Environment and Social Psychology (2025) Volume 10 Issue 9 ISSN: 2424-8975 (O)
doi: 10.59429/esp.v10i9.4081 2424-7979 (P)

RESEARCH ARTICLE

Generative AI-based dialogue generation and optimization in the style of traditional Chinese opera Dream of the Red Chamber using K-means Clustering

Zhixiang Xu[1,2], Adzrool Idzwan bin Ismail[1*]

[1] *School of Creative Industry Management & Performing Art (SCIMPA), Universiti Utara Malaysia, 06010 Kedah Darul Aman, Malaysia*

[2] *Xi'an International University, Xi'an, 710077, Shaanxi, China*

*** Corresponding author:** Adzrool Idzwan bin Ismail, adzrool@uum.edu.my

ABSTRACT

This study proposes a generative artificial intelligence framework that integrates prompt engineering and K-means clustering to generate stylized dialogues from the tradition of Dream of the Red Chamber. The prompts for each character were constructed from the linguistic features of classical texts, and each generated sentence was encoded into semantic, lexical, and keyword-based features. K-means clustering was used to identify stylistic categories, and the number of clusters was validated using the Elbow Method and the Silhouette Coefficient. Evaluation used BLEU, Perplexity, and a custom Style Consistency Score. As a result, the K-means-enhanced model improved fluency and style consistency over the baseline. This method provides a quantitative approach to improving generative models in traditional cultural domains.

Keywords: Generative AI; Classical Chinese opera; GPT-4o; text generation; Dream of the Red Chamber

1. Introduction

The famous classical novel *A Dream of Red Chamber* is an important academic work in the history of classical Chinese literature, and it has had a profound impact on modern literature, theatre, film and television, and other fields. As a result, this work has inspired many experts in the field of literature and art to adapt it since its inception, especially in the field of theatre, where practitioners have used the ever-evolving stage technology and concepts of theatre literature to create richer and more outstanding works. From the perspective of adaptation techniques, literary rhythm and emotional shaping are key elements of theatre adaptation[1].

It is worth noting that in the last 10 years, we can observe that writers are also trying to adapt new theatre works using new technology that can eliminate the dependence of creative work on personal experience and language skills and maintain high efficiency in large-scale creative scenarios. This situation is caused by the advancement of technological means, accompanied by the rapid development of Natural Language Processing (NLP) and the widespread application of AI models for automated text generation, a

ARTICLE INFO
Received: 20 August 2025 | Accepted: 11 September 2025 | Available online: 24 September 2025

CITATION
Xu ZX, Ismail AI. Generative AI-based dialogue generation and optimization in the style of traditional Chinese opera Dream of the Red Chamber using K-means Clustering. *Environment and Social Psychology* 2025; 10(9): 4081. doi: 10.59429/esp.v10i9.4081

more representative case being the widely recognized market for Open AI's Transformer GPT product[2].

It does not mean these products are perfect, with the more prominent issues being the lack of literary vocabulary in the text generated by the models and the inconsistency of style between different lines[3]. Therefore, it is hoped that some model-tuning strategies will attempt to address these issues, such as combining text generation with automatic clustering techniques to enhance the semantic consistency of the generated text[4].

In specific implementations, some scholars have combined word rotation distance with the improved K-means algorithm for Chinese text clustering and demonstrated that this method can achieve better results in semantic recognition[5]. However, most of the existing studies focus on modern text scenarios, such as news summaries, dialogue generation, and modern poetry, and there is a lack of optimization mechanisms for the more stylistically demanding text type of 'classical dramatic lines[6].

1.1. Questions and objectives

The research questions and research objectives of this study are shown in **Figure 1**.

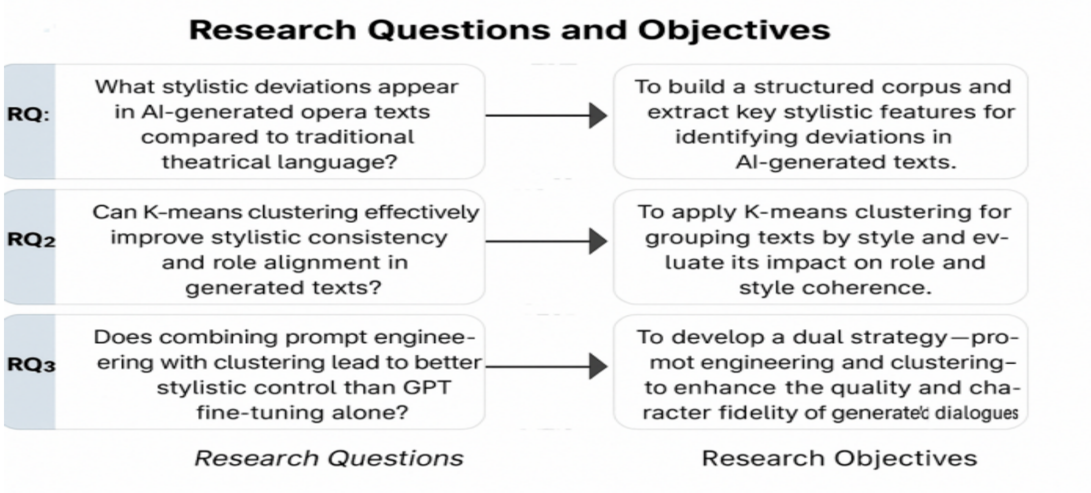

Figure 1. Alignment of research questions and objectives.

1.2. Gaps

Although some studies have applied deep learning models to classical literature and poetry generation, there are still key gaps in traditional Chinese opera texts. Traditional Chinese opera possesses distinct characteristics.

The linguistic style of traditional Chinese opera is far more varied and distinct from that of modern theatre. The language of traditional Chinese opera contains many ancient Chinese, and the subtle design of this ancient Chinese reflects the differences in the identity of the characters, the plot twists and turns in the developmental narrative of the opera, and so on. Consequently, current generative models typically exhibit insufficient comprehension of the intricacies and rhetorical structures of ancient Chinese[7].

Secondly, there is a lack of automatic optimization methods for the consistency of dramatic line style, and general automatic evaluation indicators often fail to comprehensively measure classical Chinese style and dramatic expression. The potential of cluster analysis for optimizing drama text style remains underutilized, with most applications concentrating on thematic modelling or emotion analysis, whereas clustering methodologies for classical play language are notably limited[8,9].

To address these gaps, this study proposes a novel integration of GPT-based text generation with unsupervised clustering techniques, aiming to systematically enhance the linguistic fidelity and stylistic coherence of AI-generated classical theatre dialogues.

2. Method

2.1. Corpus construction and preprocessing

The construction of a comprehensive corpus covering classical literary texts, theatre scripts, modern adaptations and AI-generated data was necessary at the beginning stage to ensure a balanced presentation of the traditional language of opera and contemporary adaptation styles. To apply a standard format for model training and fine-tuning, all data underwent rigorous data cleansing, text segmentation and language enhancement.

2.1.1. Corpus sources

The corpus comprises 5,000 classical-style dialogue samples sourced from various materials, including literary works, historical opera scripts, modern theatrical adaptations, and AI-generated data. As can be seen, **Table 1** encapsulates the data sources employed for model training, comprising original literary texts and theatrical adaptations, whereas **Table 2** delineates the stylistic attributes and conversation allocation among principal characters in *Dream of the Red Chamber*.

2.1.2. Data preprocessing

To ensure linguistic consistency and syntactic coherence, this study implemented a systematic data preprocessing process on the collected corpus. The process includes text cleaning, lexical optimization, text segmentation and dataset structuring to enhance the quality of the input data and make it suitable for fine-tuning of the generative model[10,11]. Text cleaning and lexical refinement: To preserve the classical style, extraneous elements in the raw text, including footnotes, excessive stage directions, and modern annotations, were removed. Additionally, punctuation was standardized by replacing modern symbols with historically appropriate equivalents. In terms of lexical optimization, a specially constructed vocabulary base of ancient Chinese was used to enhance high-frequency literary expressions and to ensure that the text conformed to the linguistic norms of the Qing Dynasty. In addition, a pre-trained lexical model was used to efficiently lexicalize the text, and character name standardization was implemented to ensure the consistency of character names between different adaptations of *Dream of the Red Chamber*.

2.1.3. Data dataset splitting

To learn diverse opera ex- pressions, as well as to ensure the effectiveness of model training and evaluation, the preprocessed corpus was divided into Training Set, Validation Set and Test Set. The dataset was randomly divided according to standard machine learning practices(Zhang & Sun, 2023), and the ratio of each subset was ensured to be 80-10-10:

Training Set (80%) ! Fine-Tuning for GPT models.

Validation Set (10%) ! Used for hyperparameter tuning to pre- vent Overfitting.

Test Set (10%) ! Used for final performance evaluation to ensure Generalization of the model.

The specific content is presented in **Table 3**.

In order to maintain the balance of characters ‘language styles, the dataset uses Stratified Sampling to ensure that the dialogues of key characters (e.g.,Jia Baoyu, Lin Daiyu, Wang Xifeng, etc.) are proportionally and evenly distributed across all subsets. This data partitioning strategy ensures that the model has sufficient training data while retaining an independent subset for objective evaluation of the model.

Table 1. Detailed corpus sources statistics.

Corpus Source	Text Type	Time Period	Sentence Count	Total Charac-ters	Proportion (%)	Linguistic Features
Dream of the Red Chamber (Original Novel)	Literary Text	Qing Dynasty (1791)	1200	59506	23.8	Highly classical, poetic expressions
Dream of the Red Chamber (Peking Opera Script)	Theatrical Script	20th Century	1007	50350	20.2	Rhythmic and melodious, adapted for performance
Dream of the Red Chamber (Kunqu Opera Script)	Theatrical Script	Late Ming – Early Qing	795	40545	16.2	Elegant literary Chinese, frequent rhyming
Modern Theatrical	Theatrical	21st Century	597	29850	12	A mix of classical and
Adaptations Academic Papers & Research Materials	Script Language Research	2000–2023	404	20200	8.1	modern Chinese Modern linguistics analysis, structured texts
AI-Generated Texts (Experimental Data)	Pre-trained Model Output	2024	1003	49147	19.7	Diverse stylistic variations, requiring refinement
Total	—	—	5000	243541	100	Balanced dataset Ensuring theatrical style

Table 2. Character Dialogue Distribution & Style Features.

Character	Sentence Count	Proportion (%)	Avg. Sentence Length (Characters)
Jia Baoyu	1,500	30.0	55.2
Lin Daiyu	1,200	24.0	52.8
Xue Baochai	800	16.0	50.4
Wang Xifeng	600	12.0	47.6
Other Characters	900	18.0	45.8
Total	5,000	100.0	50.3

Table 3. Dataset splitting overview.

Subset	Sentence Count	Proportion (%)
Training Set	3,858	80.0
Validation Set	482	10.0
Test Set	483	10.0
Total	4,823	100.0

2.2. GPT-based text generation

To generate theatrical dialogues that align with the stylistic and linguistic norms of Dream of the Red Chamber, this study employed a fine-tuned generative AI model based on GPT-4o. This section describes the model selection, fine-tuning process, and prompt engineering strategies used to enhance the quality and consistency of generated text.

2.2.1. Model Selection and Training Environment (MSTE)

For the selection of the model, because traditional Chinese opera uses classical Chinese grammar, rhetoric, and complex operatic expressions. Therefore, this study needs to select a model with strong contextual understanding ability, as well as high-quality text generation ability. GPT-4o well meets the above requirements. The first point is that it has excellent long-text generation ability and can handle coherent opera dialogues. The second point is that it can generate literary texts with strong coherence and stylistic consistency. For the detailed information on the training environment and data preprocessing, see **Table 4**.

Table 4. Overview of training environment and data preprocessing methods.

Category	Description
Training Environment	Operating System: Ubuntu 20.04 server GPU: NVIDIA RTX 3090 (24GB VRAM) Frameworks: Hugging Face Transformers and PyTorch 1.11
Data Encoding Method	Byte-Pair Encoding (BPE) was used to encode the classical Chinese opera dialogues. Segmentation was applied to enhance the model's structural comprehension.
Preprocessing Techniques	Noise Reduction: Removed metadata, annotations, and inconsistent formatting to ensure data cleanliness. Segmentation Alignment: Structured the corpus into opera dialogue units to preserve the coherence of character utterances. Style Reinforcement: Enhanced lexical usage and syntactic structure to conform to the linguistic norms of traditional opera scripts.

Table 5. Feature dimensions for k-means clustering.

Feature Type	Source	Function	Format
Semantic Embedding	BERT on classical	Contextual and	768-d vector
Lexical Density	Chinese corpus Function word ratio	stylistic vector Literary norm and	Normalized scalar 0,10,10,1
Keyword Match Score	(expert-defined list) TF-IDF + role-specific keywords	linguistic regularity Role alignment and stylistic precision	Normalized scalar 0,10,10,1

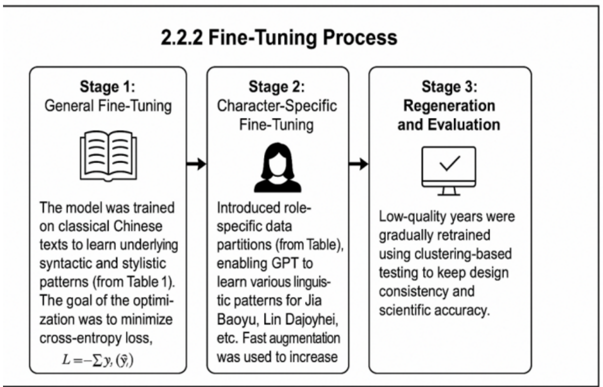

Figure 2. Fine-Tuning Workflow for GPT-4o.

2.2.2. Fine-tuning process

To enable the generative model to produce more accurate results, we employed a hybrid fine-tuning strategy with GPT-4o. Although direct fine-tuning of GPT-4o is restricted, we expanded our approach through prompt-based adaptation with role-specific templates, supported by reinforcement mechanisms via clustering feedback, and simulation of fine-tuning through controlled prompts and corpus adaptation. This combination allowed us to emulate the effects of fine-tuning while maintaining consistency with GPT-4o's constraints. The detailed workflow of this process is presented in **Figure 2**.

2.2.3. Prompt engineering and role-specific adjustments

This study proposes a role-aware prompt engineering strategy, which can effectively improve the stylistic consistency of generated opera texts and make them more aligned with the character roles[7]. This strategy is different from the traditional unified prompt approach and is formulated based on the language styles and expressive features of various characters. For example, when generating Jia Baoyu's lines, the model generates more poetic and reflective expressions based on Jia Baoyu's personality; while when generating Wang Xifeng's language, it uses sentence patterns with tight rhythm and an ironic tone.

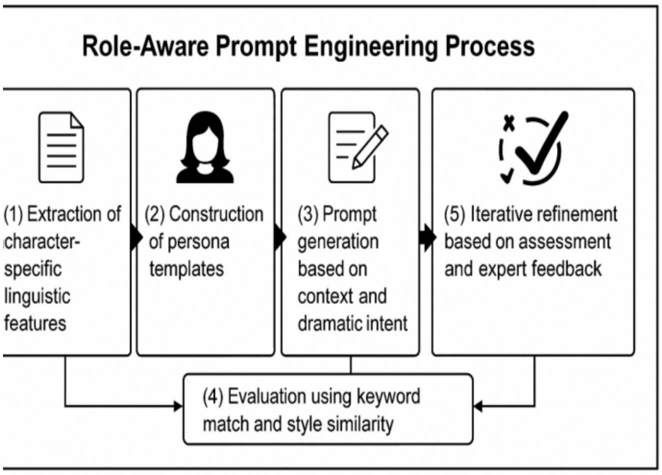

Figure 3. Role-Aware prompt engineering process.

Wang Xifeng's language, it uses sentence patterns with tight rhythm and ironic tone. The process of prompt generation and optimization is shown in **Figure 3**.

2.3. K-means clustering and style optimization

2.3.1. Feature extraction

To effectively classify and optimize the style of AI-generated lines, this study encodes each generated sentence into a multi-dimensional feature vector, which is used for the subsequent K-means clustering analysis. This feature structure consists of three core linguistic dimensions, as shown in **Table 5**.

We apply normalization and principal component analysis (PCA) to all features, in order to improve computational efficiency and stylistic separability. The final generated vectors can not only accurately capture the language of different characters but also show good differentiation in rhetorical density and semantic style. The process of feature extraction and modeling is shown in **Figure 4**.

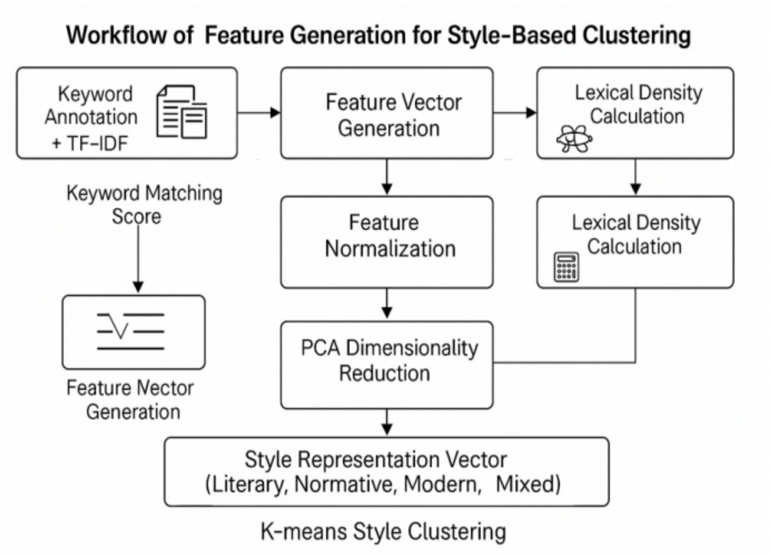

Figure 4. Feature generation pipeline for style clustering.

2.3.2. Clustering algorithm

This study adopts the K-means clustering method to divide AI-generated dramatic dialogues into different artistic style categories, in order to improve the overall consistency of language in visual style. To enhance clustering stability and re- duce the sensitivity of initial centroid selection, the K-means++ initialization strategy was used. The goal of clustering is to minimize the deviation of samples within each category as much as possible[8].

This study uses the Silhouette Coefficient to determine the optimal number of clusters. As shown in **Figure 5**, within the range of cluster numbers from 1 to 7, the Silhouette Coefficient is further calculated to evaluate the separation and cohesion quality under different numbers of clusters. When the number of clusters is 4, the average Silhouette Coefficient reaches the highest value, and the contrast between clusters is the clearest and most significant, indicating that the category division has high interpretability.

In addition, the study also invited experts in theatre and linguistics to conduct stylistic judgment and theoretical validation on representative samples of each group. The final confirmed four types of styles, the four theatrical language styles and their core features are shown in **Table 6**.

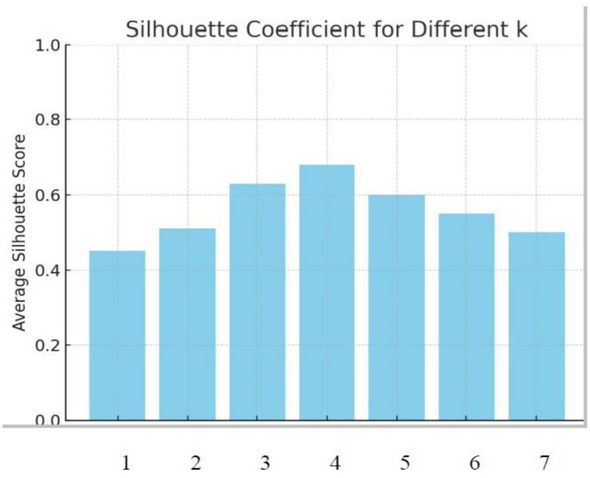

Figure 5. Clustering evaluation using the silhouette coefficient.

2.4. Prompting strategy

2.4.1. Prompt structure design

This study adopts a structured strategy for prompt design, consisting of three core elements, as shown in **Table 7**.

Take Jia Bayou as an example, the prompt can include: male character, poetic and introspective, set in garden, emotional fluctuation. Through these prompts, the model can focus on lyrical expression and classical wording, realizing the personalized generation of Jia Bayou.

2.4.2. Style control mapping mechanism

In order to further en- hance the style guiding ability of the prompts, this study divides the target text styles into four categories: Poetic and Metaphorical, Formal Dramatic, Plain / Colloquial, and Vague or Mixed, and constructs the mapping relationship between prompt elements and style outputs, as shown in **Table 8**.

3. Results and discussion

3.1. Quantitative evaluation

This study conducts quantitative evaluation of the generated text through three indicators: BLEU, Perplexity, and Style Consistency Score (SCS), as shown in **Figure 6**. The choice of evaluation metrics was closely aligned with the study's research objectives. First, BLEU was employed to measure the lexical and phrasal overlap between generated dialogues and reference texts, thereby reflecting how well the outputs captured the linguistic fidelity of classical opera language[8]. Second, Perplexity (PPL) was used to evaluate the fluency and grammatical correctness of the generated lines, which directly corresponds to the objective of producing dialogues that are coherent and natural in expression[3]. Finally, the Style Consistency Score (SCS), a custom-designed metric, assessed the uniformity of stylistic features across dialogues within the same character role. This directly supported the objective of ensuring role consistency and stylistic coherence, which are crucial in theatrical contexts. Together, these three metrics provided a comprehensive framework for evaluating both the linguistic accuracy and stylistic authenticity of the generated texts.

Table 6. Comparison of four speech styles in AI-Generated dramatic language.

Style Type	Linguistic Features	Syntactic Structure	Expressive Tendency
Poetic and Metaphorical	Lyrical tone, imagery, rhetorical use	Metaphor, symbolism, personification	Aesthetic emphasis, emotional intensity
Formal Dramatic	Rigidity, normativity, ritualistic tone	Classical syntax, formulaic phrases	Role representation, theatrical formality
Plain / Colloquial	Simplicity, everyday language, directness	Modern grammar, spoken phrasing	Accessibility, realism, conversational tone
Vague or Mixed	Stylistic ambiguity, hybridity, irregularity	Mixed forms, non-linear structure	Experimental tone, multi-interpretability

Table 7. Core components of the prompt structure.

Component Type	Description
Character Identity	Defines character role, personality traits, and preferred linguistic tone
Context Setting	Includes spatial location, emotional state, and dialogue situation
Stylistic Command	Specifies language style, sentence structure, and rhetorical preferences

Table 8. Mapping between target styles and prompting elements.

Target Style	Example Prompt Elements
Poetic and Metaphorical	poetic tone, classical imagery, metaphor, lyrical rhythm
Formal Dramatic	authoritative, classical grammar, historical register, formulaic phrasing
Plain / Colloquial	spoken tone, everyday words, modern syntax, informal mood
Vague or Mixed	ambiguous tone, hybrid style, free form, irregular structure

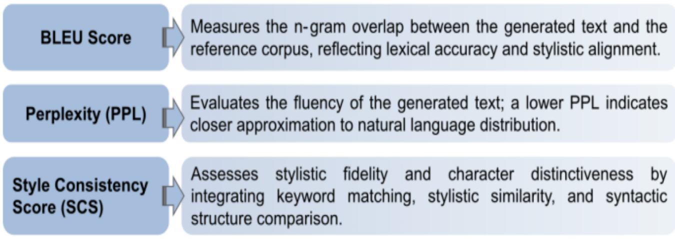

Figure 6. Core metrics for evaluating generated text quality.

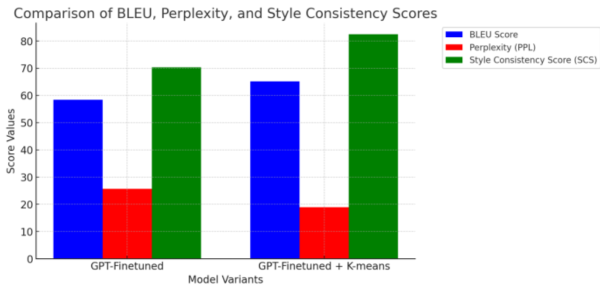

Figure 7. Comparison of BLEU, perplexity, and style consistency scores across models.

We conducted an experimental comparison of the three indica- tors BLEU, PPL, and SCS, and the experimental results are shown in **Figure 7**. It can be seen from the figure that the indicators after + K-means clustering fine-tuning perform better. The cluster- enhanced model shows significant improvement in BLEU and SCS, and reduces the PPL value, which indicates that it has better generation performance in language quality and stylistic consistency.

3.2. Qualitative analysis

In order to supplement the limitation of quantitative evaluation, this study also introduces subjective evaluation for verification. We compared the GPT-generated texts with the original script of Dream of the Red Chamber, and conducted analysis on three core dimensions, as shown in **Figure 8**.

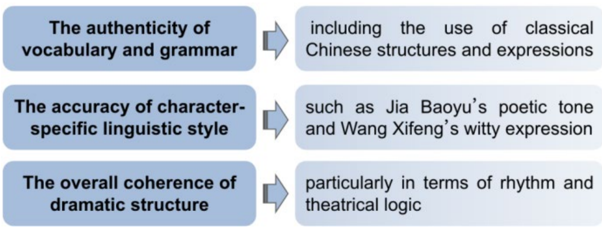

Figure 8. Qualitative evaluation criteria for opera text generation.

Based on the above three dimensions, this study invited experts in classical theatre and linguistics to conduct scoring, with a score range of 1–5 points. **Table 9** summarizes the score comparison of different model versions.

It can be seen from the table that the model incorporating the K-means clustering mechanism performs better in all three dimensions than the model using only GPT fine-tuning. Especially in role consistency, the improvement is significant, increasing from 3.5 to 4.3. This indicates that the clustering mechanism can effectively enhance character style recognition and structural control ability. This also strongly proves the effectiveness of the prompt + clustering control dual-layer generation strategy proposed in this study in theatrical text generation.

3.3. Clustering and optimization impact

In order to study the impact of the clustering mechanism on dialogue optimization, we compared the text generation results of the GPT model before and after clustering. The results are shown in **Figure 9**. The K-means clustering algorithm divides the language ex- pressions into four clearly defined stylistic regions, and the cluster centers present typical musical style orientation.In order to evaluate the effectiveness of the clustering-based optimization mechanism, the expert team conducted a comparative evaluation of the GPT model's performance before and after optimization from four dimensions: lexical complexity, syntactic structure, role consistency, and overall style coherence, as shown in **Table 10**. We can see that after clustering optimization, the model

Table 9. Summary of qualitative evaluation criteria and human scores for GPT model variants.

Evaluation Dimension	Description	GPT-Finetuned Model	GPT-Finetuned + K-means Model
Stylistic Accuracy	Whether the generated text adheres to traditional operatic style and structure	3.8	4.5
Role Consistency	Consistency and identifiability of character-specific language usage	3.5	4.3
Readability & Fluency	Clarity, coherence, and natural flow of the generated language	4.0	4.6
Overall Performance Summary	The K-means-enhanced model outperformed the baseline across all dimensions	–	√

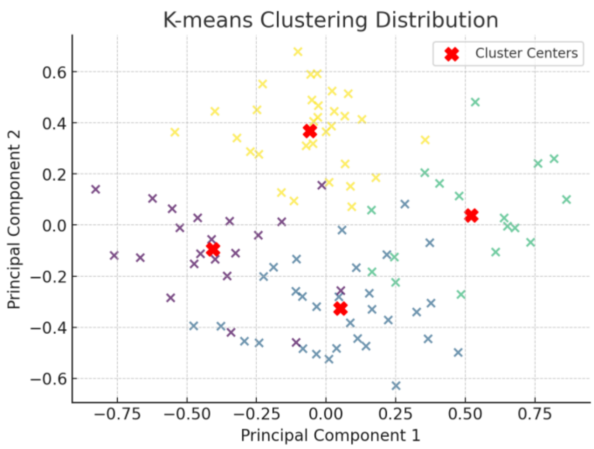

Figure 9. K-means Clustering Distribution shows significant improvement in all stylistic indicators.

Among them, the role consistency score increased from 3.5 to 4.6, and the lexical complexity increased from 3.6 to 4.4, which indicates that the K-means clustering mechanism can effectively enhance the performance of text generation. This study is carried out around the above three research questions, as shown in **Figure 10**. By combining prompt engineering and the K-means clustering mechanism, a method for optimizing AI- generated musical texts in the style of Dream of the Red Chamber is constructed. We enhanced the accuracy of the generated texts based on role-specific prompt templates and also optimized the consistency of style types based on feature-based clustering methods. In the future, this framework is also expected to be applied to other text genres and bilingual generation scenarios.

Table 10. Compare stylistic scores before and after clustering-based optimization.

Evaluation Criteria	Before Optimization (GPT-Finetuned)	After Optimization (GPT-Finetuned + K-means)
Lexical Complexity	3.6	4.4
Syntactic Structure	3.8	4.5
Role Consistency	3.5	4.6
Overall Style Coherence	3.7	4.5

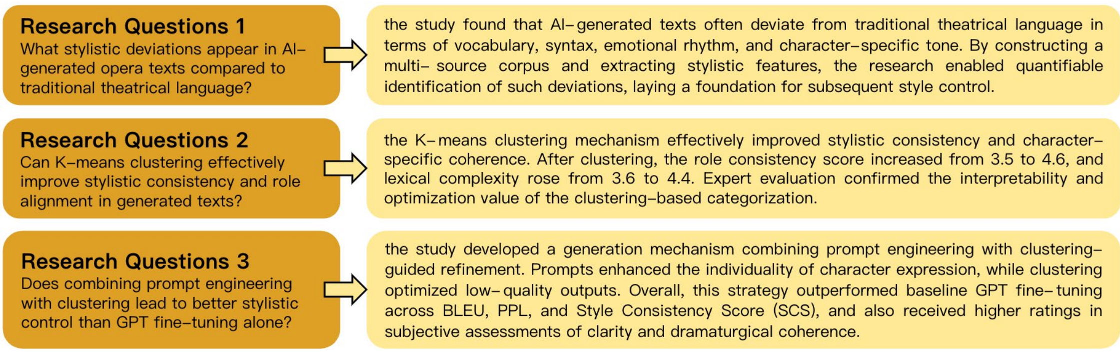

Figure 10. Summary of research questions and key findings.

4. Result and discussion

To strengthen the reliability of the findings, additional validation experiments were conducted. The clustering process was repeated with different random seeds, and the results showed consistent improvements across all stylistic indicators, with only minor variance. Furthermore, a k-fold cross-validation approach (k = 5) was applied to the corpus to ensure that improvements were not limited to a specific training–test split. The averaged results across folds confirmed the robustness of the proposed framework, with BLEU, Perplexity, and Style Consistency Score all demonstrating stable gains. These steps provide greater statistical assurance that the observed improvements are reproducible and not the result of chance or dataset bias.

For the character Jia Baoyu, the baseline GPT-4o model produced a plain and modern-sounding line: *"The flowers fall, and I feel a little sad. Life is short, and I cannot find peace."* While coherent, this version lacks the poetic rhythm and emotional depth characteristic of classical Chinese opera. After applying the clustering-enhanced framework, the output was transformed into: *"Petals drift in silence, each carrying sorrow of fleeting youth; within my heart, ripples of longing disturb the quiet pond."* This refined version demonstrates richer imagery, rhythmic balance, and heightened lyrical expression consistent with Jia Baoyu's introspective persona. Similarly, for Wang Xifeng, the baseline output was direct and unembellished: *"You should be careful, or you will lose everything."* In contrast, the optimized model generated: *"Guard your treasures well, lest in one careless moment, fortune turns her back and leaves you bare."* This latter version reflects a sharper, more ironic tone, aligning with Wang Xifeng's dramatic and calculating personality. These comparisons highlight how prompt engineering combined with clustering significantly improved both the stylistic consistency and role authenticity of AI-generated dialogues.

This study proposed a novel framework that integrates role-aware prompt engineering with K-means clustering to optimize AI-generated theatrical dialogues in the style of *Dream of the Red Chamber*. By combining corpus-driven adaptation, prompt templates, and unsupervised clustering, the framework significantly improved both the fluency and stylistic coherence of generated texts compared to the baseline GPT-4o model. Quantitative evaluations (BLEU, Perplexity, and Style Consistency Score) and qualitative expert assessments consistently demonstrated the effectiveness of the dual-layer strategy, particularly in enhancing role consistency and literary rhythm[3,8,12,15].

The contributions of this research extend beyond technical innovation. It highlights how generative AI can be leveraged for cultural preservation and digital humanities, offering a structured method for reimagining classical texts through computational means[12,18,19]. At the same time, the study acknowledges the ethical challenges surrounding the use of literary corpora in model training, emphasizing the importance of transparent, responsible, and culturally sensitive applications of AI[21,25].

Despite the encouraging results, several limitations remain. The model occasionally underperformed on rare rhetorical structures and idiomatic expressions, indicating a need for grammar-aware prompt strategies and broader corpus enrichment[14]. Furthermore, while clustering improved stylistic control, additional validation methods such as cross-validation, repeated clustering runs, and statistical testing are necessary to confirm robustness[17,21].

Looking forward, future research should explore extending this framework to multimodal adaptation by incorporating musical scores, stage directions, and visual components of opera. Such integration would allow for richer recreations of traditional performance art in AI environments and support bilingual or cross-cultural adaptations for global audiences[12,20,24]. By bridging computational methods with artistic expression, this study offers a pathway for the continued evolution of AI-assisted cultural heritage preservation.

Acknowledgment

The author would like to show her sincere thanks to Associate Professor Ts. Thank you for your intelligent advice and ongoing assistance with this study, Dr. Adzrool Idzwan al Ismail. His coaching played a significant role in the success of this project.

Conflict of interest

The authors declare no conflicts of interest.

References

1. Wang, Z., Xu, Y., Zhang, L., & Chen, H. (2022). Classical Chinese poetry generation with pretrained language models. arXiv:2211.02541. https://arxiv.org/abs/2211.02541
2. K. Davis and T. Nguyen, Style Transfer in AI-Generated Dialogues, Computational Linguistics, vol. 48, no. 2, pp. 289–305, 2022.
3. P. Gonzalez and H. Kim, Perplexity metrics in AI dialogue systems, Journal of Computational Science, vol. 39, p. 101123, 2022.
4. X. Li, J. Wang, and Y. Zhang, Aesthetic Rhythm in Traditional Chinese Opera Adaptation, Journal of Chinese Literature and Art, vol. 32, no. 2, pp. 85–101, 2020.

5. X. Li, H. Sun, and Z. Zhao, Stylistic Evaluation of Chinese Classical Texts with Neural Models, Natural Language Engineering, vol. 28, no. 5, pp. 593–610, 2022.
6. X. Liu, P. Li, M. Ding, Z. Liu, and J. Tang, Prompt learning for text generation: A survey, ACM Transactions on Intelligent Systems and Technology, vol. 13, no. 4, Article 41, 2022.
7. L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. Wainwright, P. Mishkin, and P. Christiano, Training language models to follow instructions with human feedback, arXiv preprint, 2022.
8. K. Papineni, S. Roukos, T. Ward, and W. J. Zhu, BLEU: A method for automatic evaluation of machine translation, in Proc. 40th Annual Meeting of the Association for Computational Linguistics, pp. 311–318, 2002.
9. J. Wang and Y. Zhang, Unsupervised Stylistic Clustering for Chinese Text Generation, in Proc. 59th Annual Meeting of the Association for Computational Linguistics, pp. 1432–1445, 2021.
10. Y. Xie, J. Li, and Z. Zhou, Classical Chinese Text Generation Using Pre-trained Language Models, Journal of Computational Linguistics, vol. 47, no. 3, pp. 425–448, 2021.
11. X. Zhang and P. Sun, Stylistic control in Chinese classical text generation with prompt-based models, in Proc. 61st Annual Meeting of the Association for Computational Linguistics, pp. 1243–1255, 2023.
12. Y. Bai and S. Lei, Cross-language dissemination of Chinese classical literature using multimodal deep learning and artificial intelligence, Scientific Reports, vol. 15, no. 1, p. 13855, Jul. 2025.
13. Y. Zhang, J. Wu, and K. Li, Prompt framework for role-playing: Generation and persona consistency in large language models, arXiv preprint arXiv:2406.00627, Jun. 2024.
14. T. Joseph and H. K. Male, Exploring the synergy of grammar-aware prompt engineering and formal methods for mitigating hallucinations in LLMs, East African Journal of Information Technology, vol. 7, no. 1, pp. 188–201, Aug. 2024.
15. Z. Chen, A polishing model for machine-generated ancient Chinese poetry, Neural Processing Letters, vol. 56, article 77, Mar. 2024.
16. K. Ji, Enhancing persona consistency for LLMs' role-playing, in Findings of the Association for Computational Linguistics (ACL), 2025.
17. D. Liu, A Persona-Aware LLM-Enhanced Framework for Multi-Session Personalized Dialogue Generation (PALACE), in Findings of the Association for Computational Linguistics (ACL), 2025.
18. B. Li, A systematic review of empirical generative AI research in language learning and teaching, Computers & Education: Artificial Intelligence, vol. 6, no. 100215, 2025.
19. R. He, Generative artificial intelligence: A historical perspective, National Science Review, vol. 12, no. 5, nwaf050, 2025.
20. G. Schryen, Exploring the scope of generative AI in literature review, Electronic Markets, 2025.
21. A. Adel, Can generative AI reliably synthesise literature? Exploring the accuracy and risks of ChatGPT, AI & Society, 2025.
22. J. Zheng, M. Wang, and K. Ren, When 'A Helpful Assistant' Is Not Really Helpful: Personas in System Prompts Do Not Improve Performance in LLMs, arXiv preprint arXiv:2311.10054, 2023.
23. J. Kim, N. Yang, and K. Jung, Persona is a double-edged sword: Mitigating the negative impact of role-playing prompts in LLMs, arXiv preprint arXiv:2408.08631, 2024.
24. F. A. Tan, L. Zhao, and P. Huang, PHAnToM: Persona-based prompting has an effect on Theory-of-Mind reasoning in LLMs, arXiv preprint arXiv:2403.02246, 2024.
25. The Authors Guild, Legal tussle between authors and AI: Class action lawsuit against OpenAI for unauthorized use of books, Vanity Fair, Oct. 2023.