

# Dialogue Agents with Literary Character Personality Traits

Junzhe Zhao  
Newcastle University  
zhaojunzhe\_bit@163.com

Huizhi Liang  
Newcastle University  
huizhi.liang@newcastle.ac.uk

Nicolay Rusnachenko  
Newcastle University  
nicolay.rusnachenko@newcastle.ac.uk

**Abstract**—To enhance the engagement of chatbots, we need to imbue them with unique personalities and speech patterns. However, creating a high-quality conversational training dataset for this task can be time-consuming and labor-intensive. To address this, we propose an automated system that uses character dialogues from literary works. We used the Chinese classic, 《红楼梦》 (*Dream of the Red Chamber*), as our data source. Our system efficiently extracts dialogues and personality traits from the book, creates a personality map for each character, and generates responses that reflect these traits. It allows us to utilize the distinct linguistic styles that authors assign to characters with different personalities in classic books, making chatbots more human-like. Our evaluations demonstrate the efficacy of our system, with chatbots possessing personality traits proving more engaging than those lacking them. For transparency, we have made all our datasets and models available<sup>1</sup>.

**Index Terms**—Chatbots, Dialogue Dataset, Automated Extraction from Literary Works

## I. INTRODUCTION

Chatbots, computer programs engineered to simulate human-like interactions, can be categorized into question-answering (QA), task-oriented, or social chatbots, depending on the tasks they are designed to fulfill [1]. OpenAI’s ChatGPT, a prominent example of a QA bot, is so adaptable that it has been considered a viable substitute for search engines [2]. In contrast, task-oriented bots execute specific user tasks, from booking movie tickets to playing music or setting alarms. Social chatbots are designed to interact more emotionally with humans than perform specific tasks. For instance, individuals with autism often use social chatbots for relaxed, stress-free conversations to help alleviate their symptoms [3]. However, creating social chatbots presents significant challenges, mainly because they cannot rely on predefined responses. Fortunately, advancements in deep learning techniques, especially in neural machine translation [4], now allow chatbots to generate conversations and responses that resemble human discourse in naturalness and fluency. However, the need for chatbots to provide more consistent responses during user engagement remains a central focus in chatbot research [5].

Researchers have begun incorporating human personality traits into social bots to augment consistency, enrich the user experience and engagement, and make them more stylized and emotionally expressive [6]. Dialogue agents can maintain a consistent style across multiple rounds of interaction between

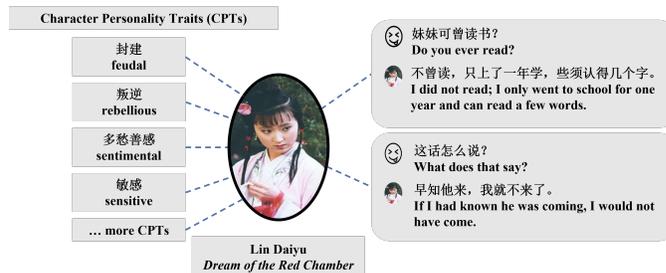


Fig. 1: Example of the character Lin Daiyu’s personality traits and associated dialogues. On the left, Lin Daiyu’s personality traits are displayed. On the right, representative conversations that reflect these personality traits are presented.

users and stylized chatbots. High-quality, personalized dialogue datasets are crucial for enhancing these generative models. Nevertheless, compiling dialogue datasets with personality traits takes time and effort. It has led some researchers to enlist volunteers to generate data in controlled environments [7]. Additional dialogue datasets featuring personality traits have been introduced into digital humanities to train dialogue agents and increase the diversity of stylized chatbots. These sources include dialogues from television [8] and film characters [9]. However, dialogues in literary books, a rich source of human personality traits, have not yet been adequately explored.

In literature, authors use dialogue to portray vibrant characters and engage readers [10]. Classic literature characters often encapsulate typical human personality traits [11], and their dialogues frequently reflect these traits. For example, Hamlet from Shakespeare’s play of the same name is depicted as *philosophical*, *contemplative*, and *indecisive* in his famous soliloquy, “To be, or not to be - that is the question.” Similarly, Lin Daiyu (林黛玉) from 《红楼梦》 (*Dream of the Red Chamber*) [12] says, “你心中自然有妹妹, 只是见了姐姐, 就把妹妹忘了 (You have got me in your heart, but you have seen other sisters and forgotten me)”, which reflects her personality traits of *sensitivity*, *straightforwardness*, and *skepticism*. Classic literary characters often stimulate a wealth of critical analysis and discourse. For instance, *Dream of the Red Chamber* has inspired numerous academic works and discussions among general readers, leading to the formation of an entire field of study known as Redology (红学), with countless books and articles published on this area. Each

<sup>1</sup>Project Repository: [https://github.com/SuperEDG/Honglouloumeng\\_Project](https://github.com/SuperEDG/Honglouloumeng_Project)

character in this book has distinct personality traits, and a key focus within Redology is the analysis of these traits. This wealth of scholarly material could serve as a vital data source for profiling character personality traits.

In conclusion, the dialogues of literary characters and the corresponding literature reviews and research articles could provide an attractive, ready-made, and high-quality raw data source for creating dialogue agents featuring human personality traits. In this paper, we propose a comprehensive framework to compile training datasets and train dialogue agents with literary character personality traits. This framework is designed to generate three datasets: a Chinese Personality Traits Lexicon, a "Literary Character Personality Map" dataset for each selected character, and a conversational dataset with named characters. The framework also leverages theories that correlate personality traits with language styles [13] to train generative models for dialogue agents with character personality traits. For example, Figure 1 illustrates Lin Daiyu's personality traits (i.e., rebellious, sentimental, sensitive) and her corresponding conversation style. We encode characters' names as profiles and integrate them into a memory-enhanced neural network to generate a dialogue agent that can mimic the character's linguistic style. This proposed framework is universally applicable as it can be utilized for many other Chinese modern and contemporary literary books and those of other languages, provided there are equivalent lexicons of personality traits and conversation extraction methods in those languages. This paper uses the terms "dialogue" and "conversation" interchangeably. Our main contributions are as follows:

- (1) We propose a comprehensive framework to automatically construct new training datasets for dialogue agents with personality traits, using literary texts as a basis. This process requires minimal human annotation. We provide pre-trained weights for others to create datasets from other literary books.
- (2) We compile a training dataset for dialogue agents based on the renowned Chinese literary masterpiece, the *Dream of the Red Chamber*. This dataset includes a Chinese Personality Traits Lexicon, a Character Conversation dataset, and a Character Personality Map dataset.
- (3) We develop dialogue agents with personality traits based on the constructed training datasets. These agents employ generative models to create dialogues in the linguistic styles of various characters, each exhibiting different personality traits in the book.
- (4) We conduct empirical evaluations and comparative experiments on both the framework and the developed dialogue agents, with and without integrating personality traits.

## II. RELATED WORK

ELIZA marked the beginning of dialogue systems, heavily relying on hand coding. Most early dialogue systems employed keyword-matching techniques to select the appropriate responses from a predefined database, precisely aiming to characterize the user's intent [14]. Specific dialogue systems

implemented Markov decision processes due to their capability to optimize model aspects based on user goals and intent [15]. These systems were primarily utilized in vertical applications, such as chatbot services for e-commerce platforms [16], but often lacked chat settings, charisma, and individuality.

Deep learning techniques like recurrent neural networks (RNN) and long short-term memory (LSTM) have ushered in a new era for AI dialogue systems. End-to-end deep learning-based methods have gained traction, particularly in generating conversations using sequence-to-sequence (Seq2seq) approaches [17], [18]. To counter the shortcomings of preceding methods like CNN and RNN, Transformers were proposed [19], progressively becoming the foundational model for conversational agent generation. Transfer learning has also been introduced to dialogue generation tasks, such as BERT (Bidirectional Encoder Representations from Transformers) [20] and GPT-2 [21]. Trained on vast data, these models offer generalized knowledge about language, which proves indispensable for downstream tasks like classification and generation.

A primary challenge for chatbots is to ensure more consistent responses during user interactions, an active area of research for social chatbots [5]. There has been a recent inclination towards integrating personality trait profiles into social bots. These profiles, once integrated and stored in memory-enhanced neural networks, enable the generation of more engaging responses than those produced by character-free models [22], [23]. Dialogue agents can maintain a consistent style across multiple communication rounds between users and stylized chatbots.

Collecting dialogue datasets imbued with personality traits is both labor-intensive and time-consuming. To mitigate this issue, researchers have utilized various dialogue datasets [6], added human feature attributes (e.g., occupation, hobbies) to the conversation dataset [7], trained the model using dialogue sets with gender-age region features [24], and introduced dialogue datasets from TV [8] and movie characters [9].

Despite the substantial progress in chatbot technology, the potential of utilizing literary dialogues for enhancing chatbot personality traits still needs to be explored. Our work addresses this gap by proposing a comprehensive framework that employs dialogues from classic literature to train dialogue agents with personality traits. This approach augments the chatbots' conversational style richness and provides a more efficient method of constructing dialogue datasets for training purposes.

## III. THE PROPOSED FRAMEWORK

In this study, we introduce a novel, four-component framework aimed at the creation of *Dialogue Agents with Literary Character Personality Traits* (refer to Figure 2). The proposed framework comprises the following components:

- Component 1: *Chinese Personality Traits Lexicon*. This component creates an expandable lexicon of Chinese personality traits, including various words and idioms that describe specific traits. This lexicon allows us to measure

character personality traits and choose characters based on set trait criteria.

- Component 2: *Literary Character Dialogue Extraction*. This module meticulously extracts dialogues and identifies the corresponding character names. These gathered dialogues are then used as input data to build dialogue agents.
- Component 3: *Literary Character Personality Map*. This map sorts characters into different personality groups, using specific words or idioms that show personality traits. Its main goal is to find characters with different personalities in the collected dialogue. It helps us create a high-quality dialogue dataset. This map is readily interpretable for humans.
- Component 4: *Literary Character Language Style Generation*. This component focuses on generating conversational responses that accurately mirror the language style and personality traits distinctive to a specific character.

Subsequent subsections will delve into a detailed description of each component of the proposed framework.

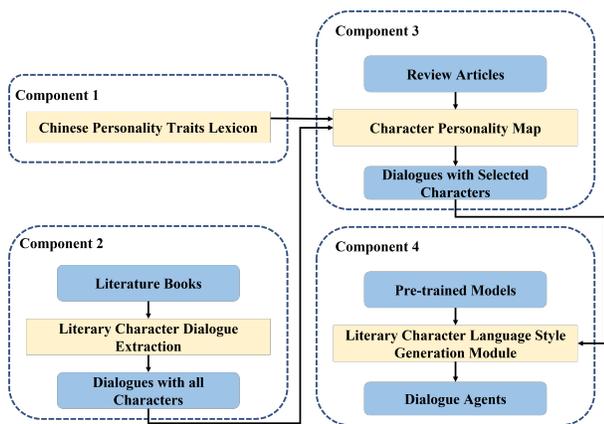


Fig. 2: Illustration of the proposed framework for Dialogue Agents with Literary Character Personality Traits.

### A. Chinese Personality Traits Lexicon

Personality, defined by persistent attitudes and consistent behaviors in response to one’s environment, significantly influences an individual’s actions and responses, including their linguistic habits and styles [13], [25], [26]. Authors often employ these stylistic variations to depict characters with differing personalities [27]. As such, a character’s linguistic style in a novel can be a powerful instrument for personality analysis. Therefore, a comprehensive lexicon must accurately capture and depict these unique personality-induced language styles.

In this work, we utilize the Simplified Chinese C-LIWC lexicon [28], the Emotional Vocabulary Ontology from the Information Retrieval Research Lab at the Dalian University of Technology [29], and the praise and criticism dictionary from Tsinghua University [30] as sources for personality indicative words. Linguistic experts manually filtered these

lexicons, resulting in a *Chinese Personality Traits Lexicon*<sup>2</sup>, which includes 10,234 words. This lexicon consists of single-character words, idioms of two-to-four characters, and phrases of more than four characters. Words and phrases that are four characters or fewer make up 99.48% of the lexicon. This lexicon is a foundation for Chinese personality indicative words and can be expanded and refined.

### B. Literary Character Dialogue Extraction

In dialogue agent training, constructing a high-quality dataset is critical. This section explores the automatic extraction of character dialogues from literary books, a task often complicated by narrative intricacies [31]. We chose 《红楼梦》(Dream of the Red Chamber)<sup>3</sup> for our analysis due to its vivid character description and linguistic similarity to Mandarin Chinese. As a landmark Qing Dynasty literature piece, it offers a linguistic similarity to modern Mandarin Chinese, bolstered by the extensive existing research literature.

Our methodology is composed of two stages. Initially, we extract dialogues and speaker names using double inverted commas (“ ”), a prevalent marker of dialogues in modern Chinese writing. By filtering sentences with these markers, we can accurately extract dialogues that follow the structure 名字 (*speaker/character name*): : ”对话内容 (*dialogue sentence*)。 ” In this initial phase, we successfully gathered 9855 dialogues, of which 1559 were manually annotated. However, given the limited size of our annotated data, we applied Long-Gang Pang’s data augmentation method<sup>4</sup>. This method involves randomly replacing dialogue words while excluding the speaker’s name. For instance, a sentence like ‘贾宝玉笑道: (*Jia Baoyu laughed:*)’ could be transformed into ‘林黛玉笑道: (*Lin Daiyu laughed:*)’, thereby effectively creating new dialogue samples. By applying this technique, we added 500,000 dialogue samples, which notably enhanced the diversity of our training data.

In the second stage, we isolate speaker names from dialogues. This task is challenging due to the many characters in literary books. We treat it as a Named Entity Recognition (NER) task, where a portion of the dialogues and corresponding speaker names are manually labeled according to the IOB (inside, outside, beginning) tagging format. For instance, in the dialogue 何为金陵十二钗正册 from 宝玉拣自己的家乡封条看, 只见那边厨上封条上大书, 宝玉问道: “何为金陵十二钗正册?”, the character name 宝玉 (**Baoyu**) is identified, with “宝”, “玉” tagged as “**B-name**”, “**I-name**”, and the rest as “**O**”. We utilized the BERT-BiLSTM-CRF model [32] in a supervised learning context for predictions and speaker extraction. The model’s high accuracy automates speaker extraction from unlabeled dialogues, significantly reducing the manual annotation workload. Importantly, this trained model can be applied to dialogue extraction in other literary books and

<sup>2</sup>Data source: [https://github.com/SuperEDG/Honglouloumeng\\_Project/blob/main/data/processed/chineseTraits.json](https://github.com/SuperEDG/Honglouloumeng_Project/blob/main/data/processed/chineseTraits.json)

<sup>3</sup>Data source: [https://github.com/SuperEDG/Honglouloumeng\\_Project/blob/main/data/resources/Honglouloumeng.txt](https://github.com/SuperEDG/Honglouloumeng_Project/blob/main/data/resources/Honglouloumeng.txt)

<sup>4</sup>Method Code: <https://gitlab.com/snowwhitiger/speakerextraction.git>

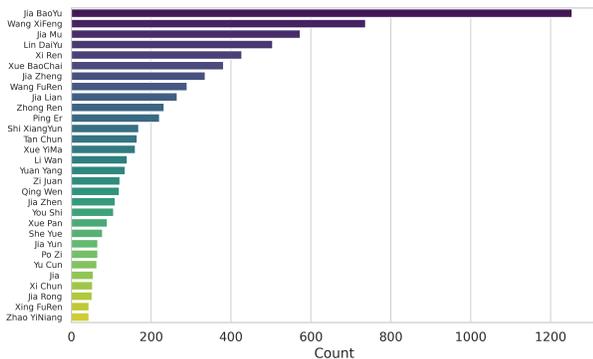


Fig. 3: Top 30 characters by dialogue count in the extracted dataset.

improved with additional annotated data, thereby expanding its scope of use. Our model displayed exceptional performance after five training rounds, with accuracy approximating 100% when preserved to four decimal places on a substantial test dataset. Moreover, it exhibited robust generalization capabilities by accurately identifying speaker names in randomly selected novel dialogues.

Our process resulted in the extraction of 685 characters and 9855 dialogues<sup>5</sup>. Fig 3 illustrates the extracted set’s top thirty characters ranked by dialogue count. The character with the highest number of dialogues, Jia Baoyu, features in 1254 dialogues. Considering the volume of dialogues and the characters’ importance in advancing the storyline, we selected 24 characters. It includes primary characters like **Jia Baoyu**, **Lin Daiyu**, and **Xue Baochai**, and excludes ambiguous designations such as “人 (person)”. Comprehensive insights into the evaluation process can be found in Section IV-A.

### C. Character Personality Map

To create a diversified dialogue dataset with diversified personality traits, we can select representative characters with distinct personality traits. These unique traits shape their language styles, making them invaluable for building personalized generative language models. To address this challenge, we have developed a character personality map. This map serves the dual purpose of grouping characters with similar personalities and separating those with different personality traits. It allows us to understand the structured and objective relationship between personality and linguistic style. We draw inspiration from the well-established Big Five Personality Traits [33] to classify the characters. We categorize the character groups into  $n$  ( $n$  can be any number, in this paper, we set  $n = 5$ ) distinct personality types. This section will focus on the construction process of the character personality map, which is an effective tool for quantitatively selecting characters and minimizing subjective analysis bias. We have already successfully extracted dialogues from various characters using the methodology outlined in the previous sections. The next

step is to deploy the character personality map to create our dataset.

Although scholars may hold differing perspectives on a character’s personality traits, these analyses typically share common elements. For instance, **Lin Daiyu**’s melancholic and sensitive disposition is a trait widely recognized among researchers. These commonly highlighted personality trait descriptions can function as fundamental features for quantifying and describing a character’s personality. In addition, they lend a human-like and understandable quality to our analysis. We collected numerous character analyses and commentary articles from various researchers and authors to construct the character personality map. Each of the selected 24 characters has a minimum of five commentary articles about them<sup>6</sup>. In total, we gathered 140 commentary articles, amounting to a word count of 242,648, averaging around 1,734 words per article.

Given the significant difference in the number of analysis articles and the popularity of characters, we needed a standardized approach to distilling personality traits. Therefore, we accumulated a collection of review articles and calculated the frequency of personality trait phrases, facilitating identifying and quantifying each trait phrase. For text pre-processing, we utilized Jieba [34], which matched the segmented words from the commentary with our Chinese Personality Traits Lexicon. Subsequently, we developed a comprehensive data table that records the occurrence frequency of each character’s traits, treating unobserved traits as zero. We defined the top ten most frequently occurring traits as the core personality traits for each character. These traits were then mapped onto a scale ranging from 0 to 5, with 0 signifying an unobserved trait and 5 indicating a highly prevalent trait. This resulted in the Character Personality Map, a user-friendly visualization tool that displays the primary personality traits of each character.

For illustrative purposes, Figure 4 introduces a Character Personality Map, underscoring distinct personality contrasts between Lin Daiyu, Xue Baochai, Wang Xifeng, and Shi Xiangyun. In the map, each row designates a different character, with columns representing diverse personality traits. The marked color and size signify the trait’s prevalence in character. For example, considering Lin Daiyu in the first row and Xue Baochai in the second, Lin Daiyu’s ‘feudal’ and ‘rebellious’ traits reach the peak level of 5, showing her distinctive traits. On the other hand, Xue Baochai has a ‘feudal’ trait of 4 and a ‘ladylike’ trait of 2, suggesting she significantly differs from Lin Daiyu.

The purpose of constructing the character personality map is to optimize our dialogue dataset by selecting representative characters to capture a broader range of linguistic styles. We adopted two methods to group characters into small groups: simulated annealing genetic algorithm and unsupervised clustering. The simulated annealing genetic algorithm, initially proposed by [35], is used in our work to quantify

<sup>5</sup>Data source: [https://github.com/SuperEDG/Hongloumeng\\_Project/blob/main/data/processed/Hongloumeng.json](https://github.com/SuperEDG/Hongloumeng_Project/blob/main/data/processed/Hongloumeng.json)

<sup>6</sup>Data source: [https://github.com/SuperEDG/Hongloumeng\\_Project/blob/main/data/processed/reviews.json](https://github.com/SuperEDG/Hongloumeng_Project/blob/main/data/processed/reviews.json)

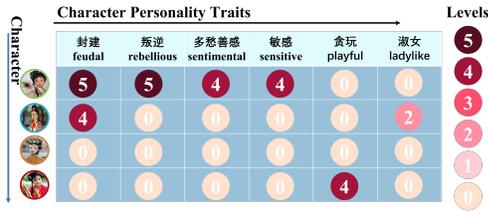


Fig. 4: Example *Character Personality Map*. Character is the vertical coordinate, and personality trait is the horizontal coordinate.

the personality differences between characters. Utilizing our personality map, we measure these differences using Euclidean distance. This approach optimizes diversity over multiple iterations, selecting characters with unique personality traits. For unsupervised clustering, we assessed three algorithms: K-means [36], DBSCAN (Density-Based Spatial Clustering of Applications with Noise) [37], and BIRCH (Balanced Iterative Reducing and Clustering Using Hierarchies) [38]. Eventually, we selected characters with the most dialogue data from the five groups generated by the K-means algorithm.

Constructing a character personality map also helps relieve the challenge of insufficient data for some characters by enabling them to leverage dialogues from other characters within the same group. Subsequently, we collated multiple sub-datasets based on different algorithms. Each sub-dataset was used to train dialogue agents, enabling them to generate responses reflecting corresponding personality traits, which were then contrasted and tested—a method we will elaborate on in the following sub-section.

#### D. Literary Character Language Style Generation Module

From the comparison in the previous section, we employed unsupervised clustering to select characters for the dialogue dataset. We chose one character from each of the five personality clusters with the most extensive corpus, structuring our dialogue dataset around five primary characters. We used BoB (BERT over BERT) [39] and GPT-2 as our base models for generating dialogue responses, further refining them for personalization by adding a unique character tag as a prefix to each query, guiding the model in generating a personalized response.

Our model, following the architecture proposed by [24] and [22], consists of three types of embeddings: word, positional, and segmentation. A token can fall into three segmentation categories: 1) Character Name, 2) Query Sentences, or 3) Response Sentences. We used an Encoder-Decoder structure based on transformers [18] for response generation. Figure 5 shows that character names are integrated with the dialogue history.

The training data is divided into three components: the character name, the query sentence, and the response. For example, the character name might be 林黛玉 (Lin Daiyu), and the query sentence could be “姐姐读过书吗？” (“Has sister read books?”). The response might be “我没读过书，

只上过一年学，认识几个字” (“I have not read, I have only attended school for a year, I know a few words.”) The query and response can span multiple rounds of dialogue. The generative model absorbs phrases representing character names and query sentences to generate response sentences.

## IV. EXPERIMENTS

In this section, we discuss the experimental setup and the evaluation of the major components of the proposed framework.

### A. Evaluation of Literary Character Dialogue Extraction

As outlined in Section III-B, punctuation provides an efficient cue for extracting dialogue content. Nonetheless, identifying the speaker’s name within dialogues poses a significant challenge due to the potential variability in sentence placement. We approach this issue as a Named Entity Recognition (NER) task and utilize widely accepted evaluation metrics such as Accuracy, Precision, and Recall to assess the performance of our predictive models.

In our experiment, we started with an initial dataset comprising 1559 entries. We expanded this dataset to include 500,000 entries using a data augmentation method [40]. This method involved substituting the ‘BI’ entities in one sentence with those from another sentence to increase the diversity of the sentences. We applied the BERT-BiLSTM-CRF model to enhance our predictions in a supervised learning setting. The dataset was tokenized using the pre-trained BERT-WWM-Chinese model [41]. We subsequently fine-tuned the BERT-BiLSTM-CRF model and evaluated its performance using the annotated dataset, with 70% allocated for training, 15% for validation, and 15% for testing. The fine-tuned model achieved perfect scores of 100% for Accuracy, Precision, and Recall on the test set. Lastly, we used the fine-tuned BERT-BiLSTM-CRF model to extract speaker names from the remaining 8296 unlabeled dialogue entries autonomously.

### B. Evaluation of Character Personality Map

The total number of personality trait phrases varies for each character. The highest number of occurrences for a single character is 12,522, while the lowest is 1,394. To evaluate the effectiveness of unsupervised methods in character classification, particularly for those characters exhibiting similar personality traits, we utilized evaluation metrics such as the Silhouette Score (SI) [42], Calinski-Harabasz index (CH) [43], and Davies-Bouldin Index (DBI) [44]. Guided by the prototypical Big Five personality theory discussed in Section III-C, we set the cluster count to five. Among the resulting performances, K-Means proved superior (SI: 0.75; CH: 96.38; DBI: 0.33), with higher SI and CH scores indicating better clustering results and a lower DBI score showing better results. It was followed closely by DBSCAN (SI: 0.73; CH: 84.68; DBI: 0.43) and BIRCH, ranking least favorably (SI: 0.55; CH: 16.11; DBI: 0.85). Consequently, we employed the K-Means clustering results to group characters exhibiting similar personality traits, as illustrated through T-SNE visualization [45] in Figure 6.

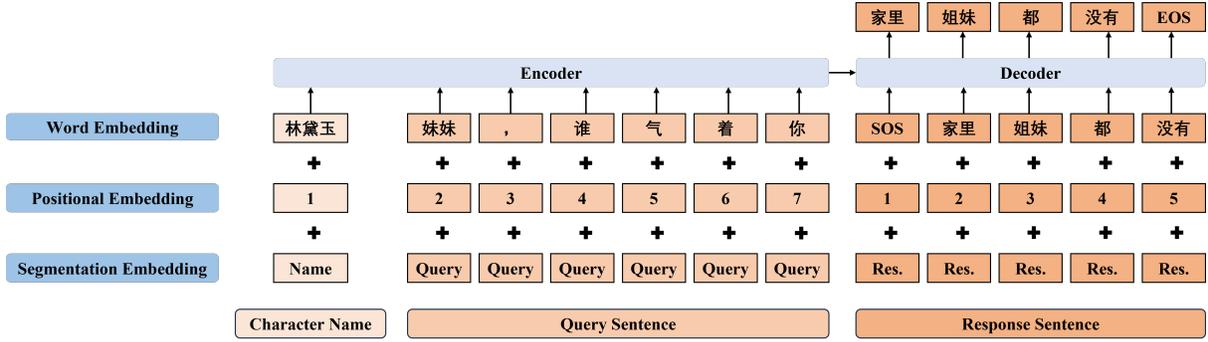


Fig. 5: An example of the encoding and decoding process for the Literary Character Language Style Generation Module.

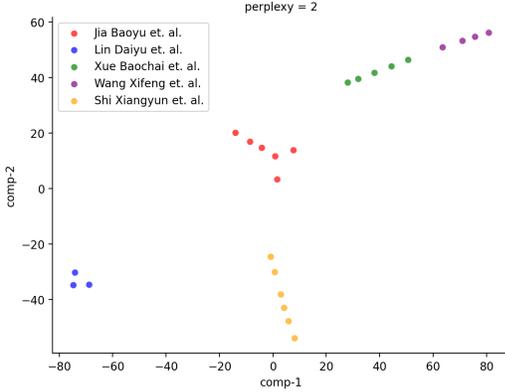


Fig. 6: T-SNE visualization of the character clustering space generated by *Character Personality Map* (CPM) with result of K-Means clustering

Cluster	Character Names
1	Lin Daiyu, Qing Wen, Miao Yu
2	Xue Baochai, Li Wan, Jia She, Wang Furen, Xing Furen
3	Shi Xiangyun, Xiang Ling, You Erjie, You Sanjie, Jia Tanchun, Jia Qiaojie
4	Wang Xifeng, Jia Mu, Jia Zheng, Jia Zhen
5	Jia Baoyu, Jia Yingchun, Jia Yuanchun, Jia Xichun, Xi Ren, Qin Keqing

TABLE I: Results of K-Means clustering approach for *character personality traits* (CPT) embedding method

Table I illustrates the results of K-Means clustering approach for CPT embedding method. The result of clustering is largely align with reader impressions, assigning the book’s five most distinctive characters (Lin Daiyu, Xue Baochai, Shi Xiangyun, Wang Xifeng, Jia Baoyu) into separate groups, consistent with the legend in Figure 6. These findings highlight the effectiveness of the clustering approach, despite minor anomalies where characters with pronounced personality differences are classified within the same group. The selection outcome from the simulated annealing algorithm included Lin Daiyu, Jia Tanchun, Wang Xifeng, Xue Baochai, and Jia Baoyu. Only one character differed from the unsupervised clustering results, with Jia Tanchun replacing Shi Xiangyun.

### C. Evaluation of Literary Character Language Style Generation Module

To assess the effectiveness of personality-imbued generative models, we compared dialogue agents with and without personality traits. We consider two generative models:

- **GPT2-Chinese** [46] is the Chinese adaptation of the GPT-2 training code, which employs a BERT tokenizer. This version is implemented using the Transformers library API from HuggingFace, and supports a comprehensive Chinese training corpus.
- **BoB** (BERT over BERT) [39] is an adaptation of BERT, fine-tuned for character dialogue agent tasks.

As illustrated in Figure 7, model inputs follow the structure “character name + query + response” as outlined in Section III-D. To investigate the affection of personality traits, as well as the discrepancies in configurations of various aspects in the automated creation of the dataset, we aim to identify the most efficient methodology. Towards this goal, we experiment with five distinct sets of dialogue-based training datasets:

- 1) Ch-A (Characters-all): is a non-filtered dialogue dataset of all characters; the character name is set to “All Characters.”
- 2) Ch-R1 (Characters-random1): is a randomly selected with Wang Xifeng and Xue Baochai set as constants (Xue Baochai, Wang Xifeng, Jia Tanchun, Jia Xichun, Jia Yinchun)
- 3) Ch-R2 (Characters-random2): is a randomly selected with Jia Baoyu and Lin Daiyu set as constants (Jia Baoyu, Lin Daiyu, Xing Furen, You Erjie, You Sanjie)
- 4) Ch-SA (Characters-SA): comprises the five characters selected through the simulated annealing algorithm (Wang Xifeng, Jia Baoyu, Xue Baochai, Jia Tanchun, Lin Daiyu)
- 5) Ch-KM (Characters-K-Means): includes the characters with the most dialogue from each group, segmented via K-Means (Wang Xifeng, Jia Baoyu, Xue Baochai, Shi Xiangyun, Lin Daiyu)

We employed five-fold cross-validation for partitioning the test set (10%), validation set (10%), and training set (80%). This division was character-based to avoid over-representing a single character’s dialogues. Base hyperparameters for both generative models were set at 50 epochs and a batch size of

<p><b>Character Name:</b> "林黛玉" (Lin Daiyu)</p> <p><b>Query:</b> "妹妹可曾读书?" (Has my younger sister ever read books?)</p> <p><b>Response:</b> "不曾读, 只上了一年学, 些许认得几个字。" (She hasn't read any, only attended school for a year and learned a few characters.)</p> <p><b>Query:</b> "妹妹尊名是哪两个字?" (What are the two characters in my sister's honorable name?)</p> <p><b>Response:</b> "无字。" (She has no characters.)</p>
<p><b>Character Name:</b> "史湘云" (Shi Xiangyun)</p> <p><b>Query:</b> "好妹妹, 替我梳上头罢。" (Dear sister, please help me with my hair.)</p> <p><b>Response:</b> "这可不能了。" (That's not possible.)</p>

Fig. 7: An example of model input and output for training.

8. We utilized automatic evaluation metrics to gauge generative model performance: perplexity of fixed-length models<sup>7</sup> (PPL) [47], F1-score, BLEU-1 (Bilingual Evaluation Understudy) [48], and Distinct-1 (dist1) / Distinct-2 (dist2) [49].

As indicated in Table II, integrating personality traits into datasets improved all results for both baseline models. Incorporating personalities into generative models as profiles can significantly enhance the performance of dialogue agents. Models, trained on the Ch-KM set demonstrated the best performance. Specifically, performance of the latter models is become improved by 26% by BLEU-1 metric compared to the one trained on dataset without personality traits. Table III displays examples of responses from the fifth set of the GPT2-Chinese, simulating the language styles of five distinctly personality characters.

Dataset	Metrics				
	PPL	F1-score	BLEU-1	dist1	dist2
<i>BoB</i>					
Ch-R1	117.25	0.0167	0.0426	0.0028	0.0060
Ch-R2	105.06	0.0403	0.0536	0.0023	0.0098
Ch-A	109.14	0.0277	0.0741	0.0071	0.0198
Ch-SA	85.84	0.0971	0.0891	0.0064	0.0318
Ch-KM	82.88	0.0994	0.0890	0.0108	0.0417
<i>GPT2-Chinese-421M</i>					
Ch-R1	117.04	0.0198	0.0507	0.0089	0.0114
Ch-R2	104.42	0.0494	0.0632	0.0097	0.0194
Ch-A	109.07	0.0347	0.0743	0.0093	0.0214
Ch-SA	85.49	0.1039	0.0897	0.0155	0.0384
<b>Ch-KM</b>	<b>82.32</b>	<b>0.1047</b>	<b>0.0943</b>	<b>0.0156</b>	<b>0.0420</b>

TABLE II: Comparison of different character selection methods based on various evaluation metrics

## V. CONCLUSION AND FUTURE WORK

This paper introduces a comprehensive framework for generating dialogue agents with personality traits, leveraging

<sup>7</sup>The lower the confusion, the better the prediction effect

Input	妹妹,又是谁气着你了? Darling, who's mad at you again?
Lin Daiyu	谁生什么气? Who's mad at what?
Xue Baochai	我又怎么了? You talked me into it again.
Shi Xiangyun	越发奇了, 也犯不上生气。 The more curious it is, the less it is necessary to be angry.
Wang Xifeng	凭他什么样儿的, 我也要见一见! No matter what he is like, I want to see him!
Jia Baoyu	家里姐姐妹妹都没有, 单我有, 我说没趣。 There are no sisters at home, only me, I say it's boring.

TABLE III: Example of using trained GPT-2 (Ch-KM) model to generate responses with different character personality traits

literature as the source. We utilize *Dream of the Red Chamber* as our dialogue source and its extensive review articles for personality trait analysis. This study underscores the importance of a high-quality dataset for building personalized generative models, proposing a method to efficiently extract high-quality dialogue datasets from literary books. We construct a training dataset founded on this framework, including a *Chinese Personality Traits Lexicon*, a Character Conversation dataset, and a *Character Personality Map* dataset. Building upon the constructed training datasets, we crafted dialogue agents enriched with personality traits. Generative models, such as BERT and GPT-2, were utilized to generate responses. Experimental results illustrate that dialogue agents with personality traits outperformed those lacking them.

The proposed framework offers substantial versatility in its potential applications. It can be directly employed in many modern and contemporary Chinese literature. Furthermore, the framework can be adapted for literature in various languages with corresponding language-specific personality trait thesauri and conversation extraction methodologies. As an immediate next step, we plan to conduct the human evaluation of dialogue generation.

In our future work, we aim to explore innovative techniques for dialogue agent development and apply this framework to more classic literature and characters across various languages. This approach will help in constructing a multilingual training corpus for dialogue agents. Our goal is to broaden this framework by creating connections between personality traits from digital humanity within the fictional virtual world and traits from humans in the real world.

## REFERENCES

- [1] J. Gao, M. Galley, and L. Li, *Neural approaches to conversational AI: Question answering, task-oriented dialogues and social chatbots*. Now Foundations and Trends, 2019.
- [2] M. U. Haque, I. Dharmadasa, Z. T. Sworna, R. N. Rajapakse, and H. Ahmad, "“ i think this is the most disruptive technology”: Exploring sentiments of chatgpt early adopters using twitter data,” *arXiv preprint arXiv:2212.05856*, 2022.

- [3] A. A. Abd-Alrazaq, M. Alajlani, A. A. Alalwan, B. M. Bewick, P. Gardner, and M. Househ, "An overview of the features of chatbots in mental health: A scoping review," *International Journal of Medical Informatics*, vol. 132, p. 103978, 2019.
- [4] N. Kalchbrenner and P. Blunsom, "Recurrent continuous translation models," in *Proceedings of the 2013 conference on empirical methods in natural language processing*, pp. 1700–1709, 2013.
- [5] I. V. Serban, R. Lowe, L. Charlin, and J. Pineau, "Generative deep neural networks for dialogue: A short review," *arXiv preprint arXiv:1611.06216*, 2016.
- [6] I. V. Serban, R. Lowe, P. Henderson, L. Charlin, and J. Pineau, "A survey of available corpora for building data-driven dialogue systems," *arXiv preprint arXiv:1512.05742*, 2015.
- [7] S. Zhang, E. Dinan, J. Urbanek, A. Szlam, D. Kiela, and J. Weston, "Personalizing dialogue agents: I have a dog, do you have pets too?," *arXiv preprint arXiv:1801.07243*, 2018.
- [8] A. W. Li, V. Jiang, S. Y. Feng, J. Sprague, W. Zhou, and J. Hoey, "Aloha: Artificial learning of human attributes for dialogue agents," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, pp. 8155–8163, 2020.
- [9] A. Sojasingarayar, "Seq2seq ai chatbot with attention mechanism," *arXiv preprint arXiv:2006.02767*, 2020.
- [10] M. Caracciolo, *Strange Narrators in Contemporary Fiction: Explorations in Readers' Engagement with Characters*. U of Nebraska Press, 2016.
- [11] N. Jacobs, *The Character of Truth: Historical Figures in Contemporary Fiction*. SIU Press, 1990.
- [12] C. Xueqin, *A dream of red mansions*. Asiapac Books Pte Ltd, 1978.
- [13] J. W. Pennebaker and L. A. King, "Linguistic styles: language use as an individual difference," *Journal of personality and social psychology*, vol. 77, no. 6, p. 1296, 1999.
- [14] S. J. Young, "Probabilistic methods in spoken–dialogue systems," *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 358, no. 1769, pp. 1389–1402, 2000.
- [15] S. Young, M. Gašić, B. Thomson, and J. D. Williams, "Pomdp-based statistical spoken dialog systems: A review," *Proceedings of the IEEE*, vol. 101, no. 5, pp. 1160–1179, 2013.
- [16] S.-P. Ma and C.-T. Ho, "Modularized and flow-based approach to chatbot design and deployment," *Journal of Information Science & Engineering*, vol. 34, no. 5, 2018.
- [17] M.-T. Luong, Q. V. Le, I. Sutskever, O. Vinyals, and L. Kaiser, "Multi-task sequence to sequence learning," *arXiv preprint arXiv:1511.06114*, 2015.
- [18] O. Vinyals and Q. Le, "A neural conversational model," *arXiv preprint arXiv:1506.05869*, 2015.
- [19] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.
- [20] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.
- [21] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever, et al., "Language models are unsupervised multitask learners," *OpenAI blog*, vol. 1, no. 8, p. 9, 2019.
- [22] S. Zhang, E. Dinan, J. Urbanek, A. Szlam, D. Kiela, and J. Weston, "Personalizing dialogue agents: I have a dog, do you have pets too?," *CoRR*, vol. abs/1801.07243, 2018.
- [23] M. Grigore, C. Rosenkranz, and J. Sutanto, "The impact of sentiment-driven feedback on knowledge reuse in online communities," *AIS Transactions on Human-Computer Interaction*, vol. 7, no. 4, pp. 212–232, 2015.
- [24] Y. Zheng, G. Chen, M. Huang, S. Liu, and X. Zhu, "Personalized dialogue generation with diversified traits," *arXiv preprint arXiv:1901.09672*, 2019.
- [25] A. H. Buss, "Personality as traits," *American psychologist*, vol. 44, no. 11, p. 1378, 1989.
- [26] F. Mairesse, M. A. Walker, M. R. Mehl, and R. K. Moore, "Using linguistic cues for the automatic recognition of personality in conversation and text," *Journal of artificial intelligence research*, vol. 30, pp. 457–500, 2007.
- [27] K. Oatley, "A taxonomy of the emotions of literary response and a theory of identification in fictional narrative," *Poetics*, vol. 23, no. 1–2, pp. 53–74, 1995.
- [28] N. Zhao, D. Jiao, S. Bai, and T. Zhu, "Evaluating the validity of simplified chinese version of liwc in detecting psychological expressions in short texts on social network services," *PLoS One*, vol. 11, no. 6, p. e0157947, 2016.
- [29] P. Y. Xu Linhong, Lin Hongfei, "The construction of emotional vocabulary ontology," *Journal of Intelligence*, vol. 27, no. 2, pp. 180–185, 2008.
- [30] J. Li and M. Sun, "Experimental study on sentiment classification of chinese review using machine learning techniques," in *2007 International Conference on Natural Language Processing and Knowledge Engineering*, pp. 393–400, IEEE, 2007.
- [31] Y. Sang, X. Mou, M. Yu, S. Yao, J. Li, and J. Stanton, "Tvshowguss: Character comprehension in stories as speaker guessing," *arXiv preprint arXiv:2204.07721*, 2022.
- [32] Z. Dai, X. Wang, P. Ni, Y. Li, G. Li, and X. Bai, "Named entity recognition using bert lstm crf for chinese electronic health records," in *2019 12th international congress on image and signal processing, biomedical engineering and informatics (cisp-bmei)*, pp. 1–5, IEEE, 2019.
- [33] D. A. Cobb-Clark and S. Schurer, "The stability of big-five personality traits," *Economics Letters*, vol. 115, no. 1, pp. 11–15, 2012.
- [34] J. Sun, "Jieba chinese word segmentation tool," 2012.
- [35] P. J. M. van Laarhoven and E. H. L. Aarts, *Simulated annealing*, pp. 7–15. Dordrecht: Springer Netherlands, 1987.
- [36] K. Krishna and M. N. Murty, "Genetic k-means algorithm," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 29, no. 3, pp. 433–439, 1999.
- [37] E. Schubert, J. Sander, M. Ester, H. P. Kriegel, and X. Xu, "DbSCAN revisited, revisited: why and how you should (still) use dbSCAN," *ACM Transactions on Database Systems (TODS)*, vol. 42, no. 3, pp. 1–21, 2017.
- [38] T. Zhang, R. Ramakrishnan, and M. Livny, "Birch: an efficient data clustering method for very large databases," *ACM sigmod record*, vol. 25, no. 2, pp. 103–114, 1996.
- [39] H. Song, Y. Wang, K. Zhang, W.-N. Zhang, and T. Liu, "Bob: Bert over bert for training persona-based dialogue models from limited personalized data," *arXiv preprint arXiv:2106.06169*, 2021.
- [40] X. Dai and H. Adel, "An analysis of simple data augmentation for named entity recognition," *arXiv preprint arXiv:2010.11683*, 2020.
- [41] Y. Cui, W. Che, T. Liu, B. Qin, S. Wang, and G. Hu, "Revisiting pre-trained models for Chinese natural language processing," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings*, (Online), pp. 657–668, Association for Computational Linguistics, Nov. 2020.
- [42] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of computational and applied mathematics*, vol. 20, pp. 53–65, 1987.
- [43] T. Caliński and J. Harabasz, "A dendrite method for cluster analysis," *Communications in Statistics-theory and Methods*, vol. 3, no. 1, pp. 1–27, 1974.
- [44] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE transactions on pattern analysis and machine intelligence*, no. 2, pp. 224–227, 1979.
- [45] L. Van der Maaten and G. Hinton, "Visualizing data using t-sne," *Journal of machine learning research*, vol. 9, no. 11, 2008.
- [46] Z. Du, "Gpt2-chinese: Tools for training gpt2 model in chinese language," *GitHub repository*, 2019.
- [47] P. F. Brown, S. A. Della Pietra, V. J. Della Pietra, J. C. Lai, and R. L. Mercer, "An estimate of an upper bound for the entropy of english," *Computational Linguistics*, vol. 18, no. 1, pp. 31–40, 1992.
- [48] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "Bleu: a method for automatic evaluation of machine translation," in *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pp. 311–318, 2002.
- [49] J. Li, M. Galley, C. Brockett, J. Gao, and B. Dolan, "A diversity-promoting objective function for neural conversation models," in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, (San Diego, California), pp. 110–119, Association for Computational Linguistics, June 2016.