*. OpenReview.net.

Ruqing Zhang, Jiafeng Guo, Yixing Fan, Yanyan Lan, and Xueqi Cheng. 2019. Outline generation: Understanding the inherent content structure of documents. In *SIGIR*, pages 745–754. ACM.

Deyu Zhou, Haiyang Xu, and Yulan He. 2015a. An unsupervised bayesian modelling approach for storyline detection on news articles. In *EMNLP*, pages 1943–1948. The Association for Computational Linguistics.

Deyu Zhou, Haiyang Xu, and Yulan He. 2015b. An unsupervised bayesian modelling approach for storyline detection on news articles. In *EMNLP*, pages 1943–1948. The Association for Computational Linguistics.

A Appendix A

Data Collection. We select high-quality long-form works that are widely read on the Internet and collect the original text of these works. As shown in Table 4. First, we divide different works into five categories according to their content and style. Then we first use the powerful LLM Deepseek-V3 to pre-generate the plot boundaries of the article. Then use expert knowledge and relevant forum content to correct the incorrect plot boundaries of the article. Satisfactory results have been obtained by using the above automatic pretreatment process. We then recruited a web-contracted writer with more than five thousand hours of writing experience and two experts with tens of thousands of hours of reading experience to review the narrative, check plot boundaries, and correct any remaining errors. In this way, we built our proposed dataset, named NovOutlineSet.

Interestingly, our analysis reveals significant genre-dependent variation in narrative structure, with swordsman and fantasy novels exhibiting extended chapter sequences compared to the more concise formats of biographical works. This structural diversity necessitates genre-specific adaptation of our retrieval threshold s_t through parameter β. Through systematic evaluation, we have established optimal β values for major literary genres, as detailed in Table 4, enabling robust performance across diverse narrative formats.

Dataset Statistics. The statistical graph related to the dataset is shown in Fig. 7 and 8. Among them, we include my category of texts, a total of 28. More than 1,000 chapters account for 50% of the total.

Dataset Contribution. To the best of our knowledge, NovOutlineSet is the first Chinese-based long-text generation outline dataset. We are the

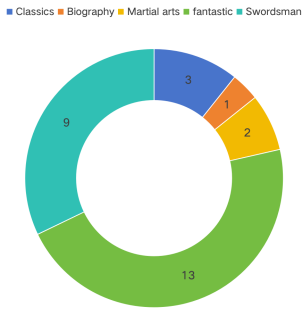


Figure 7: Text category ratio graph

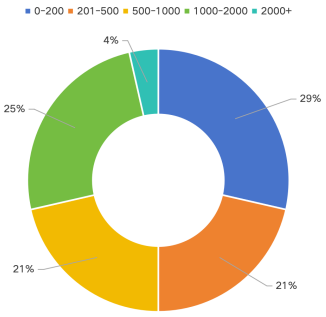


Figure 8: Distribution of the number of chapters in each text

first dataset to annotate a structured storyline that can be used to support more generation tasks. Our dataset also contains much longer individual texts. The longest text is 23.1MB and reaches 2271 chapters.

Long Text Visualization. In order to visually show the state of our long text during the calculation process, we visualized the *Fights Break Sphere*. As shown in Fig. 9. Each point represents the visual coordinates of a chapter-level graph. Then, according to the detection operator algorithm mentioned in section 3.4, we check the plot boundaries of the whole book and get the three plot boundaries shown in the figure.

Comparison of Detailed Outlines. Using the example of a famous fantasy novel, *Fights Break Sphere*, we selected the three best-performing models and presented the Outlines they generated. They are GPT 4.0, Deepseek-V3. and our method, LeStrTP. We begin by providing an outline of the standard. It not only embodies the importance of plot boundary division in the task of outline extraction and generation, but also shows the superiority of our method. In order to control the length of the appendix, we try to reduce the outline to less than 500 words. The actual generation outline does not

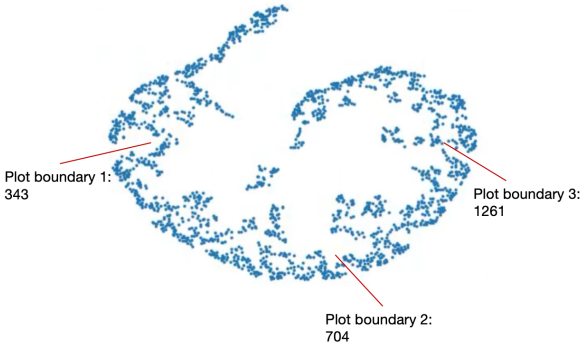


Figure 9: Visualization of *Fights Break Sphere*

limit the number of words.

We show the standard outline generated by LeStrTP in Fig. 10 and 11, and provide a simplified version of the outline and two baseline generation results in the Fig. 12 to visually show the effects of each model.

Due to the high accuracy of Deepseek, we checked its official training data and ruled out the possibility of data leakage. Such as the time capsule mechanism: building a data pipeline based on the publication dates of novel chapters, the training set is limited to chapters published between 2013 and 2019 (1-800 chapters). The content of these editions is based on the original text, and there are no artificial additions such as plot division. According to the knowledge distillation audit provided by Deepseek-R1, by comparing the original word vector with the vector after leakage protection treatment, it is further proved that the training process does not remember the original text details, and there is no plot division boundary.

Type (Chinese/English)	β value	Book's title (Chinese/English)	Plot boundary (Number of Chapters)
名著/Classics	$\beta = 0.3$	悲惨世界/Les Misérables	[9,17,25,52]
		基督山伯爵/The Count of Monte Cristo	[20,45,92,105]
		堂吉柯德/Don Quixote	[15,52,99]
人物传记/Biography	$\beta = 0.3$	慈禧全传/The Complete Biography of Cixi	[23,49,81,100]
武侠/Wuxia	$\beta = 0.4$	天龙八部/Demi-Gods and Semi-Devils	[15,28,31,40]
		鹿鼎记/The Deer and the Cauldron	[9,19,29,39]
玄幻/Fantastic	$\beta = 0.6$	最强弃少/The Abandoned Youngster Is Now the Mightiest Master	[97,190,348,575,693,956,1078, 1184,1339,1420,1520,1616,1715, 1803,1892,1990,2086,2200]
		庆余年/Joy of Life	[38,103,152,251,405,583,830]
		回到明朝当王爷/Royal Highness	[15,35,53,152,290,450]]
		斗破苍穹/Fights Break Sphere	[354,704,1261]
		佣兵天下/Sphira	[20,130,265,362,444]
		小兵传奇/The Legend of Soldier	[8,19,72,275]
		魔兽剑圣异界纵横/Swords of Warcraft	[22,52,117,301,814]
		狩魔手记/The Devil's Hand	[17,36,59,293,608]
		星峰传说/Star peak legend	[44,111,398]
		朱雀记/Suzaku	[10,138,320]
		天问/Celestial Question	[67,126,275,373]
		天逆/Celestial Inversion	[13,89,150]
		鬼吹灯3/Candle in the Tomb 3	[6,21,32,46]
仙侠/Immortal Heroes	$\beta = 0.7$	九界仙尊/Nine world fairy Buddha	[45,88,142,191,369,1695]
		莽荒纪/The Legend of JADE SWORD	[98,283,484,763,1031,1273,1390]
		斗战狂潮/Battle Frenzy	[33,74,212,509,615,1222]
		我欲封天/I Shall Seal the Heavens	[41,94,313,628,1068]
		一念永恒/The thought of eternity	[34,72,121,155,505,759,1074,1315]
		圣堂/Church	[28,60,86,159,998]
		星辰变/Legend of Immortal	[21,41,359,690]
		聊斋大圣人/Great Sage of Liaozhai	[22,81,223,609,846]
		佛本是道/Buddha is the Tao	[12,24,45,294,444]

Table 4: Details of NovOutlineSet dataset

标准大纲

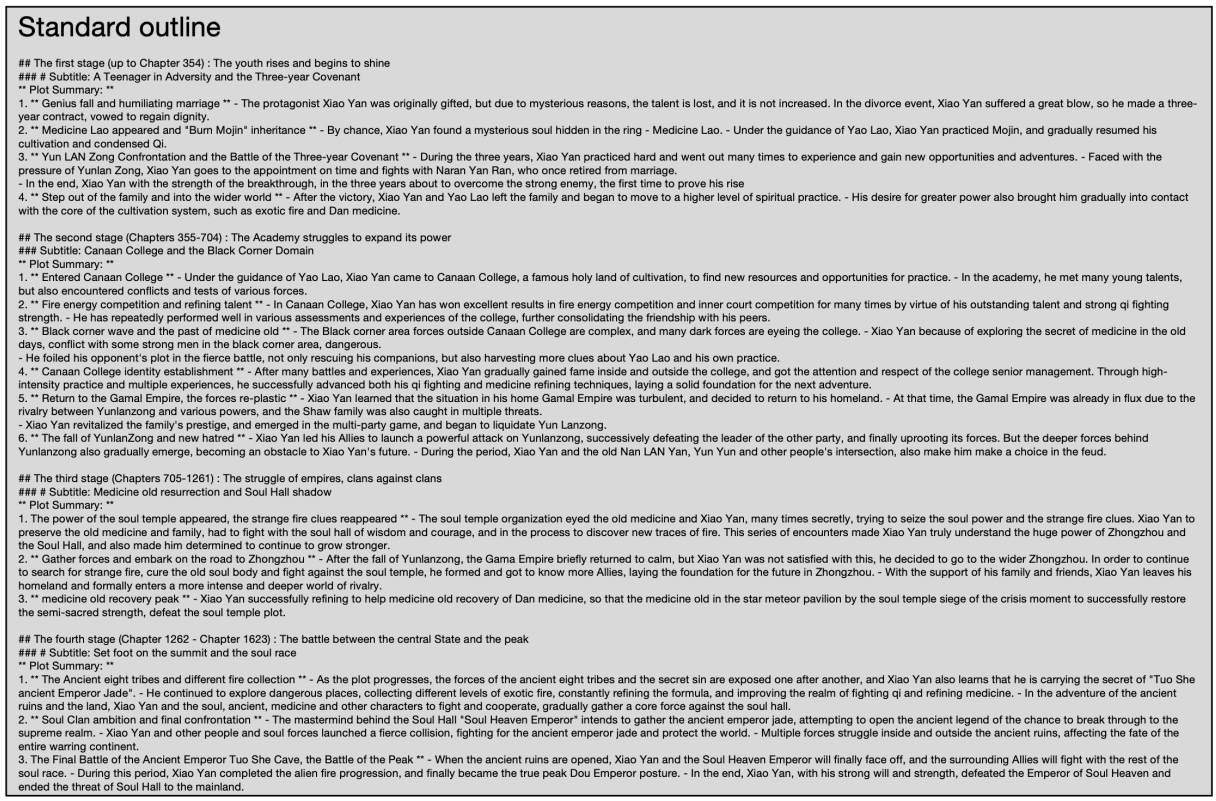
第一阶段（至第 354 章）：少年崛起，初露锋芒
小标题：逆境中的少年与三年之约
情节概述：
1. **天才陨落与耻辱婚约** - 主角萧炎原本天赋异禀，却因神秘原因天赋尽失，修为不增反降。 - 在退婚事件中，萧炎受到巨大打击，于是立下三年之约，誓要夺回尊严。
2. **药老现身与《焚诀》传承** - 机缘巧合下，萧炎发现戒指中隐藏着神秘灵魂——药老。 - 在药老的指导下，萧炎修炼《焚诀》，并逐渐恢复修为、凝聚斗气。
3. **云岚宗对峙与三年之约之战** - 三年间，萧炎刻苦修行，多次外出历练，获得新的机遇与奇遇。 - 面对云岚宗的压力，萧炎按时赴约，与曾经退婚的纳兰嫣然激战。
- 最终，萧炎凭借突破后的实力，在三年之约中力克强敌，初次证明了自己的崛起。
4. **踏出家族，走向更广阔的世界** - 胜利之后，萧炎与药老离开家族，开始走向更高层次的修行世界。 - 他对更强力量的渴望，也让他逐渐接触到异火、丹药等修炼体系的核心所在。

第二阶段（第 355 章—第 704 章）：学院争锋，扩展势力
小标题：迦南学院与黑角域
情节概述：
1. **进入迦南学院** - 萧炎在药老的指引下，来到著名修炼圣地“迦南学院”，寻找新的修行资源和机会。 - 在学院中，他结识了众多青年才俊，也遭遇了各色势力的冲突与考验。
2. **火能争夺与炼药才华** - 在迦南学院，萧炎凭借自己过人的炼药天赋与强悍斗气实力，多次在火能争夺战、内院比赛中夺得佳绩。 - 他在学院的各项考核和历练中屡屡表现出色，进一步巩固了与同伴们的友谊。
3. **黑角域风波与药老的过往** - 迦南学院外的黑角域势力复杂，许多暗势力对学院虎视眈眈。 - 萧炎因探索药老往昔的秘密，与黑角域一些强者产生冲突，险象环生。
- 他在激战中挫败了对手的阴谋，不仅救出同伴，也收获了更多关于药老和自身修行的线索。
4. **迦南学院身份地位确立** - 经过多次战斗与历练，萧炎在学院内外逐渐声名鹊起，得到学院高层的重视与器重。 他通过高强度的修行和多方历练，成功让自己的斗气和炼药术双双进阶，为下一步的冒险夯实基础。
5. **回归加玛帝国，势力再塑** - 萧炎得知家乡加玛帝国局势动荡，便决定回归故土。 - 彼时的加玛帝国已因云岚宗与各大势力的角逐而暗流涌动，萧家也陷入多重威胁中。
- 萧炎重振家族威名，并在多方博弈中崭露头角，着手对云岚宗进行清算。
6. **云岚宗的覆灭与新的仇怨** - 萧炎率领盟友对云岚宗展开强有力打击，先后击败对方头目，最终将其势力连根拔起。 - 但云岚宗背后的更深层势力也逐渐显现，成为萧炎未来的阻碍。 - 期间，萧炎与故人纳兰嫣然、云韵等人的交集，也令他在恩怨纠葛中作出抉择。

第三阶段（第 705 章—第 1261 章）：帝国之争，宗门对立
小标题：药老复生与魂殿阴影
情节概述：
1. **魂殿势力初现，异火线索再起** - 魂殿组织对药老以及萧炎虎视眈眈，多次暗中出手，企图夺取灵魂力量与异火线索。 - 萧炎为保全药老与家族，不得不与魂殿斗智斗勇，并在此过程中发掘到新的异火踪迹。 - 这一系列的交锋，使萧炎真正见识到中州与魂殿的庞大威势，也令他下定决心持续变强。
2. **凝聚势力，踏上中州之路** - 云岚宗覆灭后，加玛帝国短暂恢复平静，但萧炎已不满足于此，他决定前往更广阔的中州。 为了继续搜寻异火、救治药老魂体并对抗魂殿，他组建并结识了更多盟友，为将来在中州的闯荡奠定人脉基础。 - 在家族与朋友的支持下，萧炎离开故土，正式迈入更激烈、更深层次的斗气世界。
3. **药老恢复巅峰** - 萧炎成功炼制出帮助药老恢复的丹药，使药老在星陨阁被魂殿围攻的危难时刻成功恢复半圣实力，挫败魂殿阴谋。

第四阶段（第 1262 章—第 1623 章）：中州争锋，巅峰对决
小标题：踏上巅峰与魂族决战
情节概述：
1. **远古八族与异火收集** - 随着剧情推进，远古八族的势力、秘辛相继曝光，萧炎也得知自己身怀“陀舍古帝玉”的秘密。 - 他继续探索各处险地，收集不同等级的异火，不断淬炼焚诀，提升斗气与炼药双修的境界。 - 在远古遗迹与族地的冒险中，萧炎与魂族、古族、药族等多方人物交锋与合作，逐渐汇集一股抗衡魂殿的中坚力量。
2. **魂族野心与最终交锋** - 魂殿幕后主使“魂天帝”意欲集齐古帝玉，企图开启远古传说中的机缘，突破至无上境界。 - 萧炎等人与魂族势力展开激烈的碰撞，为争夺古帝玉与守护天下而战。 - 多方势力在远古遗迹内外鏖战，牵动整个斗气大陆的命运。
3. **决战陀舍古帝洞府，巅峰之战** - 当远古遗迹开启，萧炎与魂天帝终迎正面对决，周围盟友则与魂族余众激战不休。 - 期间，萧炎完成异火进阶，终成真正的巅峰斗帝之姿。 - 最终，萧炎凭借强大意志与力量，击败魂天帝，结束魂殿对大陆的威胁。

Figure 10: Standard syllabus for *Fights Break Sphere*



Standard outline

The first stage (up to Chapter 354) : The youth rises and begins to shine
Subtitle: A Teenager in Adversity and the Three-year Covenant
** Plot Summary: **
1. ** Genius fall and humiliating marriage ** - The protagonist Xiao Yan was originally gifted, but due to mysterious reasons, the talent is lost, and it is not increased. In the divorce event, Xiao Yan suffered a great blow, so he made a three-year contract, vowed to regain dignity.
2. ** Medicine Lao appeared and "Burn Mojin" inheritance ** - By chance, Xiao Yan found a mysterious soul hidden in the ring - Medicine Lao. - Under the guidance of Yao Lao, Xiao Yan practiced Mojin, and gradually resumed his cultivation and condensed Qi.
3. ** Yun LAN Zong Confrontation and the Battle of the Three-year Covenant ** - During the three years, Xiao Yan practiced hard and went out many times to experience and gain new opportunities and adventures. - Faced with the pressure of Yunlan Zong, Xiao Yan goes to the appointment on time and fights with Naran Yan Ran, who once retired from marriage.
- In the end, Xiao Yan with the strength of the breakthrough, in the three years about to overcome the strong enemy, the first time to prove his rise
4. ** Step out of the family and into the wider world ** - After the victory, Xiao Yan and Yao Lao left the family and began to move to a higher level of spiritual practice. - His desire for greater power also brought him gradually into contact with the core of the cultivation system, such as exotic fire and Dan medicine.

The second stage (Chapters 355-704) : The Academy struggles to expand its power
Subtitle: Canaan College and the Black Corner Domain
** Plot Summary: **
1. ** Entered Canaan College ** - Under the guidance of Yao Lao, Xiao Yan came to Canaan College, a famous holy land of cultivation, to find new resources and opportunities for practice. - In the academy, he met many young talents, but also encountered conflicts and tests of various forces.
2. ** Fire energy competition and refining talent ** - In Canaan College, Xiao Yan has won excellent results in fire energy competition and inner court competition for many times by virtue of his outstanding talent and strong qi fighting strength. - He has repeatedly performed well in various assessments and experiences of the college, further consolidating the friendship with his peers.
3. ** Black corner wave and the past of medicine old ** - The Black corner area forces outside Canaan College are complex, and many dark forces are eyeing the college. - Xiao Yan because of exploring the secret of medicine in the old days, conflict with some strong men in the black corner area, dangerous.
- He foiled his opponent's plot in the fierce battle, not only rescuing his companions, but also harvesting more clues about Yao Lao and his own practice.
4. ** Canaan College identity establishment ** - After many battles and experiences, Xiao Yan gradually gained fame inside and outside the college, and got the attention and respect of the college senior management. Through high-intensity practice and multiple experiences, he successfully advanced both his qi fighting and medicine refining techniques, laying a solid foundation for the next adventure.
5. ** Return to the Gamal Empire, the forces re-plastic ** - Xiao Yan learned that the situation in his home Gamal Empire was turbulent, and decided to return to his homeland. - At that time, the Gamal Empire was already in flux due to the rivalry between Yunlanzong and various powers, and the Shaw family was also caught in multiple threats.
- Xiao Yan revitalized the family's prestige, and emerged in the multi-party game, and began to liquidate Yun Lanzong.
6. ** The fall of YunlanZong and new hatred ** - Xiao Yan led his Allies to launch a powerful attack on Yunlanzong, successively defeating the leader of the other party, and finally uprooting its forces. But the deeper forces behind Yunlanzong also gradually emerge, becoming an obstacle to Xiao Yan's future. - During the period, Xiao Yan and the old Nan LAN Yan, Yun Yun and other people's intersection, also make him make a choice in the feud.

The third stage (Chapters 705-1261) : The struggle of empires, clans against clans
Subtitle: Medicine old resurrection and Soul Hall shadow
** Plot Summary: **
1. The power of the soul temple appeared, the strange fire clues reappeared ** - The soul temple organization eyed the old medicine and Xiao Yan, many times secretly, trying to seize the soul power and the strange fire clues. Xiao Yan to preserve the old medicine and family, had to fight with the soul hall of wisdom and courage, and in the process to discover new traces of fire. This series of encounters made Xiao Yan truly understand the huge power of Zhongzhou and the Soul Hall, and also made him determined to continue to grow stronger.
2. ** Gather forces and embark on the road to Zhongzhou ** - After the fall of Yunlanzong, the Gama Empire briefly returned to calm, but Xiao Yan was not satisfied with this, he decided to go to the wider Zhongzhou. In order to continue to search for strange fire, cure the old soul body and fight against the soul temple, he formed and got to know more Allies, laying the foundation for the future in Zhongzhou. - With the support of his family and friends, Xiao Yan leaves his homeland and formally enters a more intense and deeper world of rivalry.
3. ** medicine old recovery peak ** - Xiao Yan successfully refining to help medicine old recovery of Dan medicine, so that the medicine old in the star meteor pavilion by the soul temple siege of the crisis moment to successfully restore the semi-sacred strength, defeat the soul temple plot.

The fourth stage (Chapter 1262 - Chapter 1623) : The battle between the central State and the peak
Subtitle: Set foot on the summit and the soul race
** Plot Summary: **
1. ** The Ancient eight tribes and different fire collection ** - As the plot progresses, the forces of the ancient eight tribes and the secret sin are exposed one after another, and Xiao Yan also learns that he is carrying the secret of "Tuo She ancient Emperor Jade". - He continued to explore dangerous places, collecting different levels of exotic fire, constantly refining the formula, and improving the realm of fighting qi and refining medicine. - In the adventure of the ancient ruins and the land, Xiao Yan and the soul, ancient, medicine and other characters to fight and cooperate, gradually gather a core force against the soul hall.
2. ** Soul Clan ambition and final confrontation ** - The mastermind behind the Soul Hall "Soul Heaven Emperor" intends to gather the ancient emperor jade, attempting to open the ancient legend of the chance to break through to the supreme realm. - Xiao Yan and other people and soul forces launched a fierce collision, fighting for the ancient emperor jade and protect the world. - Multiple forces struggle inside and outside the ancient ruins, affecting the fate of the entire warring continent.
3. The Final Battle of the Ancient Emperor Tuo She Cave, the Battle of the Peak ** - When the ancient ruins are opened, Xiao Yan and the Soul Heaven Emperor will finally face off, and the surrounding Allies will fight with the rest of the soul race. - During this period, Xiao Yan completed the alien fire progression, and finally became the true peak Dou Emperor posture. - In the end, Xiao Yan, with his strong will and strength, defeated the Emperor of Soul Heaven and ended the threat of Soul Hall to the mainland.

Figure 11: The English version of the *Fights Break Sphere* standard outline

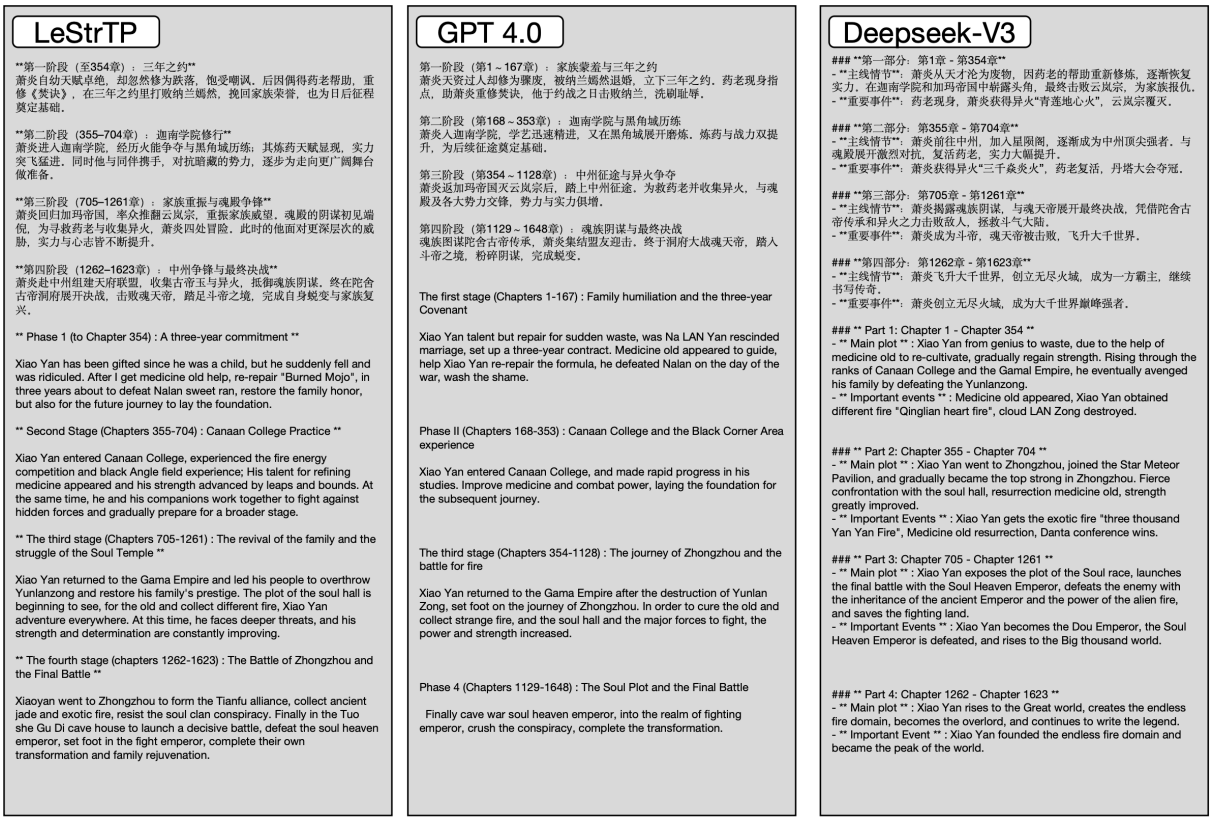


Figure 12: Comparison of outline generated by three methods.