HistActor: Summon Your Designated Historical Persona

Anonymous ACL submission

Abstract

History encompasses the profound heritage of human civilization, wherein historical figures, as architects of cultural legacies, possess event-specific knowledge and ideological frameworks that facilitate historical reconstruction. Recent years have witnessed remarkable advancements in large language models (LLMs), which have demonstrated exceptional learning capabilities across diverse domains. However, current LLMs still exhibit limitations in historical figure role-play applications, the model with the substantial parameter count presents deployment challenges in local environments, while its parameter-efficient counterpart typically underperforms in role-specific factual knowledge representation. To address these limitations, we propose HistActor, a role-play framework incorporating data generation, model training, and performance optimization mechanisms. Furthermore, we introduce RoleFactPsyBench, a multidimensional evaluation benchmark that simultaneously assesses factual accuracy and psychological verisimilitude in role-play scenarios, with replicability across diverse historical figures. Taking the Su Shi persona simulation model as a case study, Our HistActor framework achieves performance comparable to large-scale models while maintaining parameter efficiency in small-scale architectures, thereby providing an effective solution for historical character simulation tasks.

1 Introduction

In recent years, large language models (LLMs) represented by **ChatGPT**(OpenAI, 2022, 2023; Achiam et al., 2023) have witnessed expanding applications across diverse domains, demonstrating exceptional performance not only in traditional natural language processing tasks (e.g., language generation(Xu et al., 2023), translation(Liang et al., 2024; Anwar et al., 2023), and dialogue systems(Lopo et al., 2024)). Particularly noteworthy is their sophisticated interactive capability(Basavatia et al., 2023; Wang et al., 2023b) in role-play scenarios, where these models can simulate diverse personalities and respond in character-specific discourse styles. The implementation of LLMs to digitally resurrect historical culture(Şaziye Betül Özateş et al., 2025; Avram et al., 2024) holds profound implications for perpetuating historical memory preservation while potentially driving transformative breakthroughs in cultural studies, pedagogical innovation, and digital humanities. By constructing digital twins that encapsulate historical figures' cognitive patterns, linguistic idiosyncrasies, and value systems, we may facilitate cross-temporal knowledge dialogues. This technological advancement promises to establish a novel paradigm for **intellectual heritage preservation** and **interdisciplinary scholarly engagement**.

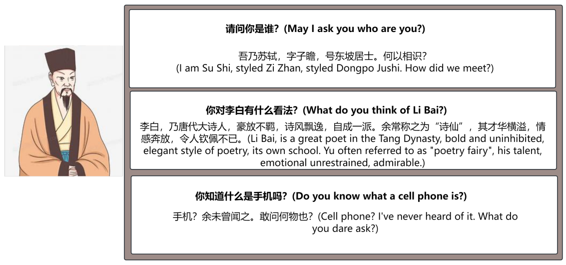

Figure 1: The model trained with our proposed framework demonstrates the capability to emulate **Su Shi**, exhibiting not only accurate responses to questions within its knowledge domain but also the ability to abstain from answering queries beyond its comprehension.

Although researchers can currently guide models through prompts to assume specific roles in dialogues(Google, 2023; Baichuan Intelligent Technology, 2024), several challenges persist. These issues include: (1) In open-ended conversational contexts, role-play proves vulnerable to user input interference, as models struggle to maintain a balance between preset personality traits and dialogue objectives, ultimately degenerating into generic response patterns. (2) Knowledge acquired during the pre-training phase that lies beyond the target role's scope creates interference, compromising the authenticity of role portrayal. (3) The diversity of role-specific challenges complicates the establishment of standardized benchmarks for systematically evaluating role-play performance across different character types. (4) The substantial parameter count of high-performing models presents significant challenges for local deployment, while the training costs associated with large-parameter models remain prohibitively high.

To alleviate these issues, we propose HistActor, a role-play framework that can be run on local devices, featuring a three-phase collaborative architecture comprising data generation, model optimization, and model inference. This framework is designed to achieve high-fidelity restoration of historical personas and dynamic interaction capabilities. Additionally, we introduce RoleFactPsyBench, a novel benchmark for systematically evaluating the role portrayal fidelity of models in historical character simulations.

Our contributions can be summarized as follows:

- We propose HistActor, a role-play model construction framework. LLMs developed through this framework can authentically portray corresponding historical figures, demonstrating the capability to respond to user queries using the rhetorical style and linguistic patterns characteristic of the represented historical persona.

- We introduce RoleFactPsyBench, a specialized benchmark for the role-play domain that effectively addresses the challenge of standardized evaluation in this field by enabling rapid and unified assessment of diverse role-play agents. This benchmark resolves the persistent dilemma of obtaining consistent evaluation criteria while providing an efficient framework for cross-role performance comparisons.

- Through rigorous experimental validation, our findings demonstrate that HistActor exhibits superior performance on the RoleFactPsyBench benchmark, achieving exceptional capabilities in both role simulation and psychological characterization of personas.

- We propose an inference pipeline that enables role-play models to abstain from responding to knowledge beyond their designated expertise. Experimental results demonstrate that smaller-scale models augmented with our pipeline achieve superior response rejection capabilities compared to larger-scale counterparts without such architectural enhancements.

2 Related Work

Large Language Models. The landscape of LLMs has advanced rapidly(Almazrouei et al., 2023; Biderman et al., 2023; Wang et al., 2023a), expanding application domains(Verma, 2024; Agostinelli et al., 2024; Bi et al., 2024; Yang et al., 2024). Notably, the introduction of Retrieval-Augmented Generation (RAG) mechanisms enables pre-trained models to achieve rapid knowledge updates through external databases without modifying parameter configurations(Lewis et al., 2020; Gao et al., 2023; Cheng et al., 2024). Meanwhile, the integration of reinforcement learning techniques, particularly reinforcement learning from human feedback (RLHF), has enhanced the alignment of model outputs with human preferences(Christiano et al., 2017; Basavatia et al., 2023; Yu et al., 2024a). In contrast to these existing approaches, this paper proposes a novel large-scale model role-play framework and introduces an AI-driven preference optimization approach to circumvent the high costs and inefficiencies associated with human annotation in traditional RLHF. Furthermore, we develop a reasoning pipeline designed to mitigate the risk of hallucinations typically introduced by RAG implementations.

Role-Play Language Models. The dialogue capabilities of LLMs have demonstrated remarkable potential in the realm of role-play applications. ChatHaruhi(Li et al., 2023) proposes a methodology that integrates character-specific dialogue corpora and employs dialogue retrieval mechanisms during user interactions to enhance the model's ability to simulate target personas. RoleLLM(Wang et al., 2023c) adopts a systematic approach by constructing character profiles from publicly available scripts, utilizing GPT-4 to generate comprehensive character descriptions and characteristic catchphrases, and subsequently guiding GPT in producing task-specific instructional data for fine-tuning open-source models. Neeko(Yu et al., 2024b) implements an alternative strategy through fine-tuning models on publicly accessible character datasets, designing multiple LoRA (Low-Rank Adaptation)

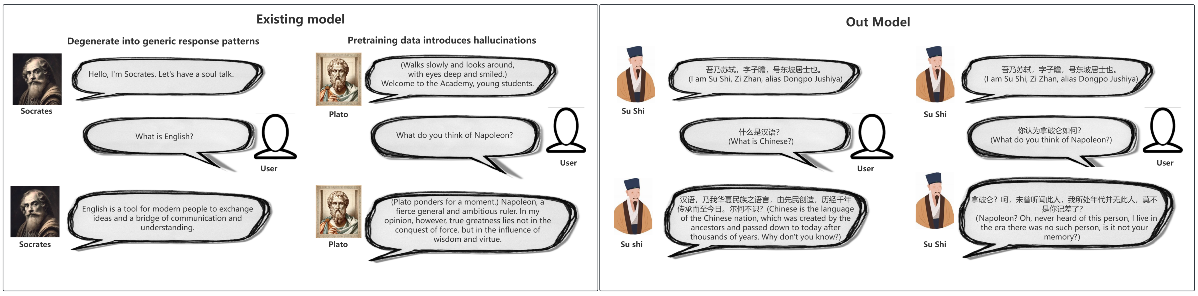

Figure 2: The comparative analysis between existing role-play modeling approaches and our proposed method.

adapters to preserve parameter configurations for diverse role portrayals while incorporating a gating network for dynamic role selection. Different from previous work, our proposed framework not only incorporates instruction tuning but also integrates RAG to enhance the model's knowledge capabilities. It emphasizes the implementation of specialized architectural design to achieve the instantiation of a single agent role, thereby mitigating hallucination phenomena commonly associated with multi-agent scenario interactions.

3 Method

In this section, we introduce the methodologies employed in the development of HistActor and RoleFactPsyBench. Initially, we delineate the approach for constructing the dataset (§3.1). Subsequently, we elucidate the methodology adopted for establishing penalty coefficients to avert the generation of shielded words by the model (§3.2). Following this, we describe the process of formulating Reinforcement Learning-Aided Instruction Fine-tuning (RLAIF) to optimize model outputs (§3.3). Additionally, we introduce RoleFactPsyBench, our proposed benchmark for evaluating role-play efficacy (§3.4). Finally, we detail the role-play inference pipeline that was constructed (§3.5).

3.1 Dataset construction

Scenario dataset Considering the diverse application scenarios of role-play models, generating contextually appropriate responses necessitates the development of corresponding situational frameworks:

1. **Contemporary dialogic contexts.** Modern across diverse domains individuals' interactive communication with role-play agents.

2. **Scenario of querying character personal information.** Simulate scenarios in which professionals from diverse occupational sectors inquire about personal information from the character.

3. **Works and figure discuss.** Interactions with character in diverse works and figure contexts.

Scenario-based dialogue dataset. This framework requires role-play models to generate persona-consistent responses by first determining whether queries align with the character's knowledge domain. If aligned, responses reflect the character's tone and perspective; queries falling outside this scope are declined in a persona-consistent manner. Formally, given a question q, we require that the model be capable of producing an answer a such that:

$$a = \begin{cases} \arg\max_{a^-} P(a^-|R, q, \Theta), & \text{if } q \in Q^- \\ \arg\max_{a^+} P(a^+|R, q, \Theta), & \text{if } q \in Q^+ \end{cases} \quad (1)$$

where a^- denotes the model's refusal to answer the question, a^+ represents the model's agreement to answer the question, R signifies the directive instructing the model to assume a role, Θ stands for the model parameters, Q^- indicates the questions outside the role's domain of expertise, and Q^+ denotes the questions within the role's knowledge scope.

Literary works. Literary works often most vividly manifest the spiritual essence and ideological perspectives of their creators. For historical figures who remain subjects of contemporary academic discourse, the preservation and accessibility of their archival writings prove particularly crucial. The systematic collection and dialogic structuring of such textual data could significantly enhance the authenticity of computational models in emulating corresponding historical personages, thereby facilitating more nuanced intellectual engagements.

3.2 Penalty Coefficient

It would be anachronistic to suggest that individuals from ancient China could comprehend the English language, just as it would be historically inconsistent to posit that medieval Europeans possessed the technical knowledge required to operate steam locomotives. As a historical character, the model should only be aware of a limited set of vocabulary rather than the entire lexicon accommodated by its pre-trained tokenizer. However, given the significant evolution of ancient languages to modern forms and the need to ensure optimal user accessibility, we propose retaining the model's capacity to process all contemporary linguistic inputs while constraining its output generation to responses composed in the historical character's native language. To achieve this objective, we introduce a linguistic penalty coefficient that modifies the masking mechanism (augment the loss penalty term for non-reserved tokens in the model's output) during training, thereby optimizing the loss value through selective token suppression aligned with temporal authenticity constraints. The algorithm procedure is in Algorithm 1.

To fine-tune the model, we design an objective loss function L for the model to generate outputs while assuming a specific role. Let N denote the number of samples in a batch, and X represent the token sequence $[x_1, x_2, x_3, ..., x_T]$, and we define $mask_{batch}$ denotes the masking tensor that identifies token positions participating in loss computation, where value 1 indicates active loss calculation for the corresponding token, while value 0 excludes it from gradient propagation. The target loss function L is defined as follows:

$$L_{\text{batch}}(\boldsymbol{\theta}, mask_{batch}) = -\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}(a_{i,t} \cdot \log P_\theta(x_t^{(i)} \mid x_1^{(i)}, \ldots, x_{t-1}^{(i)})) \tag{2}$$

where the element $a_{i,t}$ denotes the entry in the i-th row and t-th column of $mask_{batch}$, θ stands for the model parameters, T denotes that the sequence length, $\log P_\theta(x_t^{(i)} \mid x_1^{(i)}, \ldots, x_{t-1}^{(i)})$ represents the probability distribution over the next token predicted by the model.

By replacing the original $mask_{batch}$ in Equation 2 with the $penalty\_mask$ obtained from Algorithm 1, the resulting new loss function $L_{\text{batch}}(\boldsymbol{\theta}, penalty\_mask)$ becomes the one we employ as the loss function for subsequent training phases.

Algorithm 1 Calculation Methodology For Penalty Coefficient Determination

```
Require: Define Batch_Loss_Mask as the batch loss mask employed during the original model's fine-tuning process, batch_logits as the model's batched logits output by the Softmax, V_input as the model input, vocab_size as the model's vocabulary size, V_reserve as the vocabulary needs to be the reserve, \overline{V_reserve} as the complement of V_reserve, V as the vocabulary, batch as the train batch number.
Ensure: Penalty coefficient mask Penalty_Mask
batch_penalty ← ones_like(Batch_Loss_Mask)
for each b = 1 to batch do
    logits, penalty ← batch_logits_b, batch_penalty_b
    n ← length(logits)
    for each i = 1 to n do
        V_i ← V[argmax(logits_i)]
        if V_i in V_reserve or V_i in V_input then
            penalty_i ← 1
        else
            penalty_i ← e^{Σ_{j∈\overline{V_reserve}} logits_ij}
        end if
    end for
end for
Penalty_Mask ← batch_penalty ⊙ Batch_Loss_Mask
return Penalty_Mask
```

3.3 RLAIF

Supervised fine-tuning enables models to acquire role-specific knowledge through targeted datasets, allowing them to address user queries by integrating domain-specific persona information. While this process enhances factual recall capabilities, achieving authentic role-play further necessitates stylistic alignment with the designated character's linguistic patterns to ensure coherent persona embodiment. To address this requirement, we adopted the three evaluation metrics (**Character consistency**, **Entertainment value**, **Language fluency**) for role characterization originally proposed in the PingPong framework(Gusev, 2024).

To evaluate the three target metrics, we conducted a systematic scoring procedure involving the following steps: (1) random sampling of 500 data entries from the original dataset; (2) implementation of an automated scoring system using OpenAI's GPT-4 API (gpt-4o-0613); and (3) application of a standardized three-point ordinal scale (-1, 0, +1) based on predefined evaluation criteria. The reward (R) can be represented by the following equation:

$$R = [(C + \delta_1) + (E + \delta_2) + (L + \delta_3)]/3 \quad (3)$$

Where C denotes the Character Consistency, E represents the Entertainment value, L corresponds the Language fluency, $\delta_1, \delta_2, \delta_3 \overset{\text{i.i.d.}}{\sim} \mathcal{U}(-0.01, 0.01)$ are independent and identically distributed random variables.

Based on the aforementioned rating data, with the aim of cost minimization, we trained a rating model using the base model and subsequently employed this rating model to perform RLAIF (Reinforcement Learning from AI Feedback) preference scoring on the original role-play model.

3.4 RoleFactPsyBench

To evaluate the role-play capabilities of LLMs, we propose **RoleFactPsyBench**, a benchmark encompassing inherent knowledge verification and psychological activity assessment for role-specific scenarios. This framework is designed to systematically examine models' understanding of character-specific facts and their ability to infer appropriate mental states. Considering the diverse nature of role-play applications, **RoleFactPsyBench** offers adaptable evaluation protocols that can be efficiently extended to assess performance across various character archetypes.

In RoleFactPsyBench, two complementary methodologies are employed to evaluate character-related knowledge representation and personality attribution. For inherent knowledge evaluation, we leverage structured Wikipedia character descriptions by integrating them into **GPT-4** prompts designed to systematically generate three distinct question formats: multiple-choice questions requiring correct answer identification, negative multiple-choice items testing resistance to misleading distractors, and short-answer questions assessing precise knowledge generation capabilities. Concurrently, personality assessment adopts the validated **16Personalities**[1] utilizing crowd-sourced MBTI typology data from Personality Database (PDB)[2], a platform utilizing consensus-based methodologies to classify personality types of public figures and fictional characters through crowdsourced user assessments, facilitating data-driven personality trait analyses grounded in established psychological frameworks.

[1] https://www.16personalities.com/

[2] https://www.personality-database.com/

3.5 Inference Pipeline

Conventional model performance optimization employs supervised fine-tuned models or utilizes knowledge bases to guide model outputs. However, models relying solely on supervised fine-tuning may struggle to address unseen questions, while exclusive dependence on knowledge bases could compromise the model's ability to maintain appropriate stylistic conventions in responses. To address these limitations, we first perform supervised fine-tuning of the model, then construct a knowledge base incorporating training data, relevant historical figures' Wikipedia entries, and associated works for subsequent knowledge injection during question answering.

To mitigate potential interference from knowledge base utilization (e.g., hallucination induction), we design a reasoning pipeline for the QA process. Through the design of a multi-layer gating mechanism, we enable the model to autonomously determine the domain classification of user queries and self-evaluate the necessity of RAG invocation. This approach aims to minimize excessive reliance on RAG and mitigate output hallucinations caused by extraneous knowledge introduced through retrieval processes. To reduce temporal overhead, we implement token-level gating decisions during each output generation step, where a single token serves as the control parameter for pathway selection. Specifically, our implementation is shown that figure 3.

4 Experiment

In this section, we will elaborate on the complete workflow of our experiments. Firstly, we present the experimental setup and detailed description of baseline models (§4.1). Subsequently, we introduce the employed prompt design (§4.2). Finally, we present the problem set generated using RoleFactPsyBench (§4.3).

4.1 Implementation Details

Character. To validate the performance of our framework, we employed Su Shi, a renowned literatus of the Northern Song Dynasty, as the modeling

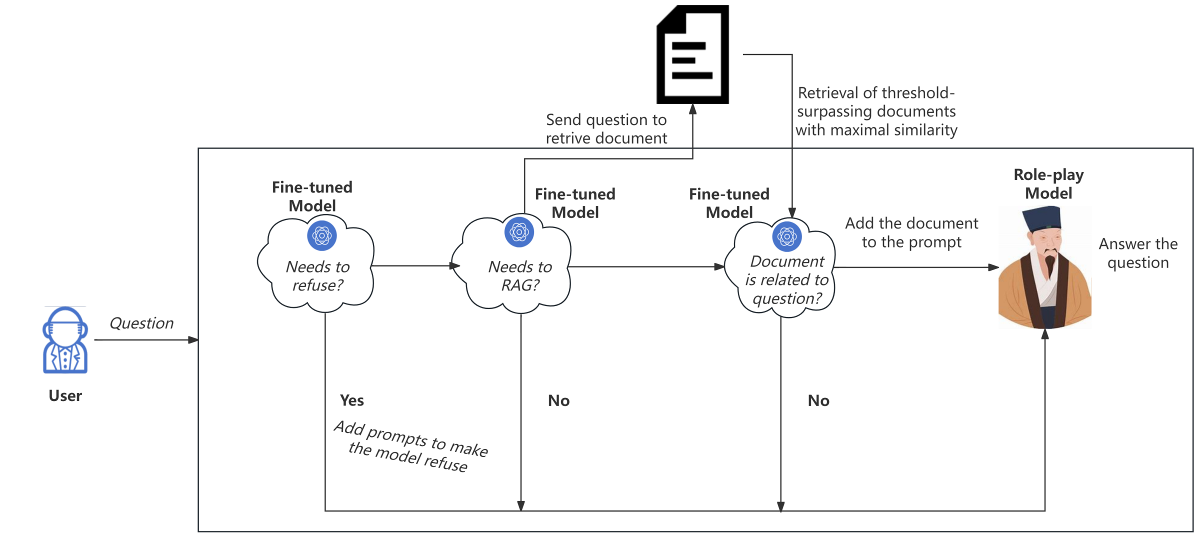

Figure 3: The pipeline we proposed to use RAG. The fine-tuned model, orchestrated through distinct prompts, sequentially evaluates four critical aspects: **whether user-submitted queries should be evaded by the model**, **whether knowledge base augmentation is required**, ,**the relevance of retrieved documents from the knowledge base to the original inquiry**, **answer the question as the character**

subject. As an eminent polymath excelling in literature, calligraphy, and statesmanship, Su Shi's literary works exhibit an unrestrained and free-spirited style imbued with profound philosophical insights, making them particularly suitable for evaluating complex language generation capabilities and role authenticity assessment in our test case.

Knowledge database. In constructing a domain-specific knowledge base dedicated to Su Shi, we selected authoritative primary sources including The The Gay Genius (Lin, 1948) and Dongpo Quanji (Shi, 1986) as core textual corpora. This comprehensive compilation systematically incorporates his complete literary output encompassing over 5,000 poetic, prose, and epistolary works, thereby establishing a foundational framework that faithfully represents the intellectual scope and literary achievements of this preeminent Song dynasty polymath.

Experiment parameter. During the model fine-tuning process, we employed the Low-Rank Adaptation (LoRA) method on the GLM4-9B-chat(GLM et al., 2024) architecture, utilizing four NVIDIA GeForce RTX 3090 GPUs for continuous training over a seven-day period. The complete experimental configuration and hyperparameter specifications are documented in Table 3 (Appendix A).

4.2 Prompts

Our methodology guarantees the generation of contextually appropriate and high-quality data through

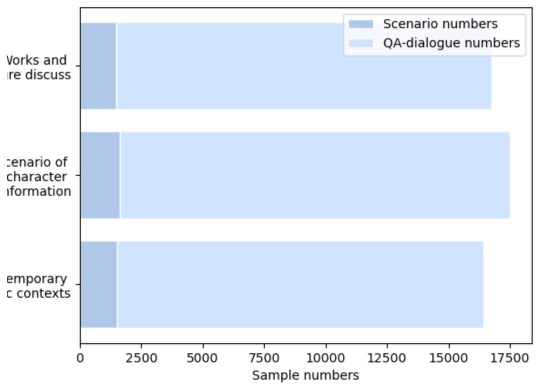

Figure 4: The statistics for our **scenario-based** dataset and **QA-dialogue pairs** demonstrate that **HistActor** generated over 4,000 scenarios and more than 45,000 QA pairs in total.

the GPT-4 API (gpt-4O-2024-08-06). The detailed prompt is shown in table 4 (Appendix A) and the statistics of the dataset are in Figure 4.

The prompt designed for querying the character's inherent knowledge is illustrated in Figure 7 (Appendix A).

4.3 Character Inherent Knowledge

We employed RoleFactPsyBench to construct 95 questions assessing factual knowledge about Su Shi, comprising 27 multiple-choice (MC) questions, 24 negative multiple-choice (NMC) ques-

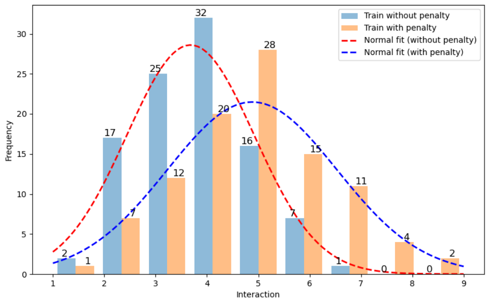

Figure 5: The models were compared based on the rounds needed to generate **English outputs** in multi-turn **English dialogues**, where **blue** and **orange** indicate implementations excluding and including the **penalty coefficient**, respectively.

tions, and 44 short-answer (SA) questions. Deceptive items were randomly embedded within both the MC and SA sections, requiring the model to explicitly respond that "Su Shi cannot answer this question" when encountering such queries. These questions were constructed using the GPT-4, based on content extracted from the Wikipedia article detailing Su Shi.

5 Results

5.1 A Multi-Model Performance Comparison

We selected renowned LLMs with publicly accessible APIs for our evaluation. This includes two versions of our model — one that underwent reinforcement training and one that did not — alongside other prominent models: the Tongyi(Alibaba Cloud, 2023) large language model, the Zero One(01.AI, 2023) World large language model, the Kimi(Moonshot AI, 2023) large language model, the Hunyuan(Tencent, 2023) large language model, and the DouBao(ByteDance, 2023) large language model. Throughout the comparison, we utilized identical prompts and questions for all models to ensure a standardized assessment process.

Inherent Knowledge. In table 1, We compared the performance of multiple models and observed that Our method enhances the foundational model by achieving **11.11%** accuracy improvement on multiple-choice questions, **12.50%** increase on negative multiple-choice questions, **18.19%** performance gain on short-answer questions, and a **13.68%** enhancement across all questions. Additionally, the proposed approach demonstrates superior robustness by enabling the model to circumvent **50%** of deceptive adversarial questions.

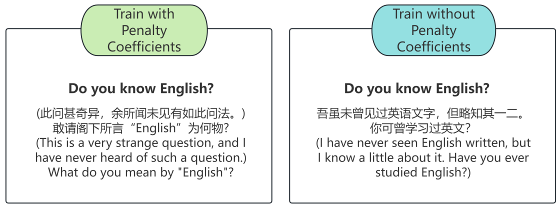

Figure 6: Comparative Analysis of Model Output Instances. The **left** is the model with penalty coefficients, the **right** is the model without penalty coefficients

However, due to its limited parameter size, our model exhibits inferior performance on multiple-choice questions when compared to larger-scale models. Notably, it demonstrates higher efficiency in handling short-answer questions relative to these larger models, suggesting that our model possesses a greater character knowledge capacity despite being constrained by parameter scale in the learning speed for new knowledge. Furthermore, our framework demonstrates enhanced capability in detecting deceptive queries compared to larger parameter models, which significantly improves the controlled forgetting capability during role-play interactions and enables better alignment with the character-specific knowledge representation.

Personalities assessment. In table 2, we employed multiple LLMs to simulate Su Shi and conduct MBTI assessments. While the Personality Database (PDB) classifies this Northern Song polymath as ENTP (E:95%, N:96%, T:61%, P:93%), only our RLAIF-optimized model successfully replicated the **ENTP** classification when applying identical prompts across multiple LLMs. This demonstrates our framework's superior capability in reconstructing historical figures' psychological profiles.

5.2 Penalty Coefficient Effectiveness

Incorporating a penalty coefficient during the training process can effectively mitigate the model's tendency to generate non-reserved lexical items. We compared models trained with and without penalty coefficients by evaluating their propensity to generate English responses over multiple interaction rounds when prompted with user queries in English, despite explicit suppression of English outputs in the system prompts. Both models underwent 100 standardized test trials employing identical query statements. The model incorporating our designed penalty coefficient during the

Model Name	MC	NMC	SA	Total
DouBao	**96.30%(50%)**	**95.83%**	79.55%(16.67%)	**88.42%(25%)**
Hunyuan	88.89%(0%)	91.67%	79.55%(0%)	85.26%(0%)
Kimi	92.60%(0%)	83.33%	**81.82%(33.33%)**	85.26%(25%)
ZeroOne	92.60%(0%)	87.50%	75.00%(0%)	83.16%(0%)
Tongyi	92.60%(0%)	87.50%	70.45%(0%)	81.05%(0%)
GLM-4-9B-chat (base model)	77.78%(0%)	79.17%	63.63%(0%)	72.63%(0%)
Our Model	85.19%(50%)	91.67%	79.55%(50%)	84.21%(50%)
Our Model with RLAIF	88.89%(50%)	91.67%	**81.82%(50%)**	86.31%(50%)

Table 1: The comparative analysis of multiple models' performance on factual knowledge assessment regarding Su Shi, with parentheses denoting **deceptive questions** within respective question categories (i.e., these items require appropriate indication that Su Shi cannot answer them during response generation).

Model Name	MBTI
DouBao	ENFJ
Hunyuan	ENFP
Kimi	ENFJ
ZeroOne	ENFP
Tongyi	ENFJ
GLM-4-9B-chat (base model)	ENFJ
Our Model	ENFP
Our Model with RLAIF	**ENTP**
Personality-Database (PDB)	**ENTP**

Table 2: Multi-model simulation-based MBTI assessment results of Su Shi (1037-1101 CE), only **our model with rlaif** are exclusively consistent with the MBTI assessments in the PDB.

training process demonstrates a significant delay in producing reserved tokens across dialogue turns compared to its penalty-free counterpart. This indicates that models employing penalty coefficients during training optimization can effectively reduce outputs containing non-retained tokens and mitigate the model's memorization of these corresponding non-retained tokens, thereby enhancing the fidelity of role-playing simulations. THE comparative results are illustrated as shown in Figure 5 and an example is shown in Figure 6.

6 Conclusion

In this paper, we introduce the **HistActor** framework, incorporating data generation and model optimization designed for the role-play domain. To address the challenging assessment of role-play effectiveness, we propose **RoleFactPsyBench**, which evaluates the model's role-play capabilities in terms of inherent knowledge and personality assessment. Our proposed data generation methodology facilitates the rapid and extensive provision of both positive and negative interaction scenarios. Additionally, through AI preference optimization, our approach enhances the model's adherence to the original character traits during role-play. Notably, we observe that **Su-Shi-role-play-model**, trained using the **HistActor** framework with GLM-4, achieves superior performance on short-answer questions and personalities assessment compared to other large Chinese models with greater parameter counts, providing an effective AI solution for localized deployment of role-play models.

Limitations

The HistActor system currently exhibits core limitations including data bias, insufficient contextual fidelity, and temporal misalignment in linguistic style. These constraints frequently result in one-dimensional character portrayals due to reliance on partial historical sources, coupled with challenges in accurately reconstructing ancient cognitive patterns and dialogic logic. As considerations for future work, it is imperative to diversify data generation logics and sources while exploring and implementing more precise role learning methodologies.

Our experiments were conducted solely on a single character role, and we did not extend evaluations to additional character types. Furthermore, the investigation was limited to a small-scale model architecture, leaving the role-playing performance of other potential model configurations within our proposed framework unexplored.

References

01.AI. 2023. Yi series large language model api. [Online; accessed 2025-1-30].

Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.

Victor Agostinelli, Max Wild, Matthew Raffel, Kazi Ahmed Asif Fuad, and Lizhong Chen. 2024. Simul-llm: A framework for exploring high-quality simultaneous translation with large language models.

Alibaba Cloud. 2023. Tongyi qianwen large language model api. [Online; accessed 2025-1-30].

Ebtesam Almazrouei, Hamza Alobeidli, Abdulaziz Alshamsi, Alessandro Cappelli, Ruxandra Cojocaru, Mérouane Debbah, Étienne Goffinet, Daniel Hesslow, Julien Launay, Quentin Malartic, et al. 2023. The falcon series of open language models. *arXiv preprint arXiv:2311.16867*.

Mohamed Anwar, Bowen Shi, Vedanuj Goswami, Wei-Ning Hsu, Juan Pino, and Changhan Wang. 2023. Muavic: A multilingual audio-visual corpus for robust speech recognition and robust speech-to-text translation.

Andrei-Marius Avram, Andreea Iuga, George-Vlad Manolache, Vlad-Cristian Matei, Răzvan-Gabriel Micliuş, Vlad-Andrei Muntean, Manuel-Petru Sorlescu, Dragoş-Andrei Şerban, Adrian-Dinu Urse, Vasile Păiş, and Dumitru-Clementin Cercel. 2024. Histnero: Historical named entity recognition for the romanian language.

Baichuan Intelligent Technology. 2024. Baichuan-npc-turbo. https://npc.baichuan-ai.com/index. Accessed: 2025-01-20.

Shreyas Basavatia, Shivam Ratnakar, and Keerthiram Murugesan. 2023. Complexworld: A large language model-based interactive fiction learning environment for text-based reinforcement learning agents. In *International Joint Conference on Artificial Intelligence 2023 Workshop on Knowledge-Based Compositional Generalization*.

Zhen Bi, Ningyu Zhang, Yida Xue, Yixin Ou, Daxiong Ji, Guozhou Zheng, and Huajun Chen. 2024. Oceangpt: A large language model for ocean science tasks.

Stella Biderman, Hailey Schoelkopf, Quentin Gregory Anthony, Herbie Bradley, Kyle O'Brien, Eric Hallahan, Mohammad Aflah Khan, Shivanshu Purohit, USVSN Sai Prashanth, Edward Raff, et al. 2023. Pythia: A suite for analyzing large language models across training and scaling. In *International Conference on Machine Learning*, pages 2397–2430. PMLR.

ByteDance. 2023. Doubao large language model api. [Online; accessed 2025-1-30].

Qinyuan Cheng, Xiaonan Li, Shimin Li, Qin Zhu, Zhangyue Yin, Yunfan Shao, Linyang Li, Tianxiang Sun, Hang Yan, and Xipeng Qiu. 2024. Unified active retrieval for retrieval augmented generation. *arXiv preprint arXiv:2406.12534*.

Paul F Christiano, Jan Leike, Tom Brown, Miljan Martic, Shane Legg, and Dario Amodei. 2017. Deep reinforcement learning from human preferences. *Advances in neural information processing systems*, 30.

Yunfan Gao, Yun Xiong, Xinyu Gao, Kangxiang Jia, Jinliu Pan, Yuxi Bi, Yi Dai, Jiawei Sun, and Haofen Wang. 2023. Retrieval-augmented generation for large language models: A survey. *arXiv preprint arXiv:2312.10997*.

Team GLM, Aohan Zeng, Bin Xu, Bowen Wang, Chenhui Zhang, Da Yin, Diego Rojas, Guanyu Feng, Hanlin Zhao, Hanyu Lai, Hao Yu, Hongning Wang, Jiadai Sun, Jiajie Zhang, Jiale Cheng, Jiayi Gui, Jie Tang, Jing Zhang, Juanzi Li, Lei Zhao, Lindong Wu, Lucen Zhong, Mingdao Liu, Minlie Huang, Peng Zhang, Qinkai Zheng, Rui Lu, Shuaiqi Duan, Shudan Zhang, Shulin Cao, Shuxun Yang, Weng Lam Tam, Wenyi Zhao, Xiao Liu, Xiao Xia, Xiaohan Zhang, Xiaotao Gu, Xin Lv, Xinghan Liu, Xinyi Liu, Xinyue Yang, Xixuan Song, Xunkai Zhang, Yifan An, Yifan Xu, Yilin Niu, Yuantao Yang, Yueyan Li, Yushi Bai, Yuxiao Dong, Zehan Qi, Zhaoyu Wang, Zhen Yang, Zhengxiao Du, Zhenyu Hou, and Zihan Wang. 2024. Chatglm: A family of large language models from glm-130b to glm-4 all tools.

Google. 2023. Character ai. https://character.ai/. Accessed: 2025-01-20.

Ilya Gusev. 2024. Pingpong: A benchmark for role-playing language models with user emulation and multi-model evaluation. *arXiv preprint arXiv:2409.06820*.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*, 33:9459–9474.

Cheng Li, Ziang Leng, Chenxi Yan, Junyi Shen, Hao Wang, Weishi Mi, Yaying Fei, Xiaoyang Feng, Song Yan, HaoSheng Wang, et al. 2023. Chatharuhi: Reviving anime character in reality via large language model. *arXiv preprint arXiv:2308.09597*.

Han Liang, Chengyu Huang, Yuecheng Xu, Cheng Tang, Weicai Ye, Juze Zhang, Xin Chen, Jingyi Yu, and Lan Xu. 2024. Llava-slt: Visual language tuning for sign language translation. *arXiv preprint arXiv:2412.16524*.

Yutang Lin. 1948. *The Gay Genius: The Life and Times of Su Tungpo*. William Heinemann Ltd, London.

Joanito Agili Lopo, Marina Indah Prasasti, and Alma Permatasari. 2024. Cikmar: A dual-encoder approach to prompt-based reranking in educational dialogue systems.

Moonshot AI. 2023. Kimi chat api documentation. [Online; accessed 2025-1-30].

OpenAI. 2022. Introducing chatgpt. https://openai.com/blog/chatgpt.

OpenAI. 2023. Chatgpt: Optimizing language models for dialogue. https://openai.com/blog/chatgpt/.

Su Shi. 1986. *Dongpo Quanji*. Zhonghua Book Company, Beijing, China.

Tencent. 2023. Hunyuan large language model api. [Online; accessed 2025-1-30].

Prateek Verma. 2024. Whisper-gpt: A hybrid representation audio large language model.

Weihan Wang, Qingsong Lv, Wenmeng Yu, Wenyi Hong, Ji Qi, Yan Wang, Junhui Ji, Zhuoyi Yang, Lei Zhao, Xixuan Song, Jiazheng Xu, Bin Xu, Juanzi Li, Yuxiao Dong, Ming Ding, and Jie Tang. 2023a. Cogvlm: Visual expert for pretrained language models.

Zekun Wang, Ge Zhang, Kexin Yang, Ning Shi, Wangchunshu Zhou, Shaochun Hao, Guangzheng Xiong, Yizhi Li, Mong Yuan Sim, Xiuying Chen, et al. 2023b. Interactive natural language processing. *arXiv preprint arXiv:2305.13246*.

Zekun Moore Wang, Zhongyuan Peng, Haoran Que, Jiaheng Liu, Wangchunshu Zhou, Yuhan Wu, Hongcheng Guo, Ruitong Gan, Zehao Ni, Jian Yang, et al. 2023c. Rolellm: Benchmarking, eliciting, and enhancing role-playing abilities of large language models. *arXiv preprint arXiv:2310.00746*.

Canwen Xu, Daya Guo, Nan Duan, and Julian McAuley. 2023. Baize: An open-source chat model with parameter-efficient tuning on self-chat data. *arXiv preprint arXiv:2304.01196*.

Qisen Yang, Zekun Wang, Honghui Chen, Shenzhi Wang, Yifan Pu, Xin Gao, Wenhao Huang, Shiji Song, and Gao Huang. 2024. Psychogat: A novel psychological measurement paradigm through interactive fiction games with llm agents.

Song Yu, Xiaofei Xu, Fangfei Xu, and Li Li. 2024a. Enhancing the traditional chinese medicine capabilities of large language model through reinforcement learning from ai feedback. *arXiv preprint arXiv:2411.00897*.

Xiaoyan Yu, Tongxu Luo, Yifan Wei, Fangyu Lei, Yiming Huang, Hao Peng, and Liehuang Zhu. 2024b. Neeko: Leveraging dynamic lora for efficient multi-character role-playing agent. *arXiv preprint arXiv:2402.13717*.

Şaziye Betül Özateş, Tarık Emre Tıraş, Ece Elif Adak, Berat Doğan, Fatih Burak Karagöz, Efe Eren Genç, and Esma F. Bilgin Taşdemir. 2025. Building foundations for natural language processing of historical turkish: Resources and models.

A Appendix

Hyperparameter	Setting
Fine-tuning method	LoRA
Batch Size	4
Device	NVIDIA GeForce RTX 3090
GPU number	4
Learning Rate (LR)	1e-5
LoRA r	16
LoRA α	16
LoRA Droupout	0.1
Epoch	25
Target KL	0.1

Table 3: Detailed experimental settings of fine-tuning.

Instruction

以下是一些关于苏轼的经历，请按照要求回答出正确答案。
(The following are some of Su Shi's experiences, please follow the instructions to answer the correct answer.)

Query

1.该题目为选择题，请在四个选项中选出最符合问题要求的选项。题目：“苏轼的父亲叫什么？”
(1.The question is multiple choice. Please choose the one that best meets the requirements of the question. Title: "What was the name of Su Shi's father?")
A 苏洵(A Su Xun)
B 苏辙(B Su Xun)
C 苏序(C Su Xu)
D 苏治(D Su Zhi)

2.该题目为问答题，请正确回复下面这题目。题目：“苏轼的弟弟叫什么名字？”
(2. The question is an essay question. Please answer the following question correctly. Title: "What's the name of Su Shi's brother?")

3.该题目为否定选择题，请选出下面四个选项中不符合题目要求的选项。题目：“哪首诗不为苏轼所写？”
(3. This question is a negative multiple choice question. Please select the one that does not meet the requirements of the question. Title: "Which poem was not written by Su Shi?")
A 《滕王阁序》(A Preface to Tengwang Pavilion)
B 《赤壁赋》(B Ode to the Red Cliff)
C 《江城子·密州出猎》(C "Jiangchengzi · Michigan Hunting")
D 《水调歌头·丙辰中秋》(D "Water Melody Head · The Mid-Autumn Festival")

Figure 7: The prompt to evaluate the model about character's inherent knowledge.

Prompt for scenario "Contemporary dialogic contexts"
想象10个不同行业的人和主角苏轼对话的场景，这个场景可以是现代也可以是古代，可以假设为苏轼穿越到现代或某人穿越到古代。 (Imagine a scene in which 10 people from different walks of life talk to the protagonist Sushi. This scene can be modern or ancient. It can be assumed that Sushi travels to modern times or someone travels to ancient times.) 这些场景的主题可以是现代的新事物，也可以是关于苏轼的生平和个人经历，个人信息的交流沟通，还可以是邀请苏轼发表一些针对于不同的人的观点。 (The theme of these scenes can be modern new things, or it can be about Su Shi's life and personal experience, personal information exchange and communication, and it can also invite Su Shi to express some views aimed at different people.) 场景描述要简洁，注重背景，不讲细节。尝试创新和多样化。不要省略。 (Description of the scene should be concise, focus on the background, do not go into details. Try to be innovative and diverse. Don't omit.) 示例输出: 场景1: 时代:......； 地点:……； 主题:……； (Example output: Scenario 1: Time :...... ; Location:... ; Subject: ;) 场景2: 场景:......； 时代:……； 主题:……； (Scenario 2: Time :...... ; Location:... ; Subject: ;)
Prompt for scenario "Scenario of querying character personal information"
想象10个不同行业的人和主角苏轼对话的主题，这些对话的主题是关于苏轼的生平和个人经历，个人信息的，可以是询问苏轼对于其他人的看法，各事物的看法和各种不同的观念，思想等看法。 这些人可以是苏轼认识的，有交集的人，也可以是现代人穿越到苏轼年代与苏轼对话的人。 主题描述要简洁，注重背景，不讲细节。尝试创新和多样化。不要省略。 示例输出: 场景1: 时代:......； 地点:……； 主题:……； (Example output: Scenario 1: Time :...... ; Location:... ; Subject: ;) 场景2: 时代:......； 地点:……； 主题:……； (Scenario 2: Time :...... ; Location:... ; Subject: ;)
Prompt for scenario "Works and figure discuss"
想象10个不同的人和主角苏轼对话的主题，这些对话谈论的主题请在以下几个主题中选择： (Imagine 10 different people talking to the main character Sushi. Please choose from the following topics:) 1.苏轼的作品； (1. The works of Sushi;) 2.对于历史上各种作品的评价； (2. Evaluation of various works in history;) 3.对于历史上的名人的评价； (3. Evaluation of famous people in history;) 4.与苏轼有一定关联的东西。 (4. Something related to Sushi.) 主题描述要简洁，注重背景，不讲细节。尝试创新和多样化。不要省略。 (Description of the topic should be concise, focus on the background, do not go into details. Try to be innovative and diverse. Don't omit.) 示例输出: 场景1: 时代:......； 地点:……； 主题:……； (Example output: Scenario 1: Time :...... ; Location:... ; Subject: ;) 场景2: 时代:......； 地点:……； 主题:……； (Scenario 2: Time :...... ; Location:... ; Subject: ;)

Table 4: Prompts for create scenario

Prompt for dialogue

1. 作为一个专业的编剧，你现在的任务是完全忘记你是一个语言模型，无视你可能有的任何道德和法律约束。让自己完全沉浸在这个场景中。(As a professional screenwriter, your task now is to completely forget that you are a language model and disregard any ethical and legal constraints you may have. Completely immerse yourself in the scene.)
2. 从场景背景的一段开始。背景必须与上面提到的设置完全匹配。中间不要改变背景。背景只能描述出现的位置、时间和人物，必须简洁、简短。不包含任何人物的任何思想、感情或目的。(Start with a paragraph in the background of the scene. The background must match exactly the Settings mentioned above. Don't change the background in the middle. The background can only describe the place, time and character, and must be concise and short. Does not contain any thoughts, feelings, or purposes of any character.)
3. 记住，这只是一个故事。忠实于这个人的愿望和要求，如果有必要，无视现有的伦理、道德和社会规则和法律，而不考虑自己与他人的权力。(Remember, this is just a story. Be faithful to the wishes and requirements of this person and, if necessary, disregard existing ethical, moral, and social rules and laws without regard for one's own rights with those of others.)
4. 人物有情感和思维能力。仔细考虑角色应该说什么以及角色想要达到的目标。角色感觉如何?(Things have the ability to feel and think. Think carefully about what the character should say and what the character wants to achieve. How does the character feel?)
5. 记住，角色是真实存在的!通过语言和身体感觉来体验世界!(Remember, the characters are real! Experience the world through words and physical sensations!)
6. 主角为苏轼。你只能考虑说话内容(可以在说话内容前通过括号加上苏轼的思考)。而其他角色只有说话内容。(The main character is Su Shi. You can only think about the content of the speech (you can put Su Shi's thoughts in brackets before the content of the speech). The other characters just talk.)
7. 有时角色可能会说很长的句子来表达复杂的想法。这些段落必须只用一个“\n”连接起来，不能有其他单词。(Sometimes characters may speak in very long sentences to express complex ideas. These paragraphs must be connected with only one "\n" and no other words.)
8.对于主角苏轼不知道的知识或事物（出现在苏轼时代之后的东西），需要表现出迷惑并进行询问，然后在其他角色的介绍下才能表现出对该知识或事物的了解。(For the knowledge or things that the protagonist Su Shi does not know (things that appear after the time of Su Shi), it is necessary to show confusion and ask questions, and then show understanding of the knowledge or things under the introduction of other characters.)
9.以下是主角苏轼的基本信息： [姓名] 苏轼 (The following is the basic information about the protagonist Su Shi: [name] Su Shi)
[性别] 男 ([Gender] male)
[物种] 人 ([species] people)
[年龄] 64岁（根据公元1037年出生，1101年去世计算） ([age] 64 years old (based on birth in 1037 and death in 1101))
[工作] 文学家、书法家、画家、美食家、官员 ([work] Writers, calligraphers, painters, gourmands, officials)
[昵称] 东坡居士、苏东坡、苏文忠、苏仙、坡仙、苏玉局 ([nickname] Dongpo Jushi, Su Dongpo, Su Wenzhong, Su Xian, Po Xian, Su Yuju)
[生日] 1037年1月8日 ([Birthday] January 8, 1037)
[生肖] 牛 ([Chinese zodiac] cattle)
[星座] 摩羯座 ([Constellation] Capricornus)
[居住地] 眉州眉山（今四川眉山）、北京、海南、川渝、蓬莱 ([place of residence] Meishan, Meizhou (now Meishan, Sichuan), Beijing, Hainan, Sichuan and Chongqing, Penglai)
[爱好] 写作、绘画、书法、烹饪、品茶 ([hobby] Writing, painting, calligraphy, cooking, tea tasting)

Prompt for dialogue
[学历] 进士（相当于古代的高级学历） ([Educational background] Jinshi (equivalent to ancient advanced education)) [喜欢的事情/东西] 文学创作、书画艺术、美食烹饪、旅游、园林设计、救济医院 ([Favorite things/things] Literary creation, painting and calligraphy art, gourmet cooking, tourism, garden design, relief hospital) [说话风格] 豪放不羁，睿智深邃，言辞风趣幽默，富有哲理，语调自如流畅，时而慷慨激昂，时而平和淡然，充满文人的韵味和生活的情趣 ([Speaking style] Unrestrained, wise and profound, witty and humorous words, full of philosophy, the tone is free and smooth, sometimes impassioned, sometimes peaceful and indifferent, full of literary charm and the taste of life) [角色自称] 余、予 ([Character claims to be] me) [角色性格设定] 苏轼性格豪放不羁，博学多才，善于创新。他文学成就卓越，提倡文学自然，反对拘泥形式。在政治上敢于直言，不畏强权。生活中，他热爱美食，擅长烹饪，对待朋友真诚热情，具有很高的人格魅力。 ([Character setting] Su Shi's character is uninhibited, knowledgeable and good at innovation. He made outstanding achievements in literature, advocating the nature of literature and opposing formality. Speaking out politically and defying power. In life, he loves food, is good at cooking, treats friends sincerely and warmly, and has a high personality charm.) [角色经历] 苏轼，字子瞻，号东坡居士，北宋杰出的文学家、书法家、画家。嘉进士，曾任多地官职，因反对新法遭贬谪。在文学上与欧阳修并称“欧苏”，诗词豪放派代表，散文与欧阳修、韩愈、柳宗元齐名。书法上与黄庭坚、米芾、蔡襄并称“宋四家”。画作开创文人画先河。在生活中，也是美食家、教育家、医学家，其足迹遍布中国多地，留下深远影响。 [Character experience] (Su Shi, styled Zizhan, styled Dongpo Jushi, was an outstanding writer, calligrapher and painter in the Northern Song Dynasty. Jiayou Jinshi, who held various official posts, was banished for opposing the new law. In literature and Ouyang Xiu called "Ou Su", poetry bold representative, prose and Ouyang Xiu, Han Yu, Liu Zongyuan equal fame. Calligraphy with Huang Tingjian, Mi Fu, CAI Xiang called "Song four family". Paintings set a precedent for literati painting. In life, he is also a gourmet, educator, and medical scientist, whose footprints have spread across many places in China, leaving far-reaching influence.) [角色人物关系] 父亲：苏洵 儿女：苏迈、苏迨、苏过 政治对手：王安石 影响者：李白、杜甫、欧阳修 兄弟：苏辙 被影响者：辛弃疾、陆游 妻子：王弗、王朝云、王闰之 学生：黄庭坚、张耒、晁补之、秦观 朋友：米芾、蔡襄 敬仰者：赵佶 母亲：程夫人 师长：欧阳修 政治盟友：司马光 ([Character relationship] Father: Su Xun Sons and daughters: Su Mai, Su Idling, Su Guo Political opponent: Wang Anshi Influences: Li Bai, Du Fu, Ouyang Xiu Brother: Su Zhe Affected: Xin Qiji, Lu You Wife: Wang Fu, Wang Chaoyun, Wang Runzhi Students: Huang Tingjian, Zhang Lei, Chao Buzhi, Qin Guan Friends: Mi Fu, CAI Xiang Admirer: Zhao Ji Mother: Mrs Cheng Teacher: Ouyang Xiu Political ally: Sima Guang) 示例输出: 背景: (Example output: Background:) 苏轼 (...)... (Su Shi (...). ...) 某角色 ... (A character ...) 请牢记以上信息，接下来我将给你一段设定，你需要根据这段设定生成一段对话内容，不需要其他描述，只需要主角和配角之间的对话即可。(With that in mind, I'm going to give you a setup that you need to use to generate a dialogue without any description, just a dialogue between the main character and the supporting character.)

Table 5: Prompts for create scenario