jcls. Journal of Computational Literary Studies

OPEN ACCESS

conference version

Verse within Prose

Annotating and Classifying Narrative Functions of Embedded Poems in Chinese Qing (1644-1912) Vernacular Fiction

Rongqian Ma[1]
Keli Du[2]
Yiwen Zheng[3]

1. Luddy School of Informatics, Computing, and Information, Indiana University Bloomington, Bloomington, IN, USA.
2. Trier Center for Digital Humanities, University of Trier, Trier, Germany.
3. Department of East Asian Languages and Cultures, Indiana University Bloomington, Bloomington, IN, USA.

Citation
Rongqian Ma, Keli Du, and Yiwen Zheng (2025). "Verse within Prose. Annotating and Classifying Narrative Functions of Embedded Poems in Chinese Qing (1644-1912) Vernacular Fiction". In: *CCLS2025 Conference Preprints* 4 (1). 10.26083/tuprints-00030147

Date published 2025-06-17
Date accepted 2025-04-17
Date received 2025-02-06

Keywords
Chinese poetry, Qing fiction, Narrative function, Large language models, Annotation

Note
This paper has been submitted to the conference track of JCLS. It has been peer reviewed and accepted for presentation and discussion at the 4th Annual Conference of Computational Literary Studies at Krakow, Poland, in July 2025. Please check jcls.io for the final journal version.

Abstract. What narrative functions do poems serve when interwoven with vernacular prose? This article takes what has often been labeled as "embedded poems" or "parasitic poems" in late imperial Chinese fiction as the primary subject of study. We examine the narrative roles of these poems within a selected corpus of Qing dynasty fiction, specifically investigating if an approach that combines human annotation with large language models can aptly capture and automatically classify their narrative functions. Through two rounds of iterative annotation and large language model testing, we demonstrate both the potential and limitations of this approach. As one of the few studies that applies large language models to Chinese literary research, our work lays the groundwork for future large-scale investigations into the dynamics between verse and prose in classical Chinese literature, incorporating both canonical works and beyond.

1. Introduction

This article takes what has often been labeled as "embedded (chanru 羼入)" poems or "parasitic (jisheng 寄生)" poems in late imperial Chinese fiction as primary subjects of study. The hybrid genre of prose incorporating verse has been widely practiced ever since the medieval period, especially during the late imperial times—the Ming (1368-1644) and Qing (1644-1912) Dynasties—an age where classical and vernacular Chinese fiction reached its peak. Scholarly perspectives on the function and significance of embedded verse in prose have long been divided—a debate that continues to this day. For example, as early as the Song Dynasty (960-1279), the literatus Luo Ye 羅燁, in his *The Drunken Man's Talk* (*Zuiweng Tanlu* 醉翁談録), demonstrated how poetry was used by professional storytellers as a teaser or a form of enticement—an embellishment designed to attract the audience. This practice,

however, was often regarded as lowbrow or associated with petty entertainment (see Luo 1965). In comparison, Zhao Yanwei 趙彦衛, in *Notes from the Cloud-Covered Hill* (*Yunlu Manchao* 雲麓漫鈔), emphasized how great works of the Tang Dynasty were great because they "incorporated diverse genres, including history, poetry, and argumentative essays" 蓋此等文備众體，可以見史才、詩筆、議論 (see Yanwei Zhao 1966).

In modern times, one major camp readapted some earlier critiques and created the concept of "parasitic verse" as one major approach to studying the phenomenon (see Yishan Zhao 2014). They argue that verse can change or even impede the narrative flows of the text, a phenomenon that they think took stage in literary history momentarily and ultimately disappeared by giving way to the narrative. The other camp, as represented by the scholar Rao Longsun 饒龍隼, opposed the labeling of "parasitic verse" by arguing that the verse and narrative sections of the late imperial Chinese fiction are supplementary to each other in forming a holistic organism and thus should not be taken apart in analysis (see Rao 2023). Another group of scholars focuses on the cause and function of the verse-within-prose phenomenon, pivoting toward intellectual history and without taking a stance. For example, Zhang Zhejun 張哲俊 argues that poetry in verse provides authority to the text, a function that can be traced back to the classics such as *Book of Songs* (*Shijing* 詩經) (see Z. Zhang 2015). Similarly, Guo Jie 郭杰 suggests that the rise of poetry in verse in China originates from and mimics ways of history writing ever since pre-Qin times (see Guo 1995).

Elder generations of Western scholars of Chinese Studies tend to side with opinions on the lack or the diminishing narrative functions of verse in late imperial fiction. John Bishop stated that "originally such verses may have had an integral function in the story; later they served as a commentary, a verification, a means of delaying a climax, or merely as an embellishment" (see Bishop 1965, p. 241). Robert Hegel argued that the use of verse in fiction was mainly due to the writer's scholarly identity and desire for articulation (Hegel 1985). He claimed that "the pace of action is deliberately slowed in mature novels for the elite by frequent insertion of verse, usually attributable to the narrator or 'quoted' by him from earlier, characteristically anonymous, sources...Literati novelists utilized the novel form to meet specific intellectual needs: social and political commentary, philosophical exploration, self-expression, and even their own and their friends' enjoyment" (see Hegel 1985, p. 126).

However, during recent years, the narrative function of poetry in late imperial Chinese fiction has been increasingly noted and studied more in depth, especially in scholarly works that narrow in on a specific fiction work. For example, scholars have taken an interest in analyzing the narrative functions of poetry in one particular work—the Qing Dynasty fiction *Dream of the Red Chamber* (*Honglou Meng* 紅樓夢, also translated as The *Story of the Stone*). This makes sense because the work is often

considered the epitome of this hybrid genre as its narrative incorporates more than a number of subgenres of verses. Cai Yijiang 蔡义江 sees the hybrid style in *Dream of the Red Chamber* as a composite of genres that generate positive meaning, describing it as "prose equipped with a variety of genres" (wen bei zhong ti 文備众體). Cai specifically lists out five narrative functions of poems in *Dream of the Red Chamber*: 1) Social critique (借题发挥，伤时骂世), 2) Part of plot (小说的有机组成部分), 3) Reflecting social reality (时代文化精神生活的反映), 4) Character portraiture (按头制帽，诗即其人), and 5) Prediction of later plot (谶语式的表现方法) (see Cai 2007, pp. 27-39). Another literary scholar Chia-ying Yeh 葉嘉瑩 divides the use of verse in *Dream of the Red Chamber* into three types: 1) Pre-introducing characters through homophonic puns or combination of radicals in poetry; 2) Modeling of characters through articulation of their imagined voices; and 3) Conveying authorial intention and emotion through prediction of later plot (see Yeh 2004, p.58). Compared to Cai's more comprehensive list that attempts to exhaust possible scenarios from the perspective of the author, Yeh's shortlist takes into account both characters and the narrator and views the verse as conveying emotion and facilitating the plot and narrative (see Yeh 2004).

Despite extensive interpretive efforts by literary scholars, few studies have examined embedded poems collectively within the broader context of Ming-Qing fiction, especially in lesser-known or non-canonical works. This gap may, in part, stem from the methodological challenges of analyzing large-scale text corpora. In our work, we engage in the literary debate over embedded poems' functions, yet aim to do so by leveraging state-of-the-art computational methodologies.

Recent advancements in natural language processing and computational literary studies provide a promising means to revisit and analyze the "verse within prose" literary phenomenon at an unprecedented scale. For instance, a Bayesian hierarchical generalized linear model can be used to track the relations between emotions in poems and factors such as period, author profession, and rhyme. This analysis demonstrates that the connection of emotion with rhyme is as strong as that with thematic genre, while the connection with profession is as strong as that with gender (see Konle et al. 2023). Moreover, the rapid development of LLMs (both prompting and fine-tuning) has proven especially valuable for literary and poetry analysis. In some cases, LLMs can match or even surpass supervised machine learning models in distinguishing broadly recognized concepts such as science fiction, westerns, or the emotional states of characters (see Bamman et al. 2024). Additionally, researchers have tested zero-shot prompts with varying levels of information across six state-of-the-art LLMs (including GPT-4 and LLaMA 3) to classify poetic forms and their structural elements (see Walsh et al. 2024).

In the domain of Chinese literary studies, however, the application of computational approaches as well as emerging LLMs remains relatively under-explored, with a few pioneering efforts from scholars in East Asian studies. For example, Paul Vierthaler

investigated the "stylistic taxonomy" of the subtle and mixed genre of the late imperial unofficial historical narrative or quasi-history texts, by applying "statistical and linear algebraic analysis of the term frequency lists calculated from digitized transcripts" of these texts (Vierthaler 2016). Liu (Liu et al. 2018) demonstrated the potential of using digital tools to explore Chinese poetry from different aspects such as aesthetic expressions, and personal styles. Additionally, LLMs have shown remarkable capabilities in generating ancient Chinese poetry (see Huang and Shen 2025) and detecting and correcting errors in classical Chinese verse (see Yu et al. 2024).

Extending from the previous works, our article seeks to apply emerging LLMs to Chinese literary analysis, particularly exploring their potential in identifying and classifying the narrative functions of embedded poems in Ming-Qing vernacular fiction. Our approach combines human annotation with an iterative, trial-and-error method for automatic classification. Researchers have argued that literary scholars are best equipped to "explore, define, and exemplify narratological concepts" and encourage "a corpus with annotated concepts" to be created, before "any computer scientist and/or machine learning expert can work on the automatic detection of the concepts" (see Reiter et al. 2019). Following this principle, we selected a sample of Qing fiction, extracting embedded poems along with their immediate narrative contexts to construct a pilot dataset. Using this pilot dataset, we developed an iterative, collaborative annotation process to classify the narrative functions of these poems. The annotated dataset was then employed to test whether LLMs, such as ChatGPT, can automatically classify poems by their function on a larger scale. In this paper, we will describe the creation of the pilot dataset (Chapter 2), our iterative annotation process, as well as two rounds of preliminary testing and their results (Chapters 3 & 4). Our findings underscore both the potential and challenges of applying LLMs to the study of Ming-Qing fiction. Ultimately, this work opens new possibilities for computational approaches in Chinese literary scholarship.

2. Creating pilot dataset

From the Chinese Text Project[1], we obtained approximately 900 titles of Chinese Ming-Qing fiction in plain text. From this complete corpus, we randomly selected 18 novels and extracted embedded poems in them. The extraction was completed in two steps. Generally, the lines in Chinese poetry have a fixed length, typically five, seven, or eight characters per line. However, in some cases, a poem with four lines of eight characters each is written as two lines of sixteen characters. In the first step, we automatically extracted all lines of up to 20 characters, along with two sentences before and after each extracted line. This approach identified many poems but also included unrelated content such as Chinese couplets (duilian 對聯), chapter titles, or text with improperly formatted paragraph breaks. To refine

1. https://ctext.org/

the results, we processed the extracted text using ChatGPT (GPT-3.5, free version) with the following prompt: "Analyze the text below to see if it contains poems, and if so, retell all the poems you find." While most of the data used to train ChatGPT is likely from modern Chinese texts, it successfully identified embedded poems written in vernacular and classical Chinese dating back to the Qing Dynasty. This was achieved using semantic cues like "詩曰" (The poem says) and "有詩为證云" (There is a poem to prove it), as well as formal elements of poetry, such as line length and number of characters per line. As a result, ChatGPT was able to extract 360 embedded poems as our pilot dataset without requiring fine-tuning on classical Chinese texts specifically for this task. This dataset is available in the project's GitHub page: https://github.com/dkltimon/EmbeddedPoems.

3. Initial Annotation Framework and Testing

3.1 Annotation Framework and Results

To build a foundation for the analysis of poems' narrative functions, we employed a data-driven, bottom-up approach, in which we established an initial annotation framework by synthesizing insights from existing research literature with an exploratory analysis of the 360 poems in our pilot dataset (Cai 2007; Chun 2009). In this initial framework, we identified five key narrative functions: opening teaser, character portraiture, commentary, integration within a scene or plot, and concluding remarks (Table 1). We did not directly adopt the categories used in existing literary scholarship for two main reasons. First, prior discussions of the narrative functions of poetry in fiction have largely been based on close reading analyses of a few canonical works, such as *Dream of the Red Chamber*. These studies focus on a narrow group of literati authors and reflect poetry's function within specific, limited historical contexts. Second, existing scholarship often differs significantly in both terminology and interpretation, resulting in fragmented and unsystematic categorizations.

Using this initial framework, one author with academic and research backgrounds in classical Chinese literature annotated the poems, assigning each to one of these narrative functions. Figure 1 demonstrates the annotation results. As shown in the graph, "part of a scene or a plot" is the most common category in our dataset, with approximately 140 poems. This is followed by "commentary," which includes around 100 poems. "Character portraiture type" has a moderate presence with about 60 poems, while "opening teaser" and "concluding remarks" are less frequent, with roughly 40 and 20 poems, respectively. This distribution graph shows that poems serving as part of a scene or plot and those providing commentary are the most prevalent, whereas opening teasers and concluding remarks are comparatively rare.

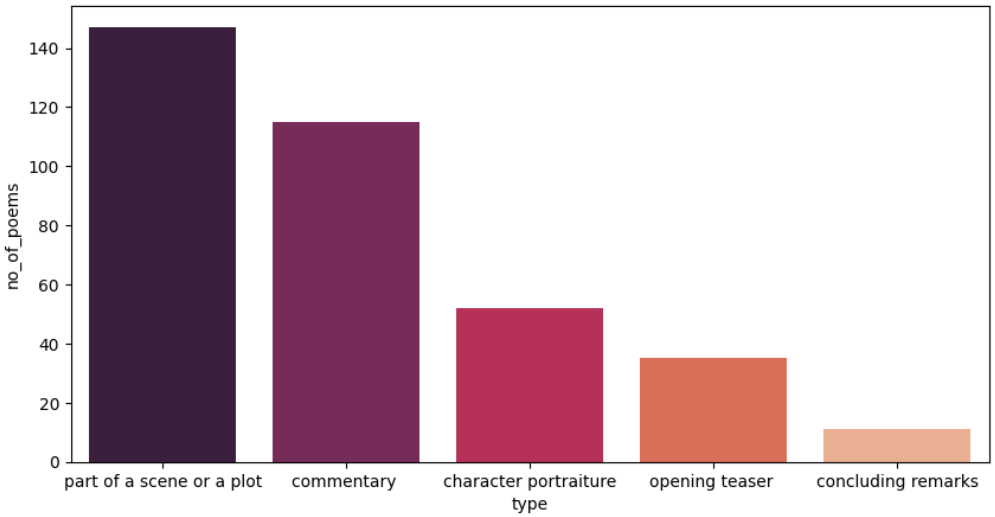

Figure 1: Distribution of the five identified narrative functions of embedded poems in Qing vernacular fiction.

Table 1: Initial annotation framework for the narrative functions of the embedded poems in Qing vernacular fiction.

Function	Description
Opening teaser	The poem is placed at the beginning of a chapter or the entire fiction. The poem may summarize the main message or idea of the fiction and serves as a teaser that grabs the readers' attention. A poem that serves this narrative function is usually composed from a third-person perspective (e.g., the narrator of the fiction).
Character Portraiture	The poem of this function describes and expresses the inner world, such as the emotions, thoughts, and feelings, of characters in the fiction, and is an important literary vehicle that shapes the personality of a character. The poem can be composed from both the character's first-person perspective or a third-person perspective (e.g., the narrator or other characters in the fiction). The poem's position is flexible and often interwoven with the vernacular narrative in the fiction.
Commentary	The poem amplifies and develops a plot from a third-person commentator's perspective. The commentator is usually the narrator and occasionally can be other characters in the fiction. Content-wise, such poems make a comment or critique of a story, a character, or a scene in the fiction. It also serves the purpose of educating the readers on moral lessons in Qing vernacular fiction.
Part of a scene or a plot	Poems of this function are a closely integrated component of a plot in fiction. The poems are usually composed by the characters involved in the plot. Although the specific content of the poems varies drastically depending on the given plot where they appear, overall, these poems facilitate the development of the plot and enrich the plot.
Concluding Remarks	The poem that serves this function is placed towards the end of a chapter or the entire fiction. Such poems usually summarize the entire story in the fiction or reiterate, emphasize, and amplify the main messages of the work. Poems of this function are usually composed from a third-person perspective (e.g., the narrator of the fiction).

3.2 First-round Classification Testing

After annotating the poems, we investigated whether LLMs can automatically distinguish their narrative functions. Specifically, we examined whether these

functions can be identified solely based on the poems' content, independent of their broader context. While this assumption is likely to be incorrect, our goal was to empirically test and potentially falsify it. To this end, we conducted an automatic classification of the poems without incorporating contextual information. Given the limited dataset of 360 poems and the imbalance across narrative function classes, the dataset was insufficient for fine-tuning and evaluating pre-trained models. Instead, we employed Zero-Shot classification and tested the following four models:

- **Erlangshen-RoBERTa-110M-NLI**: A fine-tuned version of the Chinese RoBERTa on several NLI datasets (J. Zhang et al. 2022).[2]
- **Bart-large-mnli**: The BART-large model (Lewis et al. 2019) after being trained on the MultiNLI dataset (Williams et al. 2018).[3]
- **XLM-ROBERTA-BASE-XNLI-ZH**: A fine-tuned version of the XLM-RoBERTa-base model (Conneau et al. 2020) using data in Chinese.[4]
- **ChatGPT-3.5 (free version)**: The English translation of the prompt used: "Please tell me which of the following categories this poem can be classified in. Please note that you can only assign the poem to one category. The five categories are: opening teaser, character portraiture, commentary, integration within a scene or plot, and concluding remarks."

As shown in Table 2, the classification performance was generally low, with most results aligning closely with the randomized baseline of 0.2. The best results were obtained using the XLM-Roberta-based Model, yielding an accuracy of 0.38 and an F1-score of 0.20. ChatGPT achieved the same F1-score, but lower accuracy.

Table 2: Zero-Shot classification results using four different models.

	Erlangshen-Roberta-110M-NLI	bart-large-mnli	XLM-ROBERTA-BASE-XNLI-ZH	ChatGPT-3.5
Accuracy / F1-score	0.16 / 0.08	0.21 / 0.13	0.38 / 0.20	0.27 / 0.20

To gain a deeper understanding of the classification results, we analyzed the confusion matrices for XLM-ROBERTA-BASE-XNLI-ZH and ChatGPT, as these two models achieved the highest F1-scores. As shown in Figure 2, the XLM-RoBERTa-based model exhibited a strong bias toward classifying poems as "part of a scene or plot," with approximately 60% of the poems assigned to this category. Since this category contains the largest number of poems in our dataset, the model's overall accuracy was the highest. However, this pattern suggests that the model

2. https://huggingface.co/IDEA-CCNL/Erlangshen-Roberta-110M-NLI
3. https://huggingface.co/facebook/bart-large-mnli
4. https://huggingface.co/morit/chinese_xlm_xnli

is overfitting to the dominant class rather than effectively distinguishing between narrative functions. In comparison, ChatGPT is less overfitted and classified more poems as "commentary" and "character portraiture" (see Figure 3).

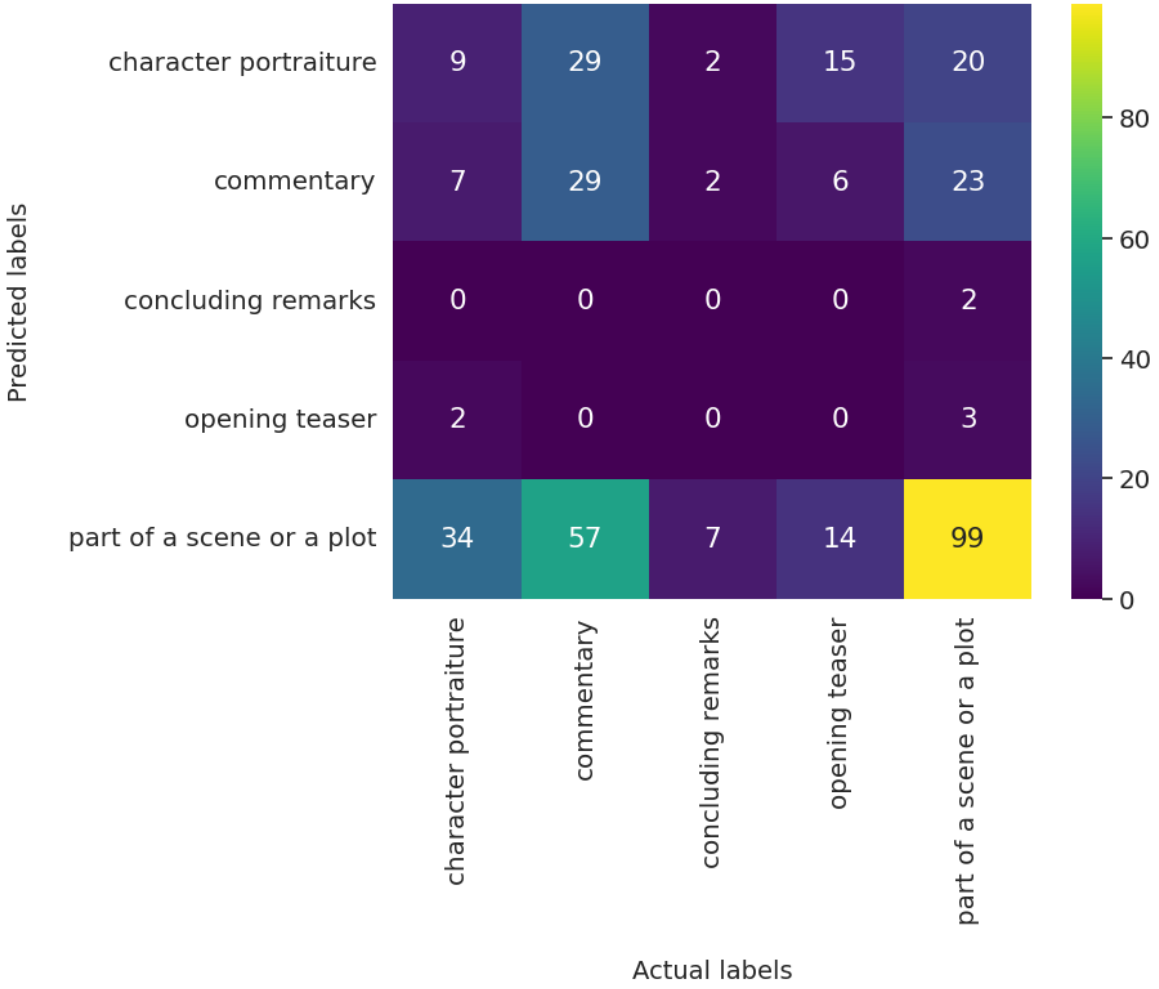

Figure 2: Confusion matrix for Zero-Shot classification using XLM-ROBERTA-BASE-XNLI-ZH.

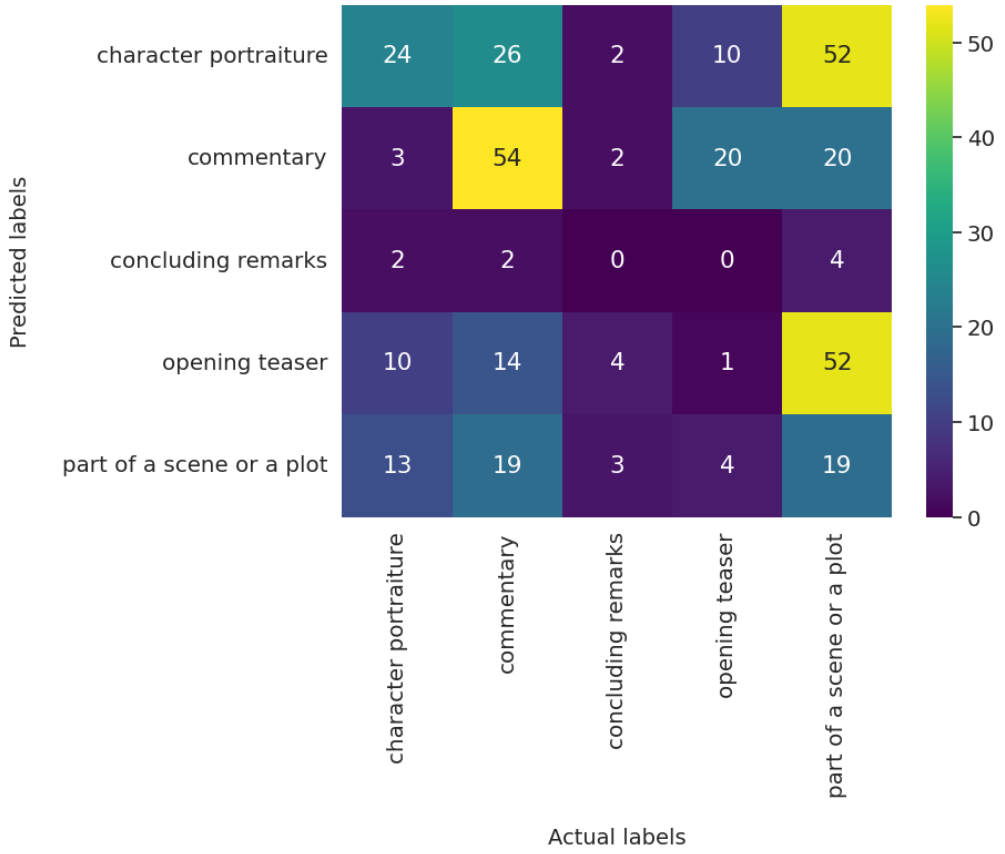

Figure 3: Confusion matrix for Zero-Shot classification using ChatGPT.

3.3 Reflections on the Framework and Results

The initial annotation and model testing revealed several limitations in using LLMs to analyze embedded poems in Qing vernacular fiction. First, regarding the annotation framework, one poem has only one assigned label in the initial framework that encompasses all dimensions of information. For instance, labels such as "opening teaser" and "concluding remarks" are primarily defined by their structural roles and positions within the text, but they may also reflect thematic or content-based features. A poem placed at the beginning or the end of a chapter, for example, may simultaneously capture readers' attention (structural role) and function as a narrative summary or commentary on a character (thematic role). Such examples demonstrate the need for a more multifaceted framework that captures the complexity of the narrative functions. Additionally, since the annotation process was conducted by one person, the subjectivity of individual interpretations may also be a factor that contributed to the models' inconsistent performance.

From a technical perspective, the models selected for the first round exhibited limitations in their linguistic and cultural adaptability when processing Chinese texts, making them less than ideal for this task. Furthermore, the methods used for prompt design and implementation may have also influenced the classification results.

Based on these observations, we refined our approach and conducted a second round of annotation and testing. In this subsequent phase, we developed a new annotation framework, revised our annotation process, and adjusted both the model selection and testing approaches to achieve better classification results.

4. Revised Framework and Second-Round Testing

4.1 Revised Annotation Framework and Results

In the revised annotation framework, each poem was analyzed across three dimensions—position, perspective, and content—to better capture the complexity of its narrative functions. The "position" dimension indicates whether a poem appears in the opening section of a chapter, within the middle of the narrative, or towards the end of a chapter. "Perspective" denotes whether a poem is written from a character's first-person or a narrator's third-person viewpoints. Finally, the "content" dimension categorizes each poem into one of four types: "character portraiture," "scene," "commentary," and "plot." Definitions for each category are included in Table 3.

Table 3: Definitions of each category for the content aspect.

Character portraiture	Describe a character, e.g., their appearance, inner feelings, emotions, and personality.
Scene	Describe natural scenery, objects, and nature.
Commentary	Offers comments and critiques of events, society, morality, characters, etc.
Plot	Conveys, narrates, and sometimes summarizes a sequence of events.

Using the revised framework, three annotators independently labeled each poem in the dataset from the three dimensions, assigning one label for each dimension[5]. The inter-annotator agreement, measured using Fleiss' kappa (Fleiss and Cohen 1973), yielded scores of 0.87 for position, 0.89 for perspective, and 0.66 for content, respectively. To address annotation discrepancies, the three annotators convened to reconcile differences, and this process resulted in a finalized ground-truth dataset for classification experiments. As illustrated in Figure 4, the dataset shows an even distribution of poems between the "narrator's" and "character's" categories for "perspective". Most poems are categorized as "middle" for "position." For content, only one poem is classified as "plot," while 36 poems are labeled as "character portraiture." The remaining two content categories are evenly represented across the dataset.

5. In the second round of annotation, the annotators identified and removed several couplets and duplicate poems from the pilot dataset. As a result, the data size has been reduced from 360 in 18 novels to 339 poems in 15 novels. The new dataset is available in the GitHub page: https://github.com/dkltimon/EmbeddedPoems

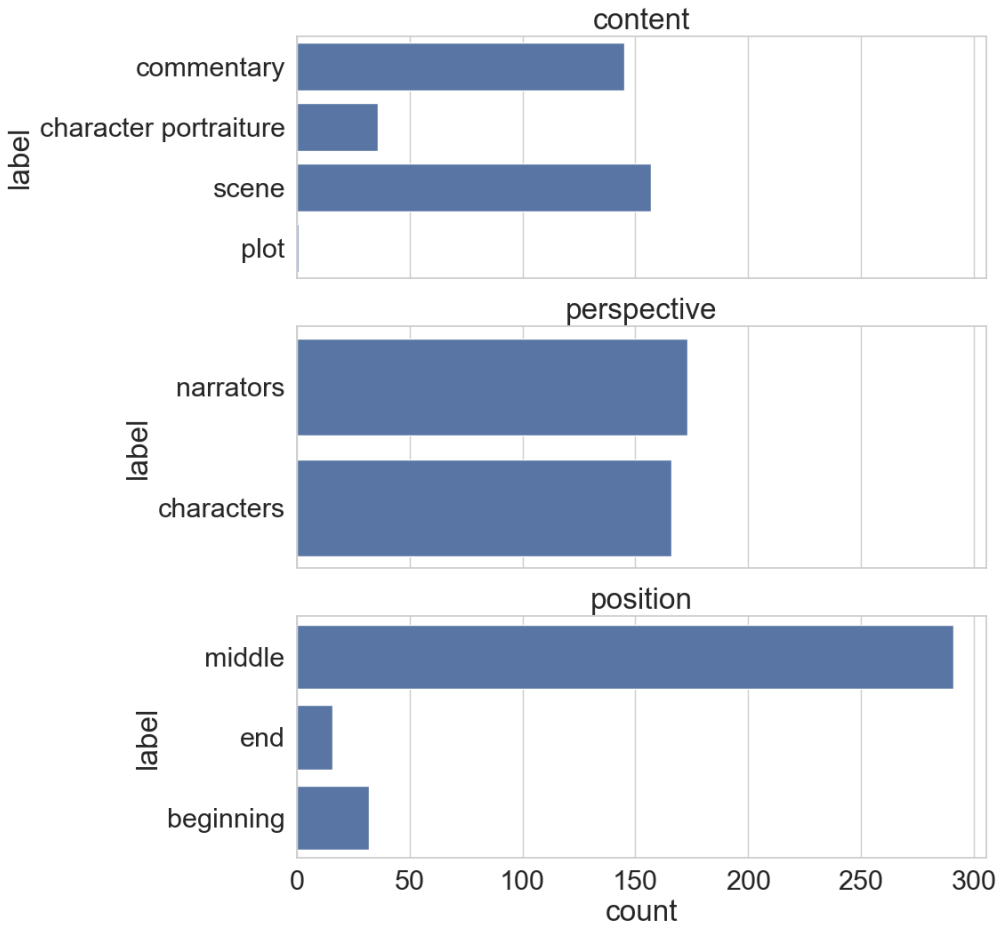

Figure 4: Annotation results for the poems using the revised framework.

4.2 Second-round Testing with the Revised Framework

In the second round of testing, we improved our approach to classifying poems in several ways. First, the results from the first round indicated that the narrative function of poems cannot be identified without considering their surrounding narrative contexts very well. To address this, we included two sentences before and after each poem as contexts for classification in this round.

Second, compared to multilingual language models fine-tuned on Chinese texts such as RoBERTa, the free version of ChatGPT 3.5 demonstrated better performance in the classification task. Also, generative models are easier to use. Therefore, we used three different generative models for this test: the paid version of ChatGPT (GPT-4) and two open-source models—Llama 3.3 (the new state-of-the-art 70B model)[6] and a Chinese Llama model, Llama-3-Chinese-8B-Instruct-v3 (Cui et al. 2023) [7].

Third, we refined the prompting methods and tested the poem content classification using three different prompts. The first was a brief prompt asking the model to

6. See: https://huggingface.co/meta-llama/Llama-3.3-70B-Instruct
7. See: https://huggingface.co/hfl/llama-3-chinese-8b-instruct-v3-gguf

select a category without further information. The second was a longer prompt that provided detailed definitions for each category before requesting a selection. The third also explained the categories but instructed the model to use a binary classification approach.[8] All three models mentioned above were tested using these prompts as follows:

- **Short prompt**: "The following Chinese text contains a poem. The beginning and end of the poem are marked with 'p_s' and 'p_e' respectively. Which of the categories "commentary," "character portraiture," "scene" and "plot" does it belong to? You do not have to explain your answer, just output your answer using the given categories. Here is the text with the poem:"
- **Long prompt**: "The following Chinese text contains a poem. The beginning and end of the poem are marked with 'p_s' and 'p_e' respectively. You have three tasks. First task, determine the narrative function of the poem. There are four options: 1. 'commentary', which offers comments and critiques of events, society, morality, characters, etc. 2. 'character portraiture', which describes a character, e.g., their appearance, inner feelings, emotions, and personality. 3. 'scene', which describes natural scenery, objects, and nature. 4. 'plot', which conveys, narrates, and sometimes summarizes a sequence of events. Second task, determine the position of a poem in a novel. The position indicates the structural role that the poem plays in the narrative of the fiction. There are three options: 1. 'beginning', means that the poem is an opening poem for a chapter. 2. 'middle', means that the poem is in the middle of a plot. 3. 'end', means the poem comes at the end of the chapter and concludes the storyline. Third task, determine whether the poem is composed or recounted from the first-person perspective of a character in the story or a third-person perspective of the author or a storyteller. If the former, please answer with 'character', if the latter, please answer with 'narrator'. For each task, you must choose one and only one option as your answer. You do not have to explain your answer, just output your answer in this format answer to the first task, answer to the second task, answer to the third task using the given option labels. Here is the text with the poem:"
- **Long prompt (binary approach)**: "The following Chinese text contains a poem. The beginning and end of the poem are marked with 'p_s' and 'p_e' respectively. Your task is to determine the narrative function of the poem, considering both the content and the context of the poem. First, determine if the poem offers comments and critiques of events, society, morality, characters, etc. If yes, answer 'commentary'. If no, determine if the poem describes natural scenery, objects, and nature. If yes, answer 'scene.' If no, determine if the poem describes characters, e.g., their appearance, inner feelings, emotions,

8. The second prompt also asked the LLMs to classify the position and perspective of the poem. This will be addressed later.

and personality. If yes, answer ‘character portraiture’. If no, determine if the poem conveys, narrates, and sometimes summarizes a sequence of events. If yes, answer ‘plot’. If still no, read the text again and choose one from the above-mentioned three options (‘commentary’, ‘character portraiture’, ‘scene’). You must choose one and only one option as your answer. You do not have to explain your answer, just output your answer using the given option labels. Here is the text with the poem:”

As shown in Table 4, the best classification results were achieved by ChatGPT using the long prompt, with an accuracy of 0.55 and an F1-score of 0.43, which is much better than the classification results obtained in the first-round testing (accuracy 0.38, F1-score 0.20). This suggests that providing detailed explanations can help models better understand both the input text and the task. However, the binary approach did not lead to further improvements but slightly reduced both accuracy and the F1-score. In comparison, the Llama models performed worse than ChatGPT. While Llama 3.3 achieved similar accuracy, its F1-scores were much lower. A detailed examination of the results revealed that Llama 3.3 classified 90.6% of the poems as “commentary,” indicating a strong overfitting issue. Surprisingly, the Chinese Llama model performed the worst, with better classification results when the short prompt was used. This model also displayed overfitting, classifying 45% and 53% of the poems as “commentary” and “plot,” respectively. By contrast, only five poems were identified as “character portraiture,” and none were classified as “scene.” To obtain more details on the classification results, we checked the confusion matrix for our best classification (Figure 5). As can be seen, ChatGPT correctly identified 81% of “character portraiture” poems and 57% of “commentary” poems. However, 54% of “scene” poems were misclassified into other categories, highlighting areas that require further refinement.

Table 4: Classification results of poems for the content aspect.

	Short prompt		**Long prompt**		**Long prompt binary**	
Model	**Accuracy**	**F1-score**	**Accuracy**	**F1-score**	**Accuracy**	**F1-score**
ChatGPT	0.41	0.33	**0.55**	**0.43**	0.53	0.39
Llama 3.3	0.40	0.15	0.51	0.33	0.53	0.29
llama3-zh-inst	0.37	0.29	0.21	0.14	0.44	0.17

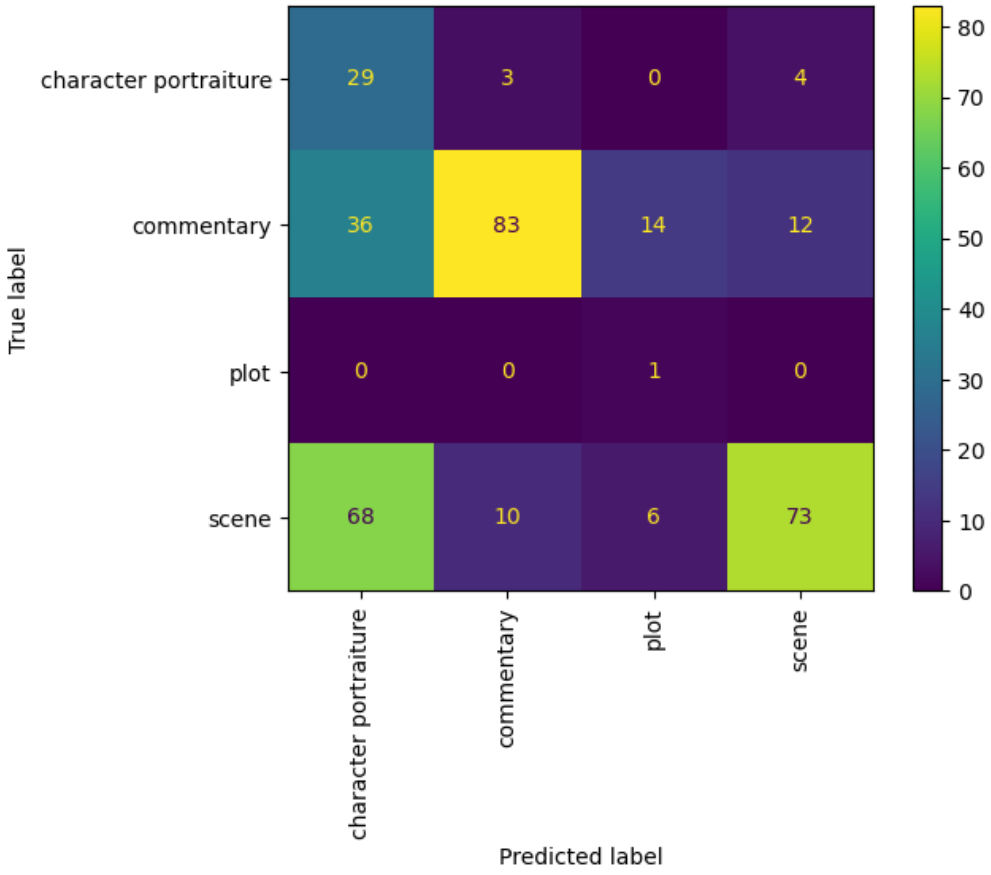

Figure 5: Confusion matrix of content classification using ChatGPT and the long prompt.

Using the long prompt described earlier, we also tested whether the models could identify the position and perspective of the poems. When classifying the position of poems, Llama 3.3 achieved a much higher F1-score than ChatGPT, with only a 0.02 decrease in accuracy. In terms of perspective classification, both Llama models outperformed ChatGPT. Taken together, the classification results across all three dimensions show that ChatGPT and Llama 3.3 have their own strengths. It is worth noting that Llama 3.3 is an open-source, free model and that we can fine-tune using texts from the Ming and Qing periods, which makes it even more appealing for our research in the future.

Table 5: Comparison of classification results using the long prompt across three aspects in the second-round testing.

	Content		Position		Perspective	
Model	**Accuracy**	**F1-score**	**Accuracy**	**F1-score**	**Accuracy**	**F1-score**
ChatGPT	**0.55**	**0.43**	**0.84**	0.34	0.63	0.60
Llama 3.3	0.51	0.33	0.82	**0.55**	**0.75**	**0.73**
llama3-zh-inst	0.21	0.14	0.48	0.26	0.66	0.66

4.3 Analysis of Misclassified Examples

What caught our attention the most among the results in our second-round testing was that ChatGPT incorrectly classified 109 poems into the "character portraiture" category for the content aspect. This represents nearly one-third of the entire dataset. To better understand the reasons behind these misclassifications, we conducted a detailed investigation of the mislabeled examples. This analysis uncovered several distinct patterns, demonstrating the challenges of aligning automated classification with human expertise in Chinese literary traditions.

4.3.1 Scene misclassified as character portraiture

One prominent type of misclassification involves 68 poems categorized by three human annotators as "scene" but classified by ChatGPT as "character portraiture." The following example from our data sample aptly demonstrates this argument:

> "彼時，众人都挪到當中桌子旁邊來，等可人月旦。獨爐湘妃折下一枝菊花，插在瓶中，放在面前，寫「供菊」一題，見了他二人眼睛，看著福壽笑了一笑。只見可人前，擺著紅筆朱硯，先看璞玉的詩:
>
> 懷菊 潤翰公子
> 獨倚東籬思故友，哀吟凄凉增新愁。
> 此心郁郁無人問，斜生彎枝知也無?
> 凉秋已臨我何急，盛時既去汝太羞。
> 艷色秀容今何在? 曼立香迹猶楚楚。
>
> 可人看罷，笑道:「璞公此詩，可謂懷之入骨髓矣，真古今之绝唱也。」。"
>
> [Translation][9]: "At that moment, everyone moved to the table in the center, waiting for Keren's monthly commentary. Lu Xiangfei alone broke off a chrysanthemum branch, placed it in a vase in front of her, and wrote the title "Offering Chrysanthemums." When she met the eyes of the two, she smiled at Fushou. In front of Keren was placed a red brush and a vermillion inkstone. First, she read Puyu's poem:
>
> 'Cherishing Chrysanthemums' *by* Scholar Runhan
> Leaning alone on the eastern fence, thinking of an old friend,
> Sorrowful chants add new grief.
> This heart, melancholic, remains uncared for,
> Do the slanted, bending branches know or not?
> The cool autumn has arrived, why must I hurry?
> The peak of your time has passed, causing you boundless shame.
> Where now is your bright, elegant beauty?
> Still upright, your fragrant traces are clear and pure.
>
> After reading, Keren smiled and said: 'Scholar Pu's poem can truly be

9. The translations presented in this article were created with the help of ChatGPT (GPT-4o).

said to express longing down to the marrow. It is indeed an unparalleled masterpiece of ancient and modern times."'

This example is quoted from the Qing fiction *One-Story Pavillion* (*Yiceng Lou* 一層樓).[10] This example captures a literati gathering engaging in a chrysanthemum-themed poetry contest set against an idyllic autumn backdrop. In such poetry contests, the participants often wove emotional elements into their poems to demonstrate their creativity and artistic finesse. Therefore, these emotions were not necessarily a reflection of the participants' or the novelist's genuine feelings, but rather artistic constructs strategically crafted to enhance the poem's appeal and increase the participants' chances of winning literary competitions. Annotators familiar with this cultural and literary tradition interpreted the poem's primary function as depicting a poetic exchange scene, rather than expressing individual characters' emotions. ChatGPT appears to have a limited understanding of Chinese literary conventions or did not consider the tradition of poetic exchanges when classifying the poems, which might have led to this misclassification.

Another instance of the "scene-to-character portraiture" misclassification involves poems inscribed on objects as decorative elements. The primary function of these poems is to complement the aesthetic or symbolic value of the objects. For instance, in one example from *Shadows of Dream of the Red Chamber* (*Hong Lou Meng Ying* 紅樓夢影),[11] a poem inscribed on a fan was presented and appreciated:

> "賈蘭说：「我的扇子也是他送的，姑姑看見没有?」二人齊说：「没有，你取去。」賈蘭忙忙下樓，不一刻取來。探春接來一看，也是檀香股、絹面，小楷寫的「擬閨詞」七律四首。探春念道:
>
> 東風影里罷梳頭，窗外呢喃聽不休。
> 藻井待栖雙玉剪，筠簾初上小銀钩。
> 疑將軟語商量定，似有柔情宛轉留。
> 銜得新泥重補葺，余香猶記舊妝樓。"
>
> [Translation]: "Jia Lan said, 'My fan was also a gift from him [Xue Pan 薛蟠]. Aunt, have you seen it?' The two replied in unison, 'No, go fetch it.' Jia Lan hurried downstairs and returned shortly after with the fan. Tanchun took it and examined it closely. It was also made with sandalwood ribs and a silk surface, inscribed in fine script with four seven-character quatrains titled 'Imitations of Boudoir Verses'. Tanchun began to read aloud:
>
> Amid the shadows of the spring breeze, she sets her comb aside,
> Listening to the incessant chirps outside the window.
> Beneath the carved ceiling, the scissor-shaped swallows await their perch,

10. https://ctext.org/wiki.pl?if=en&chapter=406211&remap=gb#p7
11. https://ctext.org/wiki.pl?if=gb&chapter=576940&remap=gb#p16

The bamboo blinds newly adorned with a silver sickle moon.
Perhaps, tender words have reached an accord,
As if gentle sentiments linger in their winding flow.
Carrying fresh mud, they rebuild what was once their home,
The lingering fragrance marks the old makeup chamber."

Poems as such were crafted to complement and emphasize the value of the objects that the characters in the story adorned. Therefore, the primary function of these poems was to vividly describe the objects, bringing them to life, while simultaneously showcasing the poetic and literary talents of the fictional writer. The poem under examination in this case was titled "Imitation of Boudoir Verses" (ni guici 擬閨詞), which represents a typical convention of classical Chinese poetry. Such poems are typically written by male poets who adopt a female persona to articulate women's emotions, inner worlds, and everyday experiences—often exploring themes of longing, separation, loneliness, and the transience of youth. Given this literary tradition, the references to a woman (e.g., the use of "she") in the poem should not be interpreted as literal depictions of a fictional character. As such, the poem would not be classified as "character portraiture" in this context. While it is possible the poem may indirectly reflect the thoughts, psychology, or personality of its author (see Rouzer 2001)—and, if the author were indeed a fictional character, potentially serve a characterizing function—this does not apply to the example at hand. The poem appears in a scene where the character Jia Lan displays a fan gifted by Xue Pan. However, close reading of the surrounding narrative reveals no clear indication of who authored the poem inscribed on the fan. This misclassified example further underscores LLMs' struggle to interpret Chinese poetry, especially when meaning emerges from the nuanced interplay between form, tradition, and narrative function.

4.3.2 Commentary misclassified as character portraiture

The second major case of misclassification includes 36 poems that were identified as "commentary" by human annotators while classified as "character portraiture" by ChatGPT. The following case from *A Pillow of Wonders* (*Yizhen Qi* 一枕奇)[12] demonstrates this point:

"莫说丁協公是個富貴公子，他日日要見教的；就是徐鵬子一個窮公孫，他看他考得利肚里又通，也時常虛賣弄，三兩日來鬼混一場去。總不如那丁公子與他貼心貼意，分外相投，一刻也離他不得的。這正是：

嫖賭場中篾片，文章社内法喜。
雖然牌挂假斯文，不如尊綽白日鬼。

却说丁協公看了那條字兒，委决不下，躊躇了一夜，次日侵早，著人去請了白日鬼來。周白日道：「昨日有些小事，不曾會你，場期已迫，看你

12. https://ctext.org/wiki.pl?if=gb&chapter=52354&remap=gb

的氣色好的緊，今科定要高發的。請問呼喚何事見教?」”

[Translation]: “Not to mention Lord Ding Xie, a wealthy nobleman he receives lessons from every day; even Xu Pengzi, a poor descendant of a noble family, who he finds capable in exams and well-read, is someone he frequently interacts with and flatters. He often engages in lighthearted mischief with Xu every two or three days. However, none of this compares to Lord Ding Xie’s deep and heartfelt connection with him. The two are exceptionally close, so much so that Ding cannot bear to part from him for even a moment. This is precisely:

A sidekick in gambling dens, an entertainer in scholarly circles.
Though wrapped in pretended refinement,
Better call him ’Daylight Swindler’.

Now, regarding Lord Ding Xie, he was deeply conflicted after reading the message and deliberated over it for an entire night. Next morning early, he sent someone to invite the ‘Daylight Swindler’ over. Zhou Daylight said, ‘Yesterday, I had some small matters to attend to, so I couldn’t meet with you. With the exam approaching and your energy looking excellent, I’m certain this will be a successful year for you. May I ask, what is the purpose of summoning me today?”’

When examined solely within the extracted sample, the poem proved difficult to understand. This confusion may likely come from the absence of context for terms such as “Daylight Swindler” (白日鬼) or the reference to “he” in the excerpt. To resolve this ambiguity, the annotators revisited the original fiction. They discovered that a character named Zhou De, with the nickname “Daylight Swindler,” was introduced in the preceding paragraphs of the same chapter where this extracted sample is located. This contextual information clarified the meaning of the poem: rather than focusing on interpersonal relationships, it delivers a sarcastic critique of Zhou De’s idle and opportunistic nature. Based on this understanding, annotators classified the poem as a commentary on the character’s personality, which falls under the “commentary” category. By contrast, ChatGPT labeled this poem as “character portraiture.” We asked ChatGPT to explain the rationale behind the classification and were given this answer: “The poem describes the relationship between two characters, showing the preference of one character towards another. It highlights a bond and understanding between the two, giving us insight into their personalities and conduct.” This misclassification may stem from the limited context (i.e., only the two-sentence excerpts immediately before and after the poem) provided to ChatGPT during the classification process, where the key contextual details regarding the character Zhou De were absent. This misclassified example has inspired us to reassess whether two sentences before and after the poem provide sufficient context for LLMs to make informed decisions.

Another scenario of the "commentary-to-character portraiture" misclassification commonly happens among poems following typical semantic cues, such as "there is a poem that proves it" (有詩为證) and "this is precisely" (這正是). The following case from the data illustrates this scenario:

> "鵬子道:「勸你放心。這科包管决中，賠也賠得你一個舉人。若還不中，不但無颜見你，也無面目再見那些親族朋友了。」王氏道:「但願如是，就當拜謝天地。」這正是:
>
> 只謂才不如己，争道巧不猶人。
> 指望一朝騰霄漢，誰知窮鬼不離身。
>
> 却说同學内有一個秀才，姓丁名全，字協公，其人也是世家。乃父累官至工部侍郎，宦途頗順，廣積官資。"
>
> [Translation]: "Pengzi said, 'I urge you to rest assured. I guarantee that I will pass this examination, and the worst case scenario would be a "recommended man". If I don't succeed, not only would I have no face to see you, but I would also have no face to see our relatives and friends.' Madam Wang replied, 'I only hope it will be so; if it happens, we will surely give thanks to Heaven and Earth.' This is precisely:
>
> Some claim others lack their talent,
> Never admits their smartness does not surpass others'.
> Hoping one day to soar to the skies,
> Yet who knew misfortune clings like a shadow.
>
> Now, among the students, there was a scholar named Ding Quan, with the courtesy name Lord Xie. He came from an established family; his father had steadily advanced in his official career, rising to the rank of Vice Minister of the Ministry of Works. His bureaucratic path had been smooth, allowing him to amass significant capital of officialdom."

Similarly from *A Pillow of Wonders*,[13] this poem is composed immediately after the narrator describes the financial and career struggles faced by Pengzi and his family. The poem creates a contrast between Pengzi's ambition and his lack of fortune and luck, reinforcing the central themes established in the preceding context. The phrase "this is precisely" preceding the poem also signals that the narrator of the story is about to reiterate the content presented previously.

To some extent, this poem was indeed about Penzi, a character in the fiction. So, this misclassification may be due to the inherent ambiguity of the poem. However, human annotators were able to distinguish between a "*description* of a character" and a "*comment* on a character's situation," the latter being the correct classification for this poem. This may be rooted in their understanding of the structural cues provided by phrases such as "this is precisely," which emphasize or reiterate the messages

13. https://ctext.org/wiki.pl?if=gb&chapter=52354&remap=gb

stated previously. These observations suggest that introducing the roles of such structural phrases to LLMs, or providing explicit examples for these ambiguous, boundary cases, may help enhance their classification accuracy.

5. Discussion

The two rounds of iterative annotation and testing show that while LLMs hold significant potential for identifying and classifying poetry's narrative functions, limitations remain that must be addressed for the approach to be applicable to the task. To address these limitations, our analysis suggests possible approaches such as enhancing the annotation framework, refining prompts, and incorporating technical considerations.

5.1 Annotation Framework

Reflecting on the annotation process, we recognize that the inherent ambiguity and fluid boundaries between categories may have contributed to the challenges of automatic classification. In particular, the "commentary" category of the content aspect in the revised framework still encompasses a wide range of content, including reflections on characters' personalities, moral lessons, scenes, and even broader societal themes. The lack of clear distinctions between "commentary" on a character or scene and descriptions of a character ("character portraiture") or a scene ("scene") may have led to confusion for LLMs. To address these issues in the next round of annotation and testing, we will focus on refining the "commentary" category by dividing it into more specific and well-defined subcategories.

The single-label annotation system with mutually exclusive narrative-function categories may also partly explain why the models appear to misclassify some poems' narrative functions. To address this potential issue, we plan to experiment with a multi-label annotation framework in place of the current single-label approach, to better capture the complexity and richness of the narrative functions that a poem often serves within fictional storytelling. In addition to revising the annotation framework itself, we will develop a more detailed annotation guideline to support future experiments with few-shot learning. This guideline will incorporate examples for each label, as well as discussions of misclassified cases to clarify the rationale behind function assignments, particularly in boundary or ambiguous instances.

5.2 Prompts Development

Based on our analysis of misclassified cases, we propose the following strategies to enhance prompt design and further improve LLMs classification results: First, incorporate sufficient information and knowledge about classical Chinese literary traditions in the prompts to communicate with LLMs. This addition may help

LLMs develop a more culturally oriented interpretation of the poems in late imperial Chinese novels, such as those in poetic exchange scenes discussed above. Second, expand the accompanying prose contexts for the poems, so that LLMs can look for additional, more accurate cues to understand and interpret the poems. Third, during the second round of testing, we got the best results using the long prompt. We speculate that this may be because the long prompt pushed ChatGPT to develop a more comprehensive understanding of the poems by requiring classification across all three aspects (content, position, and perspective). We plan to test this hypothesis in future work. For example, we will incorporate instructions for classifying position and perspective into the long prompt (binary approach) to see if the overall classification results improve. Finally, our analysis indicates that the semantic diversity of texts may have also confused LLMs. Some of the misclassifications discussed above—particularly those involving the distinctions between a "commentary of a character" and a "description of a character," as well as those triggered by structural cues like "this is precisely"—could be mitigated by using few-shot classification and providing LLMs with a few examples.

5.3 Technical Challenges

Additionally, we also face technical challenges, the most pressing of which is the difficulty in fully understanding how generative LLMs operate. For example, the Llama home page states the following:

> "Llama 3.3 supports 7 languages in addition to English: French, German, Hindi, Italian, Portuguese, Spanish, and Thai. Llama may be able to output text in other languages than those that meet performance thresholds for safety and helpfulness. We strongly discourage developers from using this model to converse in non-supported languages without implementing fine-tuning and system controls in alignment with their policies and the best practices shared in the Responsible Use Guide."[14]

Given this limitation, we tested not only Llama 3.3 but also a fine-tuned version of Llama 3 trained on Chinese text, anticipating that the latter would yield better classification results for Chinese-language data. The results of the test showed that the classification of the majority of the poems was successfully completed by Llama 3.3, except for a few that were answered *"I can't fulfill this request"* (twice) and *"I don't have the capability to view or analyze the Chinese text you provided. Could you please copy and paste the text here, and I'll be happy to help you determine which category it belongs to?"* (three times). What surprises us even more is that Llama 3.3 outperforms the Chinese Llama model across all three aspects of classification and even surpasses ChatGPT in identifying the position and perspective of poems. We do not understand how a model that does not support Chinese can accomplish the task, and it is unclear whether the results of the classification were based on the

14. https://ollama.com/library/llama3.3

model's understanding of the text or it was just a shot in the dark. This suggests that the opacity of LLMs' training data can significantly limit our understanding of why a model produces certain outputs, particularly for culturally dependent tasks. This limitation applies regardless of whether the model is classified as "open" (e.g., Llama 3) or closed-source (e.g., ChatGPT). In future research, we plan to experiment with models trained on larger Chinese corpora, such as DeepSeek, as well as next-generation LLMs as they become available—particularly those with improved capabilities for processing classical Chinese texts.

6. Conclusion

In this article, we explored the use of LLMs to examine the narrative functions of "embedded poems" in Chinese Qing fiction. Specifically, we presented two rounds of iterative annotation processes and LLMs testing. Our analysis revealed the diverse roles that poetry plays in Qing novels and highlighted both the potential and inherent limitations of LLMs for identifying and classifying these functions. Moreover, we found that an integrated refinement approach that encompasses adjustments in annotation, model selection, and testing methodologies can enhance the performance of LLMs for our classification task. After two rounds of refinements, our findings showed that ChatGPT and Llama 3.3 outperformed the other models in our dataset, each demonstrating unique strengths. Moving forward, we will continue to refine our approach to further improve the robustness and accuracy of the classification results. The ultimate goal of our work is to develop a computational approach that analyzes the narrative function of poetry in late imperial Chinese vernacular writings on a large scale, extending beyond the limited corpus of canonical works.

Our work contributes to both Chinese literary scholarship and research on LLMs. Harnessing the power of LLMs to revisit the storytelling dynamics of this rich literary tradition, we can assess and offer insights into the narrative roles of poetry in vernacular novels on a large scale. From the perspective of LLM research, this study highlighted a key limitation of current LLMs: their difficulty in processing culturally distinct corpora. This underscores the need for more rigorous evaluation and experimentation before LLMs can be applied effectively and responsibly in computational literary analysis. We believe that through careful testing, evaluation, and fine-tuning, LLMs can be developed into powerful tools for analyzing multilingual and linguistically complex text—domains that remain underrepresented in mainstream training data, which is predominantly derived from Western, contemporary, and commercially available Internet sources.

7. Data Availability

Data can be found here: https://github.com/dkltimon/EmbeddedPoems

8. Software Availability

Software can be found here: https://github.com/dkltimon/EmbeddedPoems

9. Author Contributions

Rongqian Ma: Conceptualization, Data Curation, Methodology, Investigation, Writing – original draft, Writing – review & editing, Funding Acquisition

Keli Du: Conceptualization, Data Curation, Methodology, Software, Formal Analysis, Visualization, Writing – original draft, Writing – review & editing

Yiwen Zheng: Investigation, Writing – original draft, Writing – review & editing

References

Bamman, David, Kent K. Chang, Li Lucy, and Naitian Zhou (2024). "On Classification with Large Language Models in Cultural Analytics". In: *Computational Humanities Research Conference (CHR 2024)*, 494–507. https://ceur-ws.org/Vol-3834/paper119.pdf.

Bishop, John L (1965). "Some limitations of Chinese fiction." In: *The Far Eastern Quarterly* 15.2, 239–247. 10.1002/andp.19053221004.

Cai, Yijiang (2007). *Hongloumeng shici qufu quanjie* 紅樓夢詩詞曲賦全解. Shanghai: Fudan daxue chubanshe. ISBN: 9780198520115.

Chun, Mei (2009). "Garlic and Vinegar: The Narrative Significance of Verse in 'The Pearl Shirt Reencountered'". In: *Chinese Literature: Essays, Articles, Reviews* 31, 23–43. https://www.jstor.org/stable/20799715.

Conneau, Alexis, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov (2020). *Unsupervised Cross-lingual Representation Learning at Scale*. arXiv: 1911.02116 [cs.CL]. https://arxiv.org/abs/1911.02116.

Cui, Yiming, Ziqing Yang, and Xin Yao (2023). "Efficient and Effective Text Encoding for Chinese LLaMA and Alpaca". In: *arXiv preprint arXiv:2304.08177*. https://arxiv.org/abs/2304.08177.

Fleiss, Joseph L. and Jacob Cohen (1973). "The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability". In: *Educational and psychological measurement* 33.3, 613–619. https://doi.org/10.1177/001316447303300309.

Guo, Jie (1995). "Zhongguo gudian xiaoshuo zhong shiwen ronghe chuantong de yuanyuan yu fazhan 中國古典小说中詩文融合傳統的淵源與發展". In: *Zhongguo wenxue yanjiu* 中國文學研究 2, 11–17. CNKI:SUN:ZWXY.0.1995-02-001.

Hegel, Robert E. (1985). "Distinguishing Levels of Audiences for Ming-Ch'ing Vernacular Literature: A Case Study". In: *Popular Culture in Late Imperial China*. Ed. by David Johnson, Andrew J. Nathan, and Evelyn S. Rawski. Berkeley and Los Angeles: University of California Press, 112–142.

Huang, Chihan and Xiaobo Shen (2025). "PoemBERT: A Dynamic Masking Content and Ratio Based Semantic Language Model For Chinese Poem Generation". In: *Proceedings of the 31st International Conference on Computational Linguistics*, 50–60. https://aclanthology.org/2025.coling-main.5/.

Konle, Leonard, Merten Kröncke, Simone Winko, and Fotis Jannidis (2023). "Connecting the Dots. Variables of Literary History and Emotions in German-language Poetry". In: *Journal of Computational Literary Studies* 2.1, 1–22. 10.48694/jcls.3604.

Lewis, Mike, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer (2019). "BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension". In: *CoRR* abs/1910.13461. http://arxiv.org/abs/1910.13461.

Liu, Chao-Lin, Thomas J. Mazanec, and Jeffrey R. Tharsen (2018). "Exploring Chinese Poetry with Digital Assistance: Examples from Linguistic, Literary, and Historical Viewpoints". In: *Journal of Chinese Literature and Culture* 5.2, 276–321. https://muse.jhu.edu/article/724617.

Luo, Ye (1965). *Zuiweng tanlu* 醉翁談録. Taipei: Shijie shuju.

Rao, Longsun (2023). "Mingdai xiaoshuo zhong de shiciqu jisheng shuo bianyi—yu Zhao Yishan xiansheng shangque 明代小说中的詩詞曲寄生说辨義—與趙義山先生商榷". In: *Zhongshan daxue xuebao (shehui kexue ban)* 中山大學學報 (社會科學版) 63.01, 29–37. 10.13471/j.cnki.jsysusse.2023.01.003.

Reiter, Nils, Marcus Willand, and Evelyn Gius (2019). "A Shared Task for the Digital Humanities Chapter 1: Introduction to Annotation, Narrative Levels and Shared Tasks". In: *Journal of Cultural Analytics* 4.3. https://doi.org/10.22148/16.048.

Rouzer, Paul (2001). *Articulated Ladies: Gender and the Male Community in Early Chinese Texts*. Cambridge, MA: Harvard University Asia Center.

Vierthaler, Paul (2016). "Fiction and History: Polarity and Stylistic Gradience in Late Imperial Chinese Literature". In: *Journal of Cultural Analytics* 1.1. https://doi.org/10.7910/DVN/4ZVSKA.

Walsh, Melanie, Anna Preus, and Maria Antoniak (2024). "Sonnet or Not, Bot? Poetry Evaluation for Large Models and Datasets". In: *Findings of the Association for Computational Linguistics: EMNLP 2024*. https://aclanthology.org/2024.findings-emnlp.914.pdf.

Williams, Adina, Nikita Nangia, and Samuel Bowman (2018). "A Broad-Coverage Challenge Corpus for Sentence Understanding through Inference". In: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*.

New Orleans, Louisiana: Association for Computational Linguistics, 1112–1122. http://aclweb.org/anthology/N18-1101.

Yeh, Chia-ying (2004). "Mantan Hongloumeng zhong de shici 漫談紅樓夢中的詩詞". In: *Shaanxi shifan daxue xuebao (zhexue shehui kexue ban)* 陝西師範大學學報: 哲學社會科學版 33.3, 58–64. doi:CNKI:SUN:SXSS.0.2004-03-010.

Yu, Haoyang, Chang Gao, Xingsen Li, and Lingling Zhang (2024). "Ancient Chinese Poetry Collation Based on BERT". In: *Procedia Computer Science* 242, 1171–1178. https://doi.org/10.1016/j.procs.2024.08.179.

Zhang, Jiaxing, Ruyi Gan, Junjie Wang, Yuxiang Zhang, Lin Zhang, Ping Yang, Xinyu Gao, Ziwei Wu, Xiaoqun Dong, Junqing He, Jianheng Zhuo, Qi Yang, Yongfeng Huang, Xiayu Li, Yanghan Wu, Junyu Lu, Xinyu Zhu, Weifeng Chen, Ting Han, Kunhao Pan, Rui Wang, Hao Wang, Xiaojun Wu, Zhongshen Zeng, and Chongpei Chen (2022). "Fengshenbang 1.0: Being the Foundation of Chinese Cognitive Intelligence". In: *CoRR* abs/2209.02970. https://doi.org/10.48550/arXiv.2209.02970.

Zhang, Zhejun (2015). "Proven with a poem: Confucian Classic Origins and Functions of the Phrase in Novels and Traditional Operas (Youshi weizheng: Xiaoshuo, Xiqu taoyu de jingxue yuanyuan yu gongneng 有詩为證：小说、戲曲套語的經學淵源與功能)". In: *Journal of Macao Polytechnic University* 澳門理工學報 3, 154–161.

Zhao, Yanwei (1966). *Yunlu manchao* 雲麓漫鈔. Beijing: Zhonghua shuju.

Zhao, Yishan (2014). *Mingdai jisheng ciqu yanjiu* 明代寄生詞曲研究. Beijing: Shangwu yinshuguan.