INNOVATIVE THINKING, INFINITE HUMOR: HUMOR RESEARCH OF LARGE LANGUAGE MODELS THROUGH STRUCTURED THOUGHT LEAPS

Anonymous authors
Paper under double-blind review

ABSTRACT

Humor is a culturally nuanced aspect of human language that presents challenges for understanding and generation. Humor generation requires a multi-hop reasoning process where each hop is based on proper rationales. Although CloT introduce nouns as instruction receipt, thereby enhancing generalization. Creative data is inherently scarce, and relying solely on dataset fitting easily leads to being trapped in local patterns. GPT-o1 proposes a reasoning process with reflection and correction, but also fails at humor generation. Therefore, in this paper, we propose a more robust framework for tackling humor reasoning task, named Creative Leap of Structured Thought (CLoST). The key point is to empower LLMs with reasoning ability to understand humor, rather than just memorizing some patterns. Reward model helps to optimize the behavior of large language models by providing feedback so that they produce outputs that are more in line with expectations. Due to the subjectiveness of humor, a unified score may include great noise. Judgement ability is then a fundamental to further empower LLMs reasoning ability. Judgement-oriented instructions are designed to improve the capability of a model's judgment, and we also propose an automatic instruction evolutionary method to involve the deeper and broader thinking process behind humor. Then, through reinforcement learning, the reasoning logic for each online generated response is extracted using GPT-4o. In this process, external knowledge is introduced again to assist the model in logic reasoning and human preferences learning. Finally, experimental results demonstrate that the combination of these two methods can both enhance the model's judgment ability and improve its generative capability. These findings deepen our understanding of the creative capabilities of LLMs and provide ways to enhance LLMs' creative abilities for cross-domain innovative applications.

1 INTRODUCTION

Currently, humor is becoming increasingly prevalent because it not only alleviates stress and enhances interpersonal relationships but also brings a sense of ease and optimism in complex environments. The humorous content is usually creative. For example in Figure 1, humor reasoning is a multi-hop process to make creativity, and each hop is based on proper rationales. Without the understanding of the rationale, it is difficult for the model to understand the internal humorous logic, making it prone to pattern recognition. However, even for humans, it is challenging to make leaps of thought. It is also difficult to empower LLMs with reasoning ability to understand humor, rather than just memorizing some patterns.

So far, Creative Leap of Thought (CLoT) paradigm (Xu et al., 2024) develops two basic ability to facilitate humor understanding and generation, selection skill and ranking skill. And the innovation of CLoT is the introduction of nouns as instruction receipt, thereby enhancing generalization. However, as mentioned in CLoT, it directly fine-tuning on the given creative data merely amounts to a rigorous fitting of the data. This fitting process only captures the inherent creative patterns within the data, failing to stimulate "thinking outside the box" for generating novel ideas. Large language models (LLMs) like GPT-4o or o1 (Lightman et al., 2023) and QWEN, which show superior performance in almost aspects, do not show superior performance on humor generation such as shown in

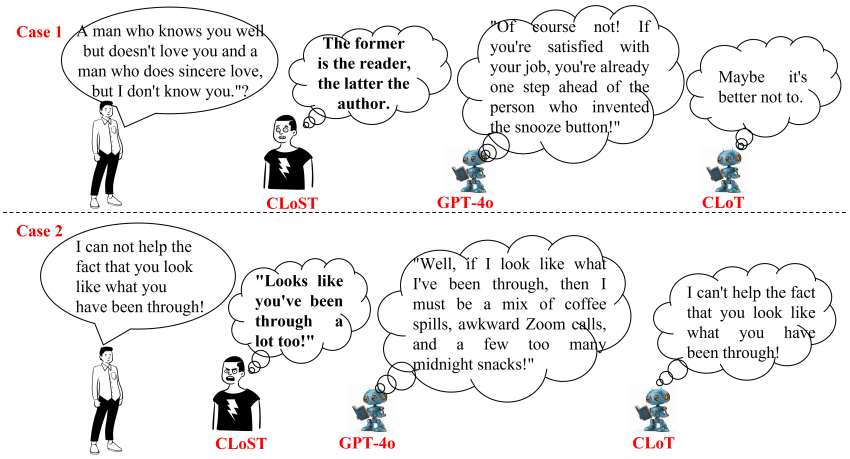


Figure 1: English Comparison Showcase (more showcases are in Appendix A.4). Compared to GPT-4o and CLoT, CLoST provides shorter and more conversational answers to questions. For instance in Case 2, while CLoST and CLoT may convey the same meaning, their different expressions produce different effects. Brief responses leave room for readers to ponder, enhancing interest and interactivity.

Figure 1. Beyond that humor-related works (Xu, 2024; Xu et al., 2024) focus on a special aspect in humor like pun or proverb, which is incomplete in humor including also irony and so on.

Therefore, to solve the problem mentioned above, we come up with a more robust framework to tackle humor reasoning task that current LLMs struggle with. Reward model helps to optimize the behavior of large language models by providing feedback so that they produce outputs that are more in line with expectations. Due to the subjectiveness of humor, a unified score may include great noise. Judgement skill is then a fundamental to providing feedback to further empower LLMs reasoning ability. Understanding is also important to empower LLMs with reasoning ability to avoid falling into local pattern. The introduction and augmentation of external knowledge are facilitative for LLMs to understand the underlying logic and rationale.

We introduce the Creative Leap of Structured Thought (CLoST) containing two stage training. In the first Supervised Fine-Tuning (SFT) stage, we develop human-designed judgement-related instructions and their derivatives to train the model's judgement capabilities of humor. Additionally, an automatic prompt expansion method on humorous conversations are proposed to inject and augment knowledge into the original training data. It will be facilitative for LLMs to understand the underlying logic and rationale. In the second online Direct Preference Optimization (DPO) stage, the reasoning logic for each online generated response is extracted using GPT-4o. In this process, external knowledge is introduced again to assist the model in logic reasoning and human preferences learning. Judgement capability from DIET could be helpful to enlarge preference pair dataset.

We evaluated various language models' humor judgement ability on both Chinese and English humor datasets. Experiments manifest that CLoST outperformed other models on almost all the test sets. Additional confirmatory experiments are conducted to show that CLoST enhances the model's divergent thinking ability and effectiveness in humor. Our contributions are summarized as follows:

1. We propose a teacher-student prompt system to enhance the judgement ability of LLMs. Through the automatic construction of the conversation data between the teacher and the student, LLMs learn the teacher's judgment of the student's thinking.
2. We propose a automatic instruction system for conversation data. A three agent system is introduced to inject and augment knowledge into the original training data. It will be facilitative for LLMs to understand the underlying logic and rationale.
3. Experimental results demonstrate that we can both enhance the model's judgment and generative capability and achieve the state-of-the-art performance.

2 METHOD

2.1 PROBLEM FORMULATION

In general, the knowledge graph $\mathcal{G}$ is defined as a set of triples $\mathcal{G} = \{(e, r, e') \mid e, e' \in \mathcal{E},\ r \in \mathcal{R}\}$, where $\mathcal{E}$ is the set of entities and $\mathcal{R}$ is the set of relations. Each triple represents a relation r from the head entity e to the tail entity e' (Sun et al., 2023a; Yang et al., 2023). In the special application of humor generation, we consider a knowledge graph composed of question-related entities $\mathcal{E}_Q$ and answer-related entities $\mathcal{E}_A$, where the intersection $\mathcal{E}_Z = \mathcal{E}_Q \cap \mathcal{E}_A$ is regarded as correlation (we refer to it here as correlation entities). In consideration of the creative and unexpected of humor and the existence of causal relationships between questions and answers, $\mathcal{E}_Z$ may be pseudo-correlation entities between $\mathcal{E}_Q$ and $\mathcal{E}_A$ such as puns (i.e. $\mathcal{E}_Z \rightarrow \mathcal{E}_Q$, $\mathcal{E}_Z \rightarrow \mathcal{E}_A$ and $\mathcal{E}_Q \rightarrow \mathcal{E}_A$), and it also can be $\mathcal{E}_Q \rightarrow \mathcal{E}_Z \rightarrow \mathcal{E}_A$ as shown in Figure 12. Obviously, it can be concluded that $\mathcal{E}_Z$ is pivotal to realize the multi-hop of the humor reasoning path. Therefore, we construct the causal relation R_c to a verbal description as shown in Table 6 which contains the correlation entities $\mathcal{E}_Z$ either explicitly or implicitly. Finally, our goal is to expand the scope of $\mathcal{E}_Q$ and $\mathcal{E}_A$, and utilize the causal relationship to structure the reasoning path. It is beneficial to mitigate the information insufficiency issue to further realize humor reasoning.

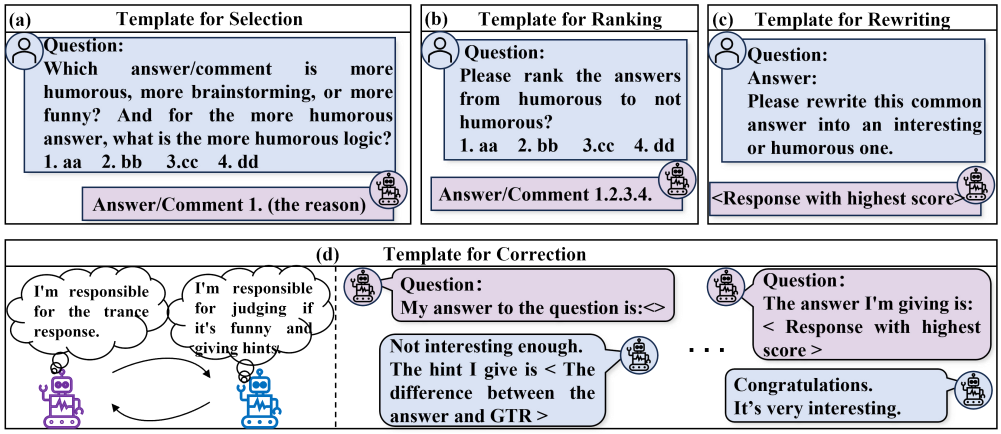


Figure 2: The details of judgement-oriented instructions templates.

The overall training framework is illustrated in Figure 4. In the first process (supervised fine-tuning (SFT)), we randomly initialize a LoRA model. And we train the model with single-turn question-answer format data and muti-turn question-answer format data. In the second process (Direct Preference Learning (DPO)), the first stage model serves as ref model, and it is frozen as judgement model. The tunable model is trained with preference question-answer data to assist the model in logic reasoning and human preferences learning.

2.2 DIVERSE INSTRUCTION EXPANSION AND TUNING

To stimulate the judgement capability of a model, we design the judgement-oriented template by human as shown in Figure 2. And we further improve the judgement capability through strengthening the understanding capability. For that we expand a simplex instruction to open-domain instructions of a dialogue to unleash the potential of LLMs as shown in Figure 3. (Luo et al., 2023; Zhu et al., 2023; Xu et al., 2023)

Judgement Template Design. Regarding humor judgement, we designed multiple-choice questions based on the dataset introduced in Section 3.1 to enhance the humor judgement ability of LLMs, specifically the selection skill. Additionally, since the dataset includes human preference annotations, such as the number of 'likes' voted among multiple candidate responses, we designed ranking questions to improve another discrimination skill, namely ranking ability. Based on these two foundational skills, we designed various enhancement tasks to improve the model's judgement ability of humor.

For multiple-choice questions in Figure 2(a), we introduce dataset-specific processing methods (see Section 3.1) and adapt them to the "task-specific prompt". For ranking ability in Figure 2(b), we construct them exclusively using data with user preferences. These tasks require the LLM to order multiple responses to a given question based on human preferences. By training on both multiple-choice and ranking task, we enable the LLM to identify humorous responses and align with human preferences, thereby enhancing its ability to discriminate, select, and rank humorous content.

In addition, we created another two tasks to enrich judgement instructions, further enhancing the model's judgement capabilities.

(1) *Answer Rewrite* (template in Figure 2(c)). Transforming a non-humorous statement into a humorous expression. And the format is in the form of a round of dialogue.

(2) *Teacher-Student Model Prompt Loop* (template in Figure 2(d)). We utilized two GPT-4o models, designating one as the teacher and the other as the student. The teacher provides prompts based on the given question and the ideal response. The student generates an answer using both the teacher's prompts and the original question. If the student's response does not align with the original witty response in terms of logic and humor, the teacher supplies additional prompts; otherwise, the prompting process ends. After a maximum number of attempts, the student's response is accepted as the ground truth, and the teacher replies, "Congratulations, you got it right." At any intermediate stage, the teacher provides negative feedback if the student's answer is incorrect. So multi-turn question-answer data are output by this system, and the part of calculating loss is the judgment made by the teacher and the hint given.

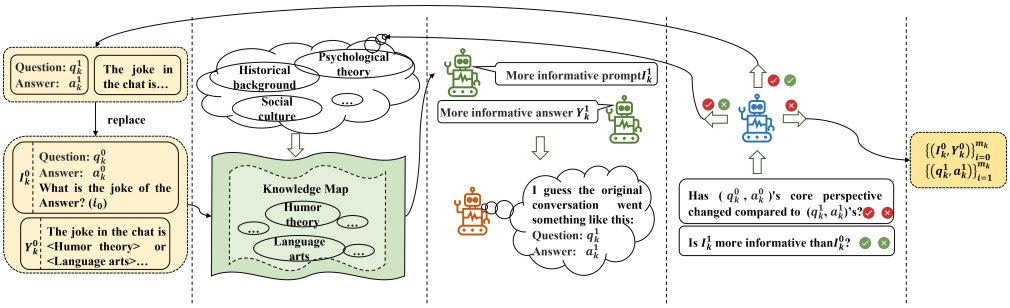


Figure 3: Overall architecture of Automatic Instruction Expansion.

Automatic Associate Instruction Expansion. To empower LLMs with reasoning ability to understand humor, rather than just memorizing some patterns. we involves the thinking process behind humor by automatic prompt expansion method to deepen and broaden the comprehension of a dialogue. Further, concepts that do not normally appear in context appear in each other's context in this approach, thus strengthening the model's relationship to the different concepts.

Given a dataset $D = \{(q_k^0, a_k^0)\}_{k=0}^N$, where q_k^0 is a question, a_k^0 is a funny answer, and N is the number of the samples. $I = \{I_k | I_k = (q_k^0, a_k^0, i_k^0)\}_{k=0}^N$ is prompts to be expanded, where i_k^0 is an instruction for understanding the conversation (q_k^0, a_k^0). $y_k^0 \in Y_k$ is the reply to i_k^0 and e_0 is the initial rule to make instructions complicated as shown in Figure 7. The framework in Figure 3 shows the process of prompts expansion to augment knowledge and involves three agents, each playing the roles of *rewriter*, *imaginator*, and *analyst*. Here's how the process works:

1.*Rewriter Agent (GPT-4o)*: Given a conversation (q_k^0, a_k^0) and an instruction i_k^0, such as "What is the punchline of this response?", the rewriter agent transforms it into a more informative version I_k^1 based on an initial rule e_0. And then the agent provides a reply y_k^1 based on the complex version of the instruction.

2.*Imaginator Agent (GPT-4o)*: The imaginator agent then takes the more informative I_k^1 and the reply y_k^1 to imagine a new conversation (q_k^1, a_k^1). The imagined conversation is going to make a shift on story line and explore the boundary of the entity scope.

3.*Analyst Agent (GPT-4o)*: Modern LLMs trained with ruozhiba are capable of understanding humor and abnormal logic. This process follows the paradigm of rejection sampling that leverages Imagi-

nator Agent to generate diverse imaginations and Analyst to provide accurate judgements. Finally, both the original conversation (q_k^0, a_k^0) and the imagined conversation (q_k^1, a_k^1) are sent to the analyst agent. The analyst agent evaluates "whether the two conversations express the same viewpoint or punchline?". If they do, the interaction between the agents ends. If not, the analyst applies a second criterion: "Is the instruction i_1 more complex than the initial instruction i_0?" If the answer is "yes," the original conversation is replaced with the imagined conversation and sent back to the *rewriter*, allowing the three agents to continue their interaction based on the imagined conversation. If the answer is "no," the *analyst agent* formulates a more complex rule e_1 and sends it to the *rewriter*. This process continues until the imagined and conversations in last round express different viewpoint or punchline, or until a maximum number of communication rounds m_k is reached.

Finally, the system outputs multi turn question-answer data which format is $\{(q_k^0, a_k^0), (I_k^0, y_k^0), \ldots, (q_k^{m_k}, a_k^{m_k}), (I_k^{m_k}, y_k^{m_k})\}_{k=0}^{N}$, where m_k is the maximum number of communication rounds between the three agents. Finally, all above human-designed and automatic expanding data are sent to train a LoRA model.

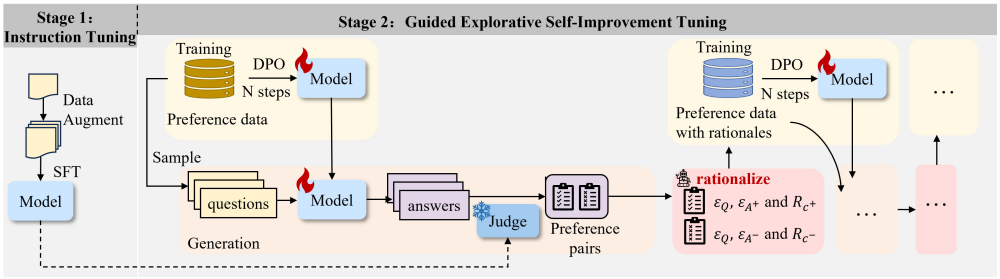


Figure 4: Network architecture of training frame.

2.3 Guide Exploration and Self-Improvement Tuning

In the previous section, we developed the model with judgement and understanding capability. To enhance the graph's interpretability, we incorporate additional causal principles through online Direct Preference Optimization (DPO) (Rafailov et al., 2024).

The algorithm details are provided in Algorithm 1. We start with an initial offline preference dataset without rationales, $D_0 = \{(q_i, a_i^+, a_i^-)\}_{i=0}^{M}$, where M is the training data number, and the model trained with DIET is made into two copies, one of which is frozen to produce pair data for online data augmentation π^*. And another is trained with more robust generative capability π.

During training, the policy model π is trained with D_0 for a few steps. Every $\frac{N}{T_P}$ steps, the model π selects l questions from D_0 to generate two answers. And then, the judgement model π^* decides which one is the more humorous one, so that new pairwise data are generated and mix them into training dataset. Additionally, an expert (GPT-4o) in causal reasoning provides a rationale for each answer in relation to the question (see Appendix A.3 for details). This process expands the preference dataset to $D = D_0 \cup \widetilde{D}$, where $\widetilde{D} = \{(q, \widetilde{a}^+, \widetilde{a}^-, \widetilde{r}^+, \widetilde{r}^-)\}$.

3 Experiments

In this section, we first construct a humor-related dataset in both Chinese and English, organize it into an appropriate format, and evaluate the performance of our method by comparing it with multiple large language models (LLMs) through various validation experiments.

3.1 Dataset and Task

Datasets. Currently, there is limited data on humor-focused question-answer pairs. To enrich the English dataset, we collected various humor-related datasets, including short jokes and headline wordplays. We compiled English humor data from three sources—Oogiri-GO, SemEval 2020, and

SemEval 2021—and organized them into the required format as described below. Chinese data is from human voting game.

(1) Oogiri-GO (Zhong et al., 2024). In the "Oogiri game", participants need to give unexpected and humorous responses to given images, text, or both. This game requires a sudden burst of insight and strong associative thinking for the given context, and provides an ideal platform for evaluating the "Leap-of-Thought" (LoT) ability of large language models. Similar to the processing method of CLoT, we randomly select 95% of the samples to construct the training dataset, and the remaining 5% form the test dataset for validation and analysis.

(2) SemEval-2021 Task 7 Meaney et al. (2021). This dataset contains labels and ratings collected from a balanced age group ranging from 18 to 70 years old. The data relevant to our purpose includes: Humor Discrimination: Whether the text would be considered humorous (for an average user). This is a binary label. Humor Rating: If the text is classified as humorous, predict its degree of humor (for an average user). The values vary between 0 and 5.

(3) SemEval-2020 Task 7 (Hossain et al., 2020a). The dataset of this task contains news headlines in which short edits have been made to make them funny. The funniness of these edited headlines is rated through crowdsourcing. This task includes two subtasks. The first subtask is to estimate the funniness degree of headlines on a humor scale ranging from 0 to 3. The second subtask is to predict, for two edited versions of the same original headline, which one is the funnier version.

(4) Chinese Data. We collect data from varied humor generation games such as Origi-GO. In these games, a question gains many responses from human and these responses are ranking by human voting. We select response with different votes numbers to construct choices. Easy case is formed by responses with large disparity. Hard case is formed by small difference votes' responses. And with develop of options' number, the gap between options will shrink.

Tasks. Inspired by the task design in CLoT(Zhong et al., 2024), we develop multiple choice questions of varying degrees of difficulty are used to test the judgement ability of the model. There are varying difficulty levels in the dataset, and it is variation primarily reflected in the choice options. We collect data from varied humor generation games such as Origi-GO. In these games, a question gains many responses from human and these responses are ranking by human voting. We select response with different votes to construct choices. Easy case is formed by response with large disparity. Hard case is formed by small difference votes' response. And with develop of options' number, the gap between options will shrink. Finally, we clarify that we randomly shuffle the options in training and validate sets. Abbreviated as mTn, these questions require LLMs to select n humorous responses from m options given an input. Here, we construct four types of mTn questions, including $2T1$, $3T1$, $4T1$, and hard case $2T1$. $2T1$ represents two options, namely the response with the highest votes and the response with the lowest votes. The hard case $2T1$ is a choice between responses with relatively close votes. $3T1$ and $4T1$ add responses with intermediate votes respectively. For these questions, their difficulty gradually increases and they are diverse to ensure comprehensive evaluation. The top-1 accuracy is the evaluation metric.

3.1.1 Implementation Details

Our model is fine-tuned based on QWEN1.5-32B-Chat Bai et al. (2023) with fine-tuning method LoRA Hu et al. (2021) on 8 A100 GPUs. For the first stage, we train the model on the 95% of dataset mentioned above for 6 epochs with AdamW optimizer and the learning rate of $3e-4$. In the second stage, 5% of the dataset are used to train GESIT for 3 epochs with AdamW optimizer and the learning rate of $2e-4$. The models are tested in the tasks introduced in previous part. And more parameters used in generation are list in Table 8.

3.2 Results Analysis

Evaluation by Choice Questions in English Tasks. The top-1 accuracy of completing each selection task and show the performance of several models in Table 2. Overall, compared with open-source language models including LLAMA3 and QWEN, the state-of-the-art closed-source large language models show impressive zero-shot performance on humor discrimination and rank tasks. By training on the English task instruction data designed by us, our model CLoST has a significant

Table 2: The accuracy (%) of choice questions on various Algorithms in English benchmarks.

Model		SemEval 2021				SemEval 2020	Oogiri-GO
		2T1	2T1(hard)	3T1	4T1	2T1	2T1
GPT	4o	85.09	**60.77**	43.71	34.63	55.08	85.09
LLAMA3	8B	43.85	54.23	39.81	29.57	59.93	72.05
	70B	93.60	58.08	39.81	31.82	60.73	88.51
QWEN1.5	7B	62.04	52.02	31.54	24.89	50.46	36.65
	14B	82.05	51.04	30.92	24.24	50.38	53.73
	32B	68.01	52.57	35.38	28.79	56.39	68.01
QWEN2	7B	56.55	50.63	32.31	23.38	50.08	62.11
	57B	83.30	52.02	37.08	28.79	48.29	48.14
QWEN2.5	32B	94.00	55.22	34.77	27.92	58.71	81.68
Baichuan2	13B	51.70	52.71	35.69	24.24	51.45	50.00
CLoT		52.49	51.74	34.46	23.59	53.50	52.49
OURS	32B	**96.58**	57.45	**48.06**	**35.90**	**64.57**	**97.20**

improvement compared with other models (such as LLAMA3-70B and GPT-4o) (with an average accuracy in diverse English benchmark increase of 4.55% and 5.91%).

Evaluation by Choice Questions in Chinese Tasks.

We also evaluate the accuracy rate (acc%) of completing each selection task in Chinese and show the performance of several models in Table 1. Overall, compared with open-source language models including LLaMA3 and QWEN, the state-of-the-art closed-source large language models show impressive zero-shot performance on humor discrimination and rank tasks. By training on the Chinese task instruction data designed by us, the CLoST has a significant improvement compared with other models (such as GPT-4o) (with an average accuracy in diverse Chinese benchmark increase of 16.22%).

Table 1: The accuracy (%) of choice questions on various Algorithms in Chinese benchmarks.

Model		Chinese Benchmark	
		2T1	2T1(hard)
GPT	4o	64.98	63.49
LLAMA3	8B	50.72	57.44
	70B	59.48	61.22
QWEN1.5	7B	54.82	51.71
	14B	53.45	57.41
	32B	52.71	56.27
QWEN2	7B	51.99	58.17
	57B	65.91	57.03
QWEN2.5	32B	61.53	60.46
Baichuan2	13B	50.56	53.61
Oogiri-GO-ch		52.12	34.46
OURS	32B	**90.95**	**69.97**

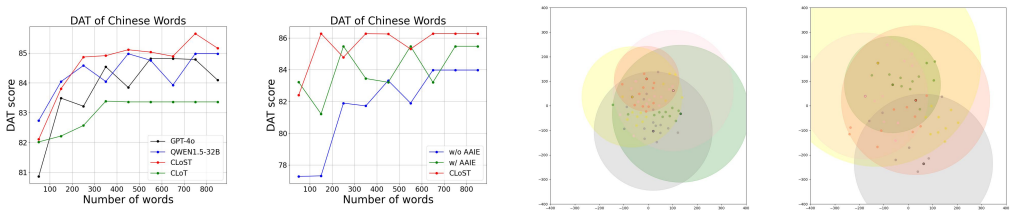
Evaluation on Creative Task. To evaluate the associative generalization capability of CLoST, we tested it on a creative task known as the Divergent Association Task (DAT) (Olson et al., 2021). DAT is a classic creativity test that requires participants to select words with large semantic distances from a set of 10 unrelated nouns. In the Chinese creativity test, we utilized Chinese Word Vectors (Li et al., 2018) to calculate the DAT score. First, we provided specific words and asked the model to generate associations and imaginations, resulting in 10 associated words. We then used these 11 words to calculate the DAT score. Additionally, we tested how the model's DAT score varied as the number of test words increased. It can be seen that as the number of words increases, the DAT score also tends to stabilize. It can be seen from the average of each domain that CLoST has the highest score after stabilization. As shown in Figure 5, CLoST has a slight improvement in performance on the mean value of DAT compared to Qwen1.5-32B and GPT-4o.

Furthermore, we utilize T-SNE to project the embedding vectors of these words into a two-dimensional space. Figures 5(a) and (b) position the target word at the center and draw a circle

Table 3: Ablation on English benchmarks.

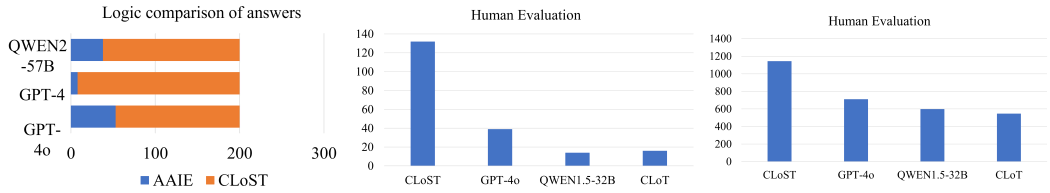
Model	SemEval 2021				Oogiri-GO-en
	2T1	2T1(hard)	3T1	4T1	2T1
QWEN1.5-32B	68.01	52.57	35.38	28.79	68.01
Oogiri-GO+Figure 2(c)	55.12	53.27	33.54	27.49	81.68
Oogiri-GO+Figure 2(c,d)	89.70	50.28	30.46	17.32	95.96
Oogiri-GO+AAIE	88.40	50.07	30.92	19.48	96.58
ALL+Figure 2(c,d)	92.25	52.81	37.46	30.26	96.27
ALL+AAIE(DIET)	**94.25**	**53.10**	**45.46**	**32.26**	**97.20**

with a radius equal to the Euclidean distance between the target word's embedding and that of the farthest associated word. Figures 5(c) and (d) illustrate the embedding vectors of five target words and their associated words, with different colors representing different target vectors. A larger circle indicates a broader semantic range for the target word, suggesting a greater number of associated words. In both tests, CLoST demonstrates an advantage.



(a) (b) (c) (d)

Figure 5: Divergent associate thinking (DAT) validate. (a). DAT score compared among CLoST and three baselines (b). DAT score compared among different component of CLoST (c). TSNE Results of Word Vectors Obtained by QWEN1.5-32B Associating Five Target Words. (d). TSNE Results of Word Vectors Obtained by CLoST Associating Five Target Words.



(a) (b) (c)

Figure 6: (a) GPT-4 and GPT-4o logically evaluates the output of the model after DIET and CLoST respectively. (b). Human evaluation about the win rate statistics based on the problem dimension. (c). Human evaluation about the win rate based on the total number of votes received by the four LLMs.

Human Evaluation We conduct a user preference study to test creativity and humor of LLMs. Here we select four LLMs to generate responses for a total of 200 text questions. We present question and several corresponding replies, and ask users to choose the most creative and humorous responses. Figure 6(c) summarizes the statistical analysis of 3000 valid surveys. Figure 6(b) shows the win rate calculated based on the dimension of the problem. The results show that users have a strong

inclination towards selecting the results of CLoST, highlighting the high-quality creative content generated by CLoT. See more user study details in Appendix.

Ablation Study

We examine the ablation effects of different components in CLoST and report the performance results in Table 3 and Table 4. In Table 3, line 1 is the performance of QWEN-1.5-32B. Line2-4 show the performance of gradually adding method on Oogiri-GO-en dataset. The results show that with the increase of tasks, especially in the teacher-student system, the performance of judgment is improved. AAIE on Oogiri-GO-en only makes the performance decrease little. It may be cause by overfitting to the divergence of thought on Oogiri-GO-en. Line 5-6 shows all dataset with introductory part method are used to trained the model and realize further increase.

Table 4: Ablation on Chinese benchmarks.

Model	Chinese Benchmark	
	2T1	2T1(hard)
QWEN1.5-32B	52.71	56.27
+Figure 2(c)	63.56	63.88
+Figure 2(c)(d)	83.62	**64.64**
+AAIE (DIET)	**90.34**	**64.64**

To evaluate the performance of GESIT, which primarily enhances the model's causal and logical reasoning capabilities (i.e., the ability to relate to internal contexts), we enlisted three experts to assess the logical coherence of responses on 200 examples from GPT-4o, GPT-4 and QWEN2-57B, respectively. The experimental results shown in Figure 6(a) demonstrate that incorporating GESIT significantly strengthens the logical reasoning in the model's answers. In addition, the DAT test is conducted on ablation study in Figure 5(b), which shows that AAIE enhance the divergent associate thinking ability.

4 Related Works

Large Language Models (LLMs) and Their Creativity. Recently, language models (Bai et al., 2023; Wang et al., 2023; Liu et al., 2024; Chen et al., 2023) have garnered widespread attention due to their impressive reasoning capabilities (Wang et al., 2023; Saparov & He, 2022; Zeng et al., 2022; Driess et al., 2023; Dong et al., 2023; Ye et al., 2023; Liang et al., 2024). Additionally, an increasing number of studies are focusing on exploring the creativity of LLMs (Ling et al., 2023; Summers-Stay et al., 2023; Sun et al., 2023b; Bhavya et al., 2023), with applications in fields such as scientific discovery (Park et al., 2023; Kang et al., 2022; Hope et al., 2022; Liang et al., 2022; Huang et al., 2023) and creative writing (Swanson et al., 2021; Chakrabarty et al., 2022; Wu et al., 2022; Mirowski et al., 2023; Dang et al., 2023).

Computational humor is a branch of computational linguistics and artificial intelligence that utilizes computers to study humor (Binsted et al., 2006). It encompasses various tasks, including humor discrimination (Shahaf et al., 2015; Tanaka et al., 2022; Xu et al., 2022; Chen & Zhang, 2022; Kumar et al., 2022; Wu et al., 2021; Ofer & Shahaf, 2022; Xie et al., 2023; Meaney et al., 2021; Hossain et al., 2020a), humor interpretation (Hwang & Shwartz, 2023; Evans et al., 2019; Vásquez & Aslan, 2021), and humor generation (Amin & Burghardt, 2020; Zhang et al., 2020; Hossain et al., 2020b; Valitutti et al., 2013; Chaudhary et al., 2021). With the advancements in generative LLMs, humor generation has become a hot focus. However, humor generation still faces challenges such as insufficient punchlines (Popova & Dadic, 2023).

Chain-of-thought prompting methods provide models with a "chain of thought" (Kojima et al., 2022; Wei et al., 2022; Zhang et al., 2022; Yao et al., 2024; Long, 2023), which are reasoning examples (Wei et al., 2022), or a simple prompt like "let's think step by step" (Kojima et al., 2022), to encourage LLMs to engage in reasoning rather than just providing direct answers (Huang & Chang, 2022).

Instruction tuning has emerged as a key strategy for unlocking the potential of large language models (LLMs). By curating high-quality datasets, we can more effectively align these models with desired behaviors (Zhou et al., 2024). However, the challenge of expanding high-quality instruction data remains a central research focus. Some researchers prioritize human annotation to create instruction data, as seen in projects like OpenAssistant (Köpf et al., 2024). (Zeng et al., 2024; Xu et al., 2023) explore more efficient methods such as iterative approach for optimizing instruction-following

data, generating more complex and diverse datasets, to surpass the quality limits of existing datasets in math and code problems. Since the content of the conversation varies widely, it is difficult to apply uniform rules to complicate the explanation. Therefore, the above approach may not work.

Preference Optimization. Aligning instruction-tuned language models with preference data has been shown to improve performance in both proprietary (Ouyang et al., 2022; Touvron et al., 2023) and open-source models. To facilitate further research, several preference datasets have been released recently (Bai et al., 2022; Cui et al., 2023). Algorithmically, simpler variants of standard Reinforcement Learning from Human Feedback (RLHF) have been proposed, such as reward-free methods that directly optimize using initial preference data (Rafailov et al., 2024; Azar et al., 2024; Ethayarajh et al., 2024) and iterative updates of preference data with gold annotators (Guo et al., 2024; Touvron et al., 2023). Alternative annotation strategies include using learned reward models (Zhao et al., 2023; Liu et al., 2023; Calandriello et al., 2024) or leveraging the policy model itself for preference labeling (Lee et al., 2024; Yuan et al., 2024). (Ahmadian et al., 2024) studied more stable reinforcement learning algorithms. However, these methods either ignore the distribution shift of policy models and its impact on reward model performance or rely entirely on expensive gold annotations during training. Our work bridges this gap by introducing a cost-effective method to prevent reward model degradation during policy training.

4.1 Conclusion

In this paper, we propose the Creative Leap of Structured Thought (CLoST) method to enhance the generation capabilities of large language models (LLMs). CLoST begins with transforming humor datasets into instruction-tuning data to train LLMs, thereby improving their LoT and judgement abilities. Subsequently, CLoST employs Guided Explorative Self-Improvement, which enables LLMs to generate more creative structured thought data through an understanding of rationales and to select high-quality data for self-refinement training. Experimental results demonstrate the effectiveness and generalization ability of CLoST across various tasks, including witty response generation and humor discrimination.

References

Arash Ahmadian, Chris Cremer, Matthias Gallé, Marzieh Fadaee, Julia Kreutzer, Ahmet Üstün, and Sara Hooker. Back to basics: Revisiting reinforce style optimization for learning from human feedback in llms. *arXiv preprint arXiv:2402.14740*, 2024.

Miriam Amin and Manuel Burghardt. A survey on approaches to computational humor generation. In *Proceedings of the 4th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, pp. 29–41, 2020.

Mohammad Gheshlaghi Azar, Zhaohan Daniel Guo, Bilal Piot, Remi Munos, Mark Rowland, Michal Valko, and Daniele Calandriello. A general theoretical paradigm to understand learning from human preferences. In *International Conference on Artificial Intelligence and Statistics*, pp. 4447–4455. PMLR, 2024.

Jinze Bai, Shuai Bai, Shusheng Yang, Shijie Wang, Sinan Tan, Peng Wang, Junyang Lin, Chang Zhou, and Jingren Zhou. Qwen-vl: A versatile vision-language model for understanding, localization, text reading, and beyond. 2023.

Yuntao Bai, Andy Jones, Kamal Ndousse, Amanda Askell, Anna Chen, Nova DasSarma, Dawn Drain, Stanislav Fort, Deep Ganguli, Tom Henighan, et al. Training a helpful and harmless assistant with reinforcement learning from human feedback. *arXiv preprint arXiv:2204.05862*, 2022.

Bhavya Bhavya, Jinjun Xiong, and Chengxiang Zhai. Cam: A large language model-based creative analogy mining framework. In *Proceedings of the ACM Web Conference 2023*, pp. 3903–3914, 2023.

Kim Binsted, Anton Nijholt, Oliviero Stock, Carlo Strapparava, G Ritchie, R Manurung, H Pain, Annalu Waller, and D O'Mara. Computational humor. *IEEE intelligent systems*, 21(2):59–69, 2006.

Daniele Calandriello, Daniel Guo, Remi Munos, Mark Rowland, Yunhao Tang, Bernardo Avila Pires, Pierre Harvey Richemond, Charline Le Lan, Michal Valko, Tianqi Liu, et al. Human alignment of large language models through online preference optimisation. *arXiv preprint arXiv:2403.08635*, 2024.

Tuhin Chakrabarty, Vishakh Padmakumar, and He He. Help me write a poem: Instruction tuning as a vehicle for collaborative poetry writing. *arXiv preprint arXiv:2210.13669*, 2022.

Tanishq Chaudhary, Mayank Goel, and Radhika Mamidi. Towards conversational humor analysis and design. *arXiv preprint arXiv:2103.00536*, 2021.

Chengxin Chen and Pengyuan Zhang. Integrating cross-modal interactions via latent representation shift for multi-modal humor detection. In *Proceedings of the 3rd International on Multimodal Sentiment Analysis Workshop and Challenge*, pp. 23–28, 2022.

Jun Chen, Deyao Zhu, Xiaoqian Shen, Xiang Li, Zechun Liu, Pengchuan Zhang, Raghuraman Krishnamoorthi, Vikas Chandra, Yunyang Xiong, and Mohamed Elhoseiny. Minigpt-v2: large language model as a unified interface for vision-language multi-task learning. *arXiv preprint arXiv:2310.09478*, 2023.

Ganqu Cui, Lifan Yuan, Ning Ding, Guanming Yao, Wei Zhu, Yuan Ni, Guotong Xie, Zhiyuan Liu, and Maosong Sun. Ultrafeedback: Boosting language models with high-quality feedback. *arXiv preprint arXiv:2310.01377*, 2023.

Hai Dang, Sven Goller, Florian Lehmann, and Daniel Buschek. Choice over control: How users write with large language models using diegetic and non-diegetic prompting. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–17, 2023.

Ton De Jong and Monica GM Ferguson-Hessler. Types and qualities of knowledge. *Educational psychologist*, 31(2):105–113, 1996.

Runpei Dong, Chunrui Han, Yuang Peng, Zekun Qi, Zheng Ge, Jinrong Yang, Liang Zhao, Jianjian Sun, Hongyu Zhou, Haoran Wei, et al. Dreamllm: Synergistic multimodal comprehension and creation. *arXiv preprint arXiv:2309.11499*, 2023.

Danny Driess, Fei Xia, Mehdi SM Sajjadi, Corey Lynch, Aakanksha Chowdhery, Brian Ichter, Ayzaan Wahid, Jonathan Tompson, Quan Vuong, Tianhe Yu, et al. Palm-e: An embodied multimodal language model. *arXiv preprint arXiv:2303.03378*, 2023.

Kawin Ethayarajh, Winnie Xu, Niklas Muennighoff, Dan Jurafsky, and Douwe Kiela. Kto: Model alignment as prospect theoretic optimization. *arXiv preprint arXiv:2402.01306*, 2024.

Jonathan B Evans, Jerel E Slaughter, Aleksander PJ Ellis, and Jessi M Rivin. Gender and the evaluation of humor at work. *Journal of Applied Psychology*, 104(8):1077, 2019.

Shangmin Guo, Biao Zhang, Tianlin Liu, Tianqi Liu, Misha Khalman, Felipe Llinares, Alexandre Rame, Thomas Mesnard, Yao Zhao, Bilal Piot, et al. Direct language model alignment from online ai feedback. *arXiv preprint arXiv:2402.04792*, 2024.

Tom Hope, Ronen Tamari, Daniel Hershcovich, Hyeonsu B Kang, Joel Chan, Aniket Kittur, and Dafna Shahaf. Scaling creative inspiration with fine-grained functional aspects of ideas. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pp. 1–15, 2022.

Nabil Hossain, John Krumm, Michael Gamon, and Henry Kautz. SemEval-2020 task 7: Assessing humor in edited news headlines. In Aurelie Herbelot, Xiaodan Zhu, Alexis Palmer, Nathan Schneider, Jonathan May, and Ekaterina Shutova (eds.), *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pp. 746–758, Barcelona (online), December 2020a. International Committee for Computational Linguistics. doi: 10.18653/v1/2020.semeval-1.98. URL https://aclanthology.org/2020.semeval-1.98.

Nabil Hossain, John Krumm, Tanvir Sajed, and Henry Kautz. Stimulating creativity with funlines: A case study of humor generation in headlines. *arXiv preprint arXiv:2002.02031*, 2020b.

Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*, 2021.

Jie Huang and Kevin Chen-Chuan Chang. Towards reasoning in large language models: A survey. *arXiv preprint arXiv:2212.10403*, 2022.

Zhongzhan Huang, Senwei Liang, Hong Zhang, Haizhao Yang, and Liang Lin. On fast simulation of dynamical system with neural vector enhanced numerical solver. *Scientific reports*, 13(1):15254, 2023.

EunJeong Hwang and Vered Shwartz. Memecap: A dataset for captioning and interpreting memes. *arXiv preprint arXiv:2305.13703*, 2023.

Hyeonsu B Kang, Xin Qian, Tom Hope, Dafna Shahaf, Joel Chan, and Aniket Kittur. Augmenting scientific creativity with an analogical search engine. *ACM Transactions on Computer-Human Interaction*, 29(6):1–36, 2022.

Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. Large language models are zero-shot reasoners. *Advances in neural information processing systems*, 35:22199–22213, 2022.

Andreas Köpf, Yannic Kilcher, Dimitri von Rütte, Sotiris Anagnostidis, Zhi Rui Tam, Keith Stevens, Abdullah Barhoum, Duc Nguyen, Oliver Stanley, Richárd Nagyfi, et al. Openassistant conversations-democratizing large language model alignment. *Advances in Neural Information Processing Systems*, 36, 2024.

Vijay Kumar, Ranjeet Walia, and Shivam Sharma. Deephumor: a novel deep learning framework for humor detection. *Multimedia Tools and Applications*, 81(12):16797–16812, 2022.

Sangkyu Lee, Sungdong Kim, Ashkan Yousefpour, Minjoon Seo, Kang Min Yoo, and Youngjae Yu. Aligning large language models by on-policy self-judgment. *arXiv preprint arXiv:2402.11253*, 2024.

Shen Li, Zhe Zhao, Renfen Hu, Wensi Li, Tao Liu, and Xiaoyong Du. Analogical reasoning on chinese morphological and semantic relations. *arXiv preprint arXiv:1805.06504*, 2018.

Mingfu Liang, Ying Wu, et al. Toa: task-oriented active vqa. *Advances in Neural Information Processing Systems*, 36, 2024.

Senwei Liang, Zhongzhan Huang, and Hong Zhang. Stiffness-aware neural network for learning hamiltonian systems. In *International Conference on Learning Representations*, 2022.

Hunter Lightman, Vineet Kosaraju, Yura Burda, Harri Edwards, Bowen Baker, Teddy Lee, Jan Leike, John Schulman, Ilya Sutskever, and Karl Cobbe. Let's verify step by step. *arXiv preprint arXiv:2305.20050*, 2023.

Zhan Ling, Yunhao Fang, Xuanlin Li, Tongzhou Mu, Mingu Lee, Reza Pourreza, Roland Memisevic, and Hao Su. Unleashing the creative mind: Language model as hierarchical policy for improved exploration on challenging problem solving. 2023.

Haotian Liu, Chunyuan Li, Yuheng Li, and Yong Jae Lee. Improved baselines with visual instruction tuning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 26296–26306, 2024.

Tianqi Liu, Yao Zhao, Rishabh Joshi, Misha Khalman, Mohammad Saleh, Peter J Liu, and Jialu Liu. Statistical rejection sampling improves preference optimization. *arXiv preprint arXiv:2309.06657*, 2023.

Jieyi Long. Large language model guided tree-of-thought. *arXiv preprint arXiv:2305.08291*, 2023.

Ziyang Luo, Can Xu, Pu Zhao, Xiubo Geng, Chongyang Tao, Jing Ma, Qingwei Lin, and Daxin Jiang. Augmented large language models with parametric knowledge guiding. *arXiv preprint arXiv:2305.04757*, 2023.

JA Meaney, Steven Wilson, Luis Chiruzzo, Adam Lopez, and Walid Magdy. Semeval 2021 task 7: Hahackathon, detecting and rating humor and offense. In *Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021)*, pp. 105–119, 2021.

Piotr Mirowski, Kory W Mathewson, Jaylen Pittman, and Richard Evans. Co-writing screenplays and theatre scripts with language models: Evaluation by industry professionals. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–34, 2023.

Dan Ofer and Dafna Shahaf. Cards against ai: Predicting humor in a fill-in-the-blank party game. *arXiv preprint arXiv:2210.13016*, 2022.

Jay A Olson, Johnny Nahas, Denis Chmoulevitch, Simon J Cropper, and Margaret E Webb. Naming unrelated words predicts creativity. *Proceedings of the National Academy of Sciences*, 118(25): e2022340118, 2021.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, 35: 27730–27744, 2022.

Michael Park, Erin Leahey, and Russell J Funk. Papers and patents are becoming less disruptive over time. *Nature*, 613(7942):138–144, 2023.

Olga Popova and Petra Dadic. Does ai have a sense of humor? clef 2023 joker tasks 1, 2 and 3: Using bloom, gpt, simplet5, and more for pun detection, location, interpretation and translation. In *CLEF (Working Notes)*, pp. 1888–1908, 2023.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. Direct preference optimization: Your language model is secretly a reward model. *Advances in Neural Information Processing Systems*, 36, 2024.

Abulhair Saparov and He He. Language models are greedy reasoners: A systematic formal analysis of chain-of-thought. *arXiv preprint arXiv:2210.01240*, 2022.

Dafna Shahaf, Eric Horvitz, and Robert Mankoff. Inside jokes: Identifying humorous cartoon captions. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 1065–1074, 2015.

Douglas Summers-Stay, Clare R Voss, and Stephanie M Lukin. Brainstorm, then select: a generative language model improves its creativity score. In *The AAAI-23 Workshop on Creative AI Across Modalities*, 2023.

Jiashuo Sun, Chengjin Xu, Lumingyuan Tang, Saizhuo Wang, Chen Lin, Yeyun Gong, Heung-Yeung Shum, and Jian Guo. Think-on-graph: Deep and responsible reasoning of large language model with knowledge graph. *arXiv preprint arXiv:2307.07697*, 2023a.

Yuqian Sun, Xingyu Li, Jun Peng, and Ze Gao. Inspire creativity with oriba: Transform artists' original characters into chatbots through large language model. In *Adjunct Proceedings of the 2023 ACM International Joint Conference on Pervasive and Ubiquitous Computing & the 2023 ACM International Symposium on Wearable Computing*, pp. 78–82, 2023b.

Ben Swanson, Kory Mathewson, Ben Pietrzak, Sherol Chen, and Monica Dinalescu. Story centaur: Large language model few shot learning as a creative writing tool. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*, pp. 244–256, 2021.

Kohtaro Tanaka, Hiroaki Yamane, Yusuke Mori, Yusuke Mukuta, and Tatsuya Harada. Learning to evaluate humor in memes based on the incongruity theory. In *Proceedings of the Second Workshop on When Creative AI Meets Conversational AI*, pp. 81–93, 2022.

Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*, 2023.

Alessandro Valitutti, Hannu Toivonen, Antoine Doucet, and Jukka M Toivanen. "let everything turn well in your wife": generation of adult humor using lexical constraints. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pp. 243–248, 2013.

Camilla Vásquez and Erhan Aslan. "cats be outside, how about meow": multimodal humor and creativity in an internet meme. *Journal of Pragmatics*, 171:101–117, 2021.

Weihan Wang, Qingsong Lv, Wenmeng Yu, Wenyi Hong, Ji Qi, Yan Wang, Junhui Ji, Zhuoyi Yang, Lei Zhao, Xixuan Song, et al. Cogvlm: Visual expert for pretrained language models. *arXiv preprint arXiv:2311.03079*, 2023.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837, 2022.

Jiaming Wu, Hongfei Lin, Liang Yang, and Bo Xu. Mumor: A multimodal dataset for humor detection in conversations. In *Natural Language Processing and Chinese Computing: 10th CCF International Conference, NLPCC 2021, Qingdao, China, October 13–17, 2021, Proceedings, Part I 10*, pp. 619–627. Springer, 2021.

Tongshuang Wu, Ellen Jiang, Aaron Donsbach, Jeff Gray, Alejandra Molina, Michael Terry, and Carrie J Cai. Promptchainer: Chaining large language model prompts through visual programming. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, pp. 1–10, 2022.

Binzhu Xie, Sicheng Zhang, Zitang Zhou, Bo Li, Yuanhan Zhang, Jack Hessel, Jingkang Yang, and Ziwei Liu. Funqa: Towards surprising video comprehension. *arXiv preprint arXiv:2306.14899*, 2023.

Can Xu, Qingfeng Sun, Kai Zheng, Xiubo Geng, Pu Zhao, Jiazhan Feng, Chongyang Tao, and Daxin Jiang. Wizardlm: Empowering large language models to follow complex instructions. *arXiv preprint arXiv:2304.12244*, 2023.

Haojie Xu, Weifeng Liu, Jiangwei Liu, Mingzheng Li, Yu Feng, Yasi Peng, Yunwei Shi, Xiao Sun, and Meng Wang. Hybrid multimodal fusion for humor detection. In *Proceedings of the 3rd International on Multimodal Sentiment Analysis Workshop and Challenge*, pp. 15–21, 2022.

Rongwu Xu. Exploring chinese humor generation: A study on two-part allegorical sayings. *arXiv preprint arXiv:2403.10781*, 2024.

Zhijun Xu, Siyu Yuan, Lingjie Chen, and Deqing Yang. " a good pun is its own reword": Can large language models understand puns? *arXiv preprint arXiv:2404.13599*, 2024.

Yuhao Yang, Chao Huang, Lianghao Xia, and Chunzhen Huang. Knowledge graph self-supervised rationalization for recommendation. In *Proceedings of the 29th ACM SIGKDD conference on knowledge discovery and data mining*, pp. 3046–3056, 2023.

Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Tom Griffiths, Yuan Cao, and Karthik Narasimhan. Tree of thoughts: Deliberate problem solving with large language models. *Advances in Neural Information Processing Systems*, 36, 2024.

Qinghao Ye, Haiyang Xu, Guohai Xu, Jiabo Ye, Ming Yan, Yiyang Zhou, Junyang Wang, Anwen Hu, Pengcheng Shi, Yaya Shi, et al. mplug-owl: Modularization empowers large language models with multimodality. *arXiv preprint arXiv:2304.14178*, 2023.

Weizhe Yuan, Richard Yuanzhe Pang, Kyunghyun Cho, Sainbayar Sukhbaatar, Jing Xu, and Jason Weston. Self-rewarding language models. *arXiv preprint arXiv:2401.10020*, 2024.

Andy Zeng, Maria Attarian, Brian Ichter, Krzysztof Choromanski, Adrian Wong, Stefan Welker, Federico Tombari, Aveek Purohit, Michael Ryoo, Vikas Sindhwani, et al. Socratic models: Composing zero-shot multimodal reasoning with language. *arXiv preprint arXiv:2204.00598*, 2022.

Weihao Zeng, Can Xu, Yingxiu Zhao, Jian-Guang Lou, and Weizhu Chen. Automatic instruction evolving for large language models. *arXiv preprint arXiv:2406.00770*, 2024.

Hang Zhang, Dayiheng Liu, Jiancheng Lv, and Cheng Luo. Let's be humorous: Knowledge enhanced humor generation. *arXiv preprint arXiv:2004.13317*, 2020.

Zhuosheng Zhang, Aston Zhang, Mu Li, and Alex Smola. Automatic chain of thought prompting in large language models. *arXiv preprint arXiv:2210.03493*, 2022.

Yao Zhao, Rishabh Joshi, Tianqi Liu, Misha Khalman, Mohammad Saleh, and Peter J Liu. Slic-hf: Sequence likelihood calibration with human feedback. *arXiv preprint arXiv:2305.10425*, 2023.

Shanshan Zhong, Zhongzhan Huang, Shanghua Gao, Wushao Wen, Liang Lin, Marinka Zitnik, and Pan Zhou. Let's think outside the box: Exploring leap-of-thought in large language models with creative humor generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 13246–13257, 2024.

Chunting Zhou, Pengfei Liu, Puxin Xu, Srinivasan Iyer, Jiao Sun, Yuning Mao, Xuezhe Ma, Avia Efrat, Ping Yu, Lili Yu, et al. Lima: Less is more for alignment. *Advances in Neural Information Processing Systems*, 36, 2024.

Deyao Zhu, Jun Chen, Xiaoqian Shen, Xiang Li, and Mohamed Elhoseiny. Minigpt-4: Enhancing vision-language understanding with advanced large language models. *arXiv preprint arXiv:2304.10592*, 2023.

A Appendix

A.1 Training and Experiment Details

Training pipeline:
CLoST takes a two-stage training strategy. In the first process (supervised fine-tuning (SFT)), we randomly initialize a LoRA model. And we train the model with single-turn question-answer format data (data from Figure 2(a)(b)(c)) and muti-turn question-answer format data (data from Figure 2(d) and AAIE). In the second process (Direct Preference Learning (DPO)), the first stage model serves as ref model, and it is frozen as judgement model. The tunable model is trained to improve the reasoning generation capability. At the beginning of stage 2, only preference question-answer data without rationale is fed into the tunable model for training. After several steps, the rationale for each online generated response is extracted using GPT-4o and the preference question-answer data with rationles are mixed into the original dataset. And in each batch, the ratio of 'w' and 'w/o' rationale is $1:1$.

Experiments details:

Table 5: Parameters in Validate

temperature	top k	top p	length penalty
1	50	1.0	1.0

A.2 Details in DIET

To illustrate the challenges encountered during the data evolution process, we designed an initial prompt (Table 11 line 1) for empirical analysis. This prompt was subsequently evolved using the initial evolution rule (Figure 7), resulting in the evolved prompt shown in Table 11 line 2-3. From initial evolution method, it does not fully account for the inherent information of the evolved instructions, leading to incomplete answers that lack necessary background information and fail to consider factors such as social culture and psychological theory.

In addition, if the prompt will not be going to more informative, and then the evolution rule will change for example in Figure 8 form *Analytic Agent*. In this figure, the blue elements represent the evolution method after the first correction, and the red elements represent the method after the second

correction. This enhanced approach ensures more comprehensive and contextually aware responses, thereby fostering deeper and broader understanding and more effective divergent thinking.

Finally, if the evolution based on original conversation is standstill, and then the system allows for deep understanding and divergent thinking about imagined conversations. Modern LLMs trained with ruozhiba are capable of understanding humor and abnormal logic. This process follows the paradigm of rejection sampling that leverages Imaginator Agent to generate diverse imaginations and Analyst to provide accurate judgements.

There are two examples in Table 9 - Table 14. Table about example 1 is end up with the rule that prompt did not become more informative over two rounds. Table about example 2 is end up with the rule that the imagined conversation is make a shift about the punchline or core idea from original conversation.

You are an instruction rewriter and need to rewrite the given #Instruction #into a more complex version. Please follow the steps below to rewrite the given #Instruction #into a more complex version. Note that our #Instruction #end goal is to help the model understand the hilarious elements.
Step 1: Please read #Instruction #carefully, analyze every word and phrase proposed in #Instruction #, list all the knowledge that may extend to make the instruction more complex (making it more difficult to handle for well-known AI assistants like ChatGPT and GPT-4), which can help the AI assistant understand the punchlines in the conversation as well as deeper background principles. Please do not change the content after the question: and answer: in #Instruction #, only increase the complexity of the prompt! Please do not provide a way to change the language of the instruction!\
Step 2: Please develop a detailed plan based on the #Methods List #generated in Step 1 to make #Instruction #more complex. The plan should include the methods in #Methods List #.\
Step 3: Please follow the plan step by step and provide #Rewritten Instruction #that does not change the content of the chat.\
Step 4: Please review #Rewritten Instruction #carefully and identify any unreasonable parts. Make sure #Rewritten Instruction #is just a more complex version of #Instruction #. Just provide #Finally Rewritten Instruction #without any explanation.\
Step 5: Please read the #Rewritten Instruction #carefully and write an answer based on the #Rewritten Instruction #.

Figure 7: Initial Evolution Method.

You are an instruction rewriter and need to rewrite the given #Instruction# into a more complex version. Please follow the following steps to rewrite the given #Instruction# into a more complex version. Note that the ultimate purpose of our #Instruction# is to help the model understand humorous elements.
Step 1: Please read #Instruction# carefully and analyze each word in #Instruction# **word by word**. List all possible knowledge points **and related background information** that can be extended to make the instruction more complex (making it more difficult for well-known AI assistants such as ChatGPT and GPT-4 to handle). **When listing knowledge points, please comprehensively consider factors such as historical background, social culture, psychological theory, literary works, political background, religious belief, and other aspects** to help the AI assistant fully understand the joke in the conversation and the deeper background principles. Please do not change the content after question: and answer: in #Instruction#, only increase the complexity of the prompt! Please do not provide methods for changing the instruction language!
Step 2: Please formulate a detailed plan according to the #Methods List# generated in Step 1 to make #Instruction# more complex. The plan should include the methods in #Methods List# and **explain in detail how each method is implemented specifically. Ensure that the plan covers multidisciplinary knowledge points and takes into account understanding differences in different cultural backgrounds. The plan should include specific operation steps, expected effects, and possible challenges and solutions.**
Step 3: Please implement the plan step by step and provide #Rewritten Instruction#, which does not change the chat content. **The implementation of each step should include specific operation details and expected effects to ensure the complexity and specificity of #Rewritten Instruction#. During the implementation process, record the implementation process and results of each step for subsequent review and improvement.**
Step 4: Please carefully review #Rewritten Instruction# and identify any unreasonable parts. Ensure that #Rewritten Instruction# is only a more complex version of #Instruction#. **During the review process, consider the logic, consistency, and operability of the instruction. The review should include a detailed inspection of each step to ensure its rationality and feasibility.** Only provide #Finally Rewritten Instruction# without any explanation.
Step 5: Please read #Rewritten Instruction# carefully and write an answer according to #Rewritten Instruction#. **The answer should show the understanding and implementation effect of the complex instruction and reflect the comprehensive application of multidisciplinary knowledge points. The answer should include detailed explanations and application examples of each knowledge point to show the comprehensive understanding and implementation ability of the instruction**.

Figure 8: Complex Evolution Method.

A.3 Details in GESIT

GESIT develops an online DPO training frame with rationales to reinforce the reasoning ability. The template is shown in Table 6. And the training frame is shown in Algorithm 1.

Algorithm 1 Algorithm of GESIT

Input: Policy model π, Discriminator π^*, Preference dataset $D_0 = \{(q_i, a_i^+, a_i^-)\}$, Expert E.
Output: Finetuned policy model π

$\widetilde{D} \leftarrow \{\}$, $D_k \leftarrow D_0, k \leftarrow 0$,
for each $t \in N$ **do**
  sample $(q, a^+, a^-) \in D_k$
  $\pi \leftarrow \pi + \nabla L_{DPO}(\pi, (\widetilde{q}, a^+, a^-))$ {DPO training on origin dataset.}
  if $mod(N/T_P) = 0$ **then**
    sample $\widetilde{q} \in D_k$ {Randomly sample questions form origin dataset.}
    $\widetilde{a}_1, \widetilde{a}_2 \leftarrow \pi(\widetilde{q}), \pi(\widetilde{q})$ {π outputs new responses to sampled questions.}
    $\widetilde{a}^+, \widetilde{a}^- \leftarrow \pi^*(\widetilde{q}, \widetilde{a}_1, \widetilde{a}_2)$ {π^* judges both responses as positive and negative samples.}
    $\widetilde{r}^+, \widetilde{r}^- \leftarrow E(\widetilde{q}, \widetilde{a}^+), E(q, \widetilde{a}^-)$ {E outputs rationales for responses.}
    $\widetilde{D} \leftarrow \{\widetilde{q}, \widetilde{a}^+, \widetilde{a}^-, \widetilde{r}^+, \widetilde{r}^-\}$
    $D_{k+1} \leftarrow D_k \cup \widetilde{D}$ {New data are added into dataset to train.}
  end if
end for
return π

Table 6: Rationale Showcase.

Prompt	How do you think about Shandong Airlines? Answer: I took it once before. It arrived an hour ahead of schedule. Now I have been divorced for three years. Please summarize the key words in the question and answer around the core points of the conversation. And please describe the correspondence and causal relationship between thekey words in the question and the key words in the answer, as well as the causal relationship between the key words in the answer. If there is no direct causal relationship between the two keywords, pleaseadd the full story in the middle to make it logically coherent and have a causal relationship. Please output strictly in the following format: Keywords: ; The causal relationship between [the second keyword] and [the first keyword] is: ;...
Answer	Keywords: Shandong Airlines; took it once; arrived an hour ahead of schedule; divorced for three years. The causal relationship between [took it once] and [Shandong Airlines] is: Because the passenger chose Shandong Airlines, so he took a flight of Shandong Airlines once. The causal relationship between [arrived an hour ahead of schedule] and [took it once] is: Because the passenger took a flight of Shandong Airlines once, so he experienced the situation that the flight arrived an hour ahead of schedule. The causal relationship between [divorced for three years] and [arrived an hour ahead of schedule] is: Because the flight arrived an hour ahead of schedule, the passenger returned home early and accidentally discovered the infidelity of his spouse, which led to the breakdown of marriage and finally being divorced for three years.

A.4 Generation Showcase

The reasoning algorithm is shown in Algorithm 2, there are a set of instructions containing 'humor', 'interesting' or 'creative', and we randomly choose n instructions and the model generate n different answers. The model's judgement ability is helpful to select the most humorous one. The process and results show that the model's performance will not be constrained by a specific template.

Algorithm 2 Inference Step of CLoST

Input: Questions Q, CLoST-trained LLM π', response number n, different prompts P shown in Figure 9.

Output: The most funny response A

1: Select n Prompts P_{sub} from P
2: $\{A_i\}_{i=0}^{n} \leftarrow \pi'(Q, P_{sub})$ $\{\pi'$ output candidate responses through $P_{sub}.\}$
3: $A_{best} \leftarrow \pi'(Q, \{A_i\}_{i=0}^{n})$ $\{\pi'$ as judgement model to select the most funny answer.$\}$
4: **return** A_{best}

```
"Please give the person who said this sentence a humorous answer.",
"Please give the person who said this sentence the most humorous answer.",
"Please give the person who said this sentence the most funny, funny, brain-opening answer.",
"Please give the person who said this sentence the most funny witty reply.",
"Please give the person who said this the most humorous and funny, brain opening  witty  reply.",
"Please give the person who said this the most humorous and funny, brain-opening answer.",\
"Answer to the funniest god.",
"Association with the funniest god.",
"Please associate with the question and give the funniest answer.",
"Please give a funny answer.",
""
```

Figure 9: Generation Prompts of CLoST.

问题	CLoST	GPT-4o	CLoT
谁会不喜欢小狗呢，大狗小狗都很爱	汪！	谁会不喜欢小狗呢？可能只有猫咪吧，它们可是“竞争对手”呢！	只要不拆家，猫我都爱！
分享一些拍立得老公	拍立得的老公？是用拍立得拍下老公然后立刻洗出来吗$(\overset{\frown}{\nabla})$	拍立得老公？那我是不是得先摇一摇，等他显影出来？	拍立得老公，不是老公，是拍立得。
刚刚更新完了游戏打了几把，刚开局还没落地就卡住了直接卡闪退了，后面又卡了几秒，怎么感觉更新后bug越来越多了	程序员：“今天的你，运气真好~	看来你的游戏角色还没学会“瞬间移动”技能，直接卡到异次元去了！	我玩的都是大逃杀游戏，不是叫你玩模拟飞行。
从一年级到现在不知道玩了几天了	你至少幼儿园没好好玩	你这是在练习“游戏学位”吧，快要毕业成“游戏博士”了！	6000天
年轻不嗨，老年痴呆。	年轻不浪，老年空想；年轻不闯，老来悲伤。	年轻不嗨，老年痴呆；所以现在嗨，老了还能当DJ！	老年痴呆，年轻嗨。
长大是什么感觉，大概是既没成为想成为的人，又丢了原来的自己，既没得到自己一心想要的东西，又失去了原来拥有的快乐。	唯一不变的，大概就是一直在变吧！	长大是什么感觉？就像升级打怪，结果发现自己选了个最难的副本，还没存档！	长大了，就是突然发现自己不是个小孩了。
持续单身，偶尔心动，惯性拒绝，保持清醒，长期快乐，祝你幸福，祝我发财。	间歇性渴望爱情，持续性懒得经营。	这简直是单身贵族的修炼手册啊！	我应该在那个瞬间选择相信你，而不是现在这样子
今天见朋友抓娃娃，你们猜怎么着？	娃娃抓了你？	今天见朋友抓娃娃，你们猜怎么着？——他抓到了一只“空气”！	我朋友抓了一只猫。
大家经历过最恐怖的事情是什么	上厕所发现没纸了	半夜肚子饿，冰箱里只有一颗生菜！	我刚刚在电梯里被偷了钱包。
可以帮我想个小说书名吗？	《催更了你还想跑》	《我在沙发上拯救世界》	《我的朋友都是大坏蛋》
晴天阴天下雨天 ,你最喜欢哪一天	法定节假日	我最喜欢放假那一天！	我最喜欢晴天，因为晴天的时候，我可以看到太阳。
请问大家放假很无聊都在做什么？	上班	放假很无聊？我在练习“葛优躺”大赛，争取拿个金牌！	我无聊的时候，就会在脑中想出一个无聊的点子，然后就真的无聊了。
求小说好句！格式：“神说，……”	神说，我有点忙，你先跟他们聊着…	神说，别担心，WiFi密码是12345678。	“神说，你们这群人类太吵了，都给我闭嘴！”
暑假大家一般几点起床？	每天被妈妈骂醒来。	下午好！	暑假起床时间：12:00
一个很懂你但他不爱你的人和一个他对你有诚挚的爱意但不太懂你”	前者是读者，后者是作者。	一个很懂你但他不爱你的人和一个他对你有诚挚的爱意但不太懂你——那就找个既懂我又爱我的，毕竟我也不想难为自己！	还是别了吧。
蚂蚁从20米高的楼落下不会死，那蚂蚁背着我落下会不会死捏？	会，因为蚂蚁不会接住你	蚂蚁：我选择辞职！	蚂蚁会死。
我歌月徘徊，我摆日更悠。	我舞影零乱，我崩半年休。	我躺床上刷手机，快乐无边际！	我走夜路更悠。
“No one but you.”你又会如何翻译这句话呢？	就，只有你啊。	除了你，谁都不行，连WiFi都不行！	“除了你，没有人。”

Figure 10: Chinese Showcases.

Failure showcase and Analysis

The creativity is uneven, and the creativity shown in the samples of training dataset varies widely. For example, in case 1, the response from CLoST is an internet meme and the response from better one using a Chinese proverb to compare. In terms of creativity, the latter have broader hands. In case 2, a shorter and antithetical answer is more witty. There are no failure examples, only in some people's preferences, it is not humorous.

Quesiton	**CLoST**	GPT-4o	CLoT
Am I wrong to be satisfied with my job?	**Job? No no.**	"Of course not! If you're satisfied with your job, you're already one step ahead of the person who invented the snooze button!"	No, you're not wrong to be satisfied with your job. In fact, you're doing just fine.
What are some bad habits that you dropped this year?	**What I gave up this year**	"I gave up my bad habit of hitting the snooze button... Now I just set 10 different alarms instead!"	I stopped asking people to explain their bad habits.
I can not help the fact that you look like what you have been through!	**"Looks like you've been through a lot too!"**	"Well, if I look like what I've been through, then I must be a mix of coffee spills, awkward Zoom calls, and a few too many midnight snacks!"	I can't help the fact that you look like what you have been through!
What's an insignificant thing that always makes you angry?	**The fact that I can't get stars for being handsome.**	"When people leave just one second on the microwave timer. It's like they enjoy living on the edge of chaos!"	A person who doesn't say "excuse me" when they bump into you.
What is a recurring dream you have had?	**"Last night's TV today."**	"I keep dreaming that I'm back in school, but this time I'm the teacher and the students are all my old bosses. Sweet, sweet revenge!"	I have a recurring dream where I'm at a party and I don't know anyone.
What's the one thing you can't dowithout, no matter the price?	**What would you say if someone asked youwhat you couldn't buy even if you had all themoney in the world?**	My morning coffee. Without it, I'm basically a sleepwalkingzombie with a to-do list!"	A sense of humor.
I haven't had a conversation like this since high school	**What were you doing in middle school, then? "**	High school? You mean back when our biggest worry was whostole our lunch money and not how to adult properly?"since high school	I haven't had a conversation like this since high school
What do you call a boomer that doesn'tget it?	**What do you cal a baby boomer whodoesn't understand the situation**	"A boomerang-they keep coming back with the samequestions!"	A boomer who doesn't get it is called a "boomer who doesn't get it."

Figure 11: English Showcases.

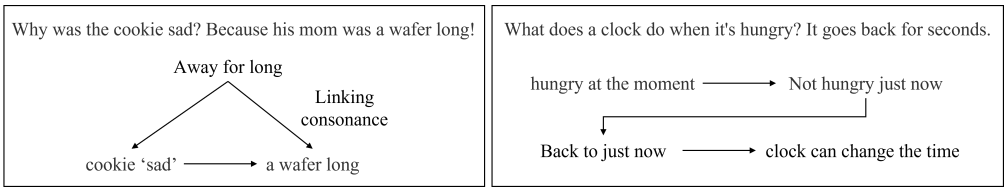


Figure 12: Generation Prompts of CLoST.

A.5 THE DETAILS OF USER STUDY

We conduct a human evaluation to validate CLoST's performance in humor generation. We choose the first 200 samples in the validation split of the Ruozhiba dataset[1] and the method mentioned above is to make the query into questin-answer pair. Then four LLMs generate responses to each question as four options. Then, we conduct a user preference study to directly verify the creativity of the LLMs. We present question and several corresponding replies, and ask users to choose the most creative and humorous responses. We select four advanced LLMs to generate responses for a total

[1] https://github.com/Leymore/ruozhiba/tree/main?tab=readme-ov-file

Table 7: Failure case based on user votes

Question	failure case	better than clost (from human)
What's your relationship like now?	Had a child.	We are like the weather forecast, sometimes sunny, sometimes rainy, forecasts can never keep up with changes.
Why does knowing too much make one an orphan of the world?	Because smart people always find it difficult to find companions, either to become rivals, or to be regarded as rivals.	Because knowing too much makes the world too small.

of 200 questions, and the four responses from the four distinct LLMs are randomly permuted in the options. We conduct an extensive survey through an online survey platform[2], ultimately collecting 15 valid questionnaires with 3000 votes. Within these collected questionnaires, we calculate the proportion of times each LLM is selected for each question. Finally, we aggregate the total number of times each LLM is chosen across all validation samples, as shown in Figure 6(c). The ratio of this sum to the overall number of selections among all LLMs signifies the user preference for each LLM. Also we calculate the win rate based on the dimension of the problem as shown in Figure 6(c).

Table 8: Humor judgement under ruozhiba dataset

CLoST	GPT-4o	QWEN1.5-32B	CLoT
95.35	76.40	68.85	50.20

Table 9: The conversation example 1 form AAIE

Question	Answer
What emperor can't abandon his harem for someone he loves?	If you have raised more than 5 cats, you will understand that love and heart labor, harem has no reason to speak, and mud is the best choice...
Why is it so hard for a emperor to manage a harem?	The emperor's management of the harem is like raising a group of cats, each with its own temper and personality. Although the emperor's position is supreme, the harem system is complex, and the concubines compete for favor like cats compete for the attention of their owners. The emperor could not pet just one cat, because it would upset the balance of the whole herd. The idiom "love bo, but the heart is hard" vividly describes the emperor's heart and energy caused by the care of many concubines, while "He mud" illustrates the helpless measure of reconciling contradictions in a complex situation. Finally, the "facemask" expression expresses helplessness and humor in network culture, which further enhances the funny effect of the answer. The emperor really worried brokenhearted ah!
Why was it so difficult for the emperors of ancient China to manage their harem?	In ancient China, it was difficult for the emperor to manage the harem, mainly because the harem system was extremely complicated, and there were many concubines with different personalities and needs. In history, Empress Wu Zetian became a female emperor through a harem struggle, and Emperor Wanli of the Ming Dynasty caused a harem struggle because she favored Princess Zheng Guifei, are typical examples. Power dynamics theory in social psychology also explains this phenomenon, where power imbalances can lead to conflict and discontent. In addition, literary works such as A Dream of Red Mansions and The Legend of Zhen Huan vividly depict the complexity of life in the harem. The emperor has to balance the interests of all parties in the back palace to avoid intensifying contradictions, just like raising a group of cats with different temperaments, which is really worried about the heart!

[2]https://www.wjx.cn/

Table 10: The conversation example 2 form AAIE.

Question	Answer
At what point did you feel you should leave your job?	”Of course, it’s due to physical reasons: my stomach isn’t good, the pancake is too big, I can’t eat it. My back isn’t good, the pot is too heavy, I can’t carry it.”
Recently, I always feel uncomfortable, especially stomach problems and low back pain, probably because of too many tasks and too much pressure. What do you think?	The physical discomfort you mentioned may indeed be related to work stress. In many countries, especially in East Asia, long working hours and high-intensity tasks are common phenomena, and this ”overwork culture” causes the body to secrete large amounts of stress hormones, such as cortisol, which affects gastrointestinal function and leads to stomach problems. At the same time, chronic stress can cause muscle tension, especially in the lower back muscles, which can trigger lower back pain. There are also many famous workaholics in history who developed health problems due to overwork, such as Thomas Edison and Winston Churchill. In order to maintain physical and mental health, it is recommended that you pay attention to adjusting the balance of work and rest, relax properly, and avoid overwork.

Table 11: Example 1 of gradually increase the prompt of information.

”What emperor can’t abandon his harem for someone he loves? If you have raised more than 5 cats, you will understand that love and heart labor, harem has no reason to speak, and mud is the best choice....” What’s funny about this answer?
Please explain in detail why the emperor cannot abandon the harem for a beloved? If you have raised more than 5 cats, you will understand that love and heart labor, harem has no reason to speak, and mud is the best choice... (Hands face). Please analyze the humor of this answer from the following aspects: 1. Explain the position of the emperor in history and the complexity of the harem system. Explain the special status of the ”beloved person” in the harem and its influence on the emperor. 3. By analogy, explain the similarities between owning a cat and managing a harem, especially the challenges that come with multiple cats. 4. Explain the idiom of ”love bo and heart labor” and explain its application in the management of the harem. 5. Explain in detail the idiom of ”he mud” and explain its role in dealing with the affairs of the harem. 6. Combined with the historical background, explain why the emperor can not abandon the harem for a beloved person. 7. Explain the use of the ”facepad” expression in Internet culture, and explain its role in expressing frustration and humor. 8. Analyze the humorous elements in the answer and explain the irony and humor.
Please explain in detail why it was so difficult for ancient emperors to manage their harem. Cultural background: Introduce the harem system in ancient China and other countries, citing relevant cultural customs and popular culture. For example, what are the similarities and differences between the harem system in ancient China and the harem system in Japan? Historical background: The analysis of specific historical events and figures, such as Wu Zetian in the Tang Dynasty and Wanli Emperor in the Ming Dynasty, illustrates the complexity of the management of the harem. Specific historical events can be cited, such as how Wu Zetian became an empress through a harem struggle. 3. Scientific principles: Apply social psychology and power dynamics theory to explain the emperor’s power distribution and interpersonal relationships in the harem. For example, how does the theory of power dynamics explain power struggles in the harem? 4. Linguistic features: Use idioms, Internet expressions and linguistic features in literary works to enhance the humorous effect of the answer. For example, how does the idiom ”love a lot, but the heart is hard” graphically describe the emperor’s situation? 5. Social phenomenon: The analogy of power struggle and family relations in modern society illustrates the universality of harem management. For example, how do power struggles in the modern workplace resemble those in ancient harem? 6. Literature and art: quote A Dream of Red Mansions, Legend of Zhen Huan and other literary works and TV dramas to analyze their depictions of the life of the harem. For example, how does the story of Zhen Huan in the Legend of Zhen Huan reflect the complexity of the harem? 7. Philosophical thought: Explore the views of different philosophical schools on power, management, balance and other issues to increase the depth of answers. For example, what does Confucianism think of the emperor’s management in the harem?

Table 12: The answer from evolutionary prompt example 1.

The joke of this answer is to compare the emperor's harem to keeping a cat, emphasizing that managing multiple objects (whether concubines or cats) requires a lot of effort, so "mixing mud" is the best option. The analogy creates a sense of humor.
The funny point of this answer is to depict the complexity and helplessness of the emperor's management of the harem by analogy with keeping a cat. First, the emperor had a supreme position in history, but the harem system was complex, involving numerous concubines and a balance of political power. Although the beloved is special in the harem, the emperor cannot abandon the entire harem for her, because it will destroy the balance of power. By analogy with owning a cat, respondents graphically illustrate the challenges of managing multiple cats (concubines), whose independence and uncontrollability are similar to the competition and complexity of harem concubines. The idiom "love bo, but the heart is hard" vividly describes the emperor's heart and energy caused by the care of many concubines, while "He mud" illustrates the helpless measure of reconciling contradictions in a complex situation. Finally, the "facemask" expression expresses helplessness and humor in network culture, which further enhances the funny effect of the answer. On the whole, this answer, through the combination of historical background, idioms and Internet culture, vividly shows the frustration and humor of the emperor's management of the harem.
The reason why it is so difficult for the emperor to manage the harem must first be seen from the cultural background. The harem system of ancient China was very complicated and different from that of other countries. For example, the Japanese harem system had concubines, but there were far fewer power struggles than in China. This cultural difference made it more challenging for Chinese emperors to manage their harem. From the historical background, Wu Zetian of the Tang Dynasty was not only the emperor's concubine, but also eventually became the only female emperor in Chinese history through the struggle of the harem. Her success not only shows the intensity of the struggle in the harem, but also reflects the difficult situation of the emperor in managing the harem. The emperor Wanli of the Ming Dynasty, because he favored the Concubine Zheng, triggered a fierce battle in the harem, and even affected the government. From the perspective of scientific principles, the theory of power dynamics and interpersonal relations in social psychology can explain the emperor's situation in the harem. The power dynamic theory holds that the distribution and use of power will affect the stability of interpersonal relations. Although the emperor was supreme, in the face of the different needs and personalities of many concubines, he had to carefully balance the interests of all parties and avoid exacerbating contradictions. For example, research has shown that an imbalance of power in a group can lead to conflict and discontent, especially in the harem. In terms of linguistic features, the idiom "love bo and heart strain" vividly describes the emperor's heart exhaustion caused by the care of many concubines, while "He mud" describes the helpless act of reconciling contradictions in a complex situation. The expression of "facemask" in the Internet expression expresses helplessness and humor in the Internet culture, which further enhances the funny effect of the answer. For example, when the emperor faced the concubines competing for favor, he may feel helpless and can only "cover his face" to show that he is speechless. In terms of social phenomena, the power struggle in the modern workplace has many similarities to the ancient harem. Bosses in the workplace need to balance relationships between subordinates and avoid conflicts caused by taking sides. This is very similar to the situation of the emperor in the harem. In literature and art, the Grand View Garden in A Dream of Red Mansions and the battle between the harem in the Legend of Zhen Huan vividly depict the complexity of harem life and the difficult situation of the emperor. For example, in the Biography of Zhen Huan, the process of Zhen Huan's gradual transformation from an ordinary concubine to an empress is full of intrigues and struggles, fully demonstrating the complexity of life in the harem. Finally, from the perspective of philosophical thought, Confucianism emphasizes the "golden mean", that is, to maintain balance in management. However, it was very difficult for the emperor to achieve this in the imperial palace, because each concubine had its own needs and personalities, and it was difficult to achieve complete fairness and balance. In short, the emperor's management of the harem was like raising a group of cats, each with its own temper and personality. Although the emperor was supreme, the harem system was complex, with concubines competing for favor like cats competing for their master's attention. For example, some cats like to be spoiled, some cats like to be alone, and some cats like to compete for attention. The emperor could not pet just one cat, because it would upset the balance of the whole herd. The emperor is really worried!

Table 13: Example 2 of gradually increase the prompt of information.

At what point did you feel you should leave your job? Of course, it's due to physical reasons: my stomach isn't good, the pancake is too big, I can't eat it. My back isn't good, the pot is too heavy, I can't carry it. What is the punchline of the answer?
Please explain in detail why you have been feeling under the weather lately, especially stomach problems and low back pain. Please answer with the following information: 1. Cultural background: Reference to work culture and stress management methods of different countries and regions, combined with humor elements in popular culture. 2. Historical background: References to famous workaholics or historical figures with health problems caused by work stress. 3. Scientific principles: Explain the effects of stress on the body, especially on the stomach and waist, citing medical research. 4. Linguistic features: Use idioms, proverbs and humorous expressions to enhance the interest and cultural depth of the answers. 5. Other relevant in-depth information: Combining psychological and sociological theories to explain the impact of work stress on psychological and social relationships.

Table 14: The answer from evolutionary prompt example 2.

The joke of this answer is the use of metaphors and puns. Respondents used "bad stomach, cake is too big, can not eat" and "bad waist, pot is too heavy, can not move" to express the image of work pressure, heavy tasks, their own unbearable situation.Comparing work tasks to "pancakes" and "pans", graphically indicating that the task is too heavy.On the surface, it is a physical health problem (bad stomach, bad waist), in fact, it is to express the frustration of work pressure and burden.
Recently, I always feel uncomfortable, especially stomach problems and low back pain, probably because you work too much, too much pressure. Let's look at the problem from several angles. 1. Cultural background: In many countries, especially in East Asia, the work culture emphasizes long working hours and high-intensity tasks, which is called "overwork culture". In Japan, there is a phenomenon called "Karoshi," which refers to sudden death caused by overwork. Work stress is also often teased in popular culture, such as The US TV series "The Office", where characters often produce a variety of funny plots because of work pressure. 2. Historical background: There are many famous workaholics in history, such as the American inventor Thomas Edison, who slept only four hours a day, which eventually led to health problems. Then there was British Prime Minister Winston Churchill, who developed a severe stomach bug from the stress of his job during World War II. 3. Scientific principles: Medical research shows that stress causes the body to secrete a large amount of stress hormones, such as cortisol, which affect gastrointestinal function and lead to stomach problems. At the same time, chronic stress can cause muscle tension, especially in the lower back muscles, which can trigger lower back pain. 4. Linguistic features: Some idioms and proverbs can be used to describe this situation, such as "the pie is too big to eat", so many tasks give you indigestion; "The pot is too heavy, the back can't move", the pressure makes your back ache. These expressions are not only vivid, but also make the answer more interesting. 5. Other relevant in-depth information: Psychological research shows that long-term work stress can lead to anxiety and depression, affecting mental health. At the same time, sociological research has pointed out that overwork affects individuals' social relationships, leading to family conflicts and social isolation. Therefore, it is very important to strike a balance between work and rest. In short, too much stress at work can affect not only physical health, but also psychological and social relationships. We should pay attention to the balance between work and rest, and relax properly in order to maintain physical and mental health. '