YNU-HPCC at SemEval-2026 Task 1: Constraint-Aware In-Context Learning for Multilingual Humor Generation

Xulong Zhang, Jin Wang and Xuejie Zhang
School of Information Science and Engineering
Yunnan University
Kunming, China
xulongzhang@stu.ynu.edu.cn, {wangjin, xjzhang}@ynu.edu.cn

Abstract

This paper describes the system developed by the YNU-HPCC team for SemEval-2026 Task 1 (Humor Generation). The task aims to generate humorous texts from given news headlines or from two unrelated words. The core challenge lies in enabling Large Language Models (LLMs) to understand human humor and align with specific humorous styles. We investigated two approaches: fine-tuning with Proximal Policy Optimization (PPO) and in-context learning with LLMs. We also employed Qwen-Max to evaluate the quality of the generated texts. In the PPO experiments, we constructed a hybrid reward model to align with humor. For our final submission based on LLMs, we used multiple advanced LLMs, along with customized few-shot prompts and a small set of gold samples, to effectively guide the models in generating jokes that resonate with human humor. Experimental results show that our system achieves competitive performance, ranking 4th in the English track, 2nd in the Chinese track, and 2nd in the Spanish track.

1 Introduction

Humor is a complex cognitive and linguistic phenomenon that relies on incongruity, context, and timing to evoke amusement (Mihalcea and Strapparava, 2005; Yang et al., 2015). Generating humorous text is significantly more challenging than standard text generation, as it requires moving beyond semantic accuracy to master stylistic nuances such as sarcasm, irony, and wit (Loakman et al., 2023). The true essence of a joke often lies in the unexpected twist or punchline, making the automated generation of jokes a frontier challenge in Natural Language Processing (NLP) (Hossain et al., 2019).

In previous computational humor research (Tomasulo et al., 2020), early methods relied heavily on template-based generation or shallow recurrent neural networks, which struggled to maintain long-range coherence or produce genuine surprise (Holtzman et al., 2020). With the advent of LLMs, researchers have increasingly adopted Reinforcement Learning from Human Feedback (RLHF) using algorithms like PPO to align model outputs with human preferences (Ouyang et al., 2022). Relying on scalar reward models often leads to reward hacking and mode collapse (Kim et al., 2025), where the model might repeat generic, high-scoring templates rather than generating diverse, context-specific wit.

Meanwhile, recent work suggests that humorous generation often benefits from mental leaps rather than strictly linear reasoning: the Leap-of-Thought (LoT) paradigm (Zhong et al., 2024) highlights creative semantic association beyond conventional chain-of-thought prompting (Wei et al., 2023).

To address these challenges, our team proposed a comprehensive humor generation system for SemEval-2026 Task 1 (Castro et al., 2026). Our approach initially explored a fine-tuning scheme based on PPO, constructing a hybrid reward model to align multilingual humor preferences, and then integrated a robust In-Context Learning (ICL) method (Li et al., 2024). This subsequent approach leverages the emergent reasoning capabilities of advanced LLMs (Brown et al., 2020), supplemented by a few-shot prompt that features a specific comedic persona to establish and effectively regulate the final humorous tone (Deshpande et al., 2023).

The rest of the paper is organized as follows. First, we present two methods for multilingual humor generation. Then, we evaluate them on the English, Spanish, and Chinese tracks. Finally, we analyze the results and conclude.

2 System Overview

This section introduces the two primary methods implemented for the multilingual humor generation

Proceedings of the The 20th International Workshop on Semantic Evaluation (2026), pages 664–670

Data Source	Count	Processing Strategy	Contribution
Human	100	Upsampling	500
V2 (Qwen)	3,400	Full retention	3,400
V2 (DeepSeek)	3,400	Full retention	3,400
V1 original	3000	Top 500 filter	500
Total	-	-	7800

Table 1: Construction strategy and composition of the V13 dataset.

task and analyzes how each framework operates.

2.1 Method 1: SFT and PPO

Our initial approach formulated humor generation as an alignment problem via RLHF. To achieve this, we developed a comprehensive pipeline encompassing Supervised Fine-Tuning (SFT), Reward Modeling (RM), and PPO.

2.1.1 SFT Data Engineering

The generation of high-quality humorous text is heavily dependent on the scale and diversity of the training data (Zhou et al., 2023). The construction utilized a tiered mixture strategy, as summarized in Table 1:

Human Anchors: We manually curated 100 high-quality human jokes from the Chinese dataset and upsampled them 5 times to serve as stylistic anchors for native Chinese humor.

Heterogeneous Synthetic: We directly utilized the aforementioned 100 high-quality human jokes as few-shot prompt demonstrations to guide both Qwen3-Max and DeepSeek-V3.1 in generating the complete set of 3,400 original task entries independently (Tunstall et al., 2023).

Filtered Baseline: This batch is the product of our early direct generation using Qwen3-Max on the 3,400 raw entries. Although these outputs lacked humor and overall text quality, we intentionally retained 500 of them as effective "penalty" options for subsequent PPO reward model training (Touvron et al., 2023).

2.1.2 Reward Model with Tier-Based Pairwise Ranking

To train a reward model capable of discerning comedic nuances without expensive manual annotation, we exploited the natural quality tiers of our V13 dataset to automatically construct preference pairs, each of which contains a chosen and rejected sample:

Stylistic Alignment: Human > Synthetic to guide the model in capturing the semantic reversals and sarcastic features unique to human expression.

Logical Optimization: Synthetic > Filtered to encourage the model to prefer comedic structures with valid setups and punchlines, suppressing flat, declarative outputs.

Hard Constraint Injection: Filtered > Violating Negatives. Even if filtered texts are comedically mediocre, as long as they satisfy the hard constraints of keyword inclusion and topical relevance, they must strictly outscore any invalid generations that miss keywords or hallucinate off-topic (Chen et al., 2025).

For the reward model backbone, we selected the Encoder-only mdeberta-v3-base over a Decoder model. Its bidirectional attention mechanism proved highly effective in understanding the contextual correlation between the prompt and the joke.

2.1.3 Construction of the Hybrid Reward System

During the exploratory PPO phase, to guide the model in generating high-quality humorous text while strictly adhering to task constraints, we initially conceptualized a hybrid reward system, theoretically defined as:

$$R_{\text{total}} = R_{\text{quality}} + R_{\text{rule}} \tag{1}$$

Here, R_{quality} is the quality score, directly output by our previously described mDeBERTa model and representing the neural network's comprehensive evaluation of the joke's humor and fluency. R_{rule} is the hard rule constraints that address the phenomena of *instruction ignoring* and *text degeneration*, which are highly prevalent during LLM reinforcement learning fine-tuning. We introduced deterministic regex-based penalties. For instance, if the model falls into a repetitive loop or omits task-mandated specific vocabulary, a severe penalty is directly applied to ensure the outputs meet the baseline of compliance.

To prevent the policy model (Actor) from experiencing catastrophic forgetting or collapsing text fluency while attempting to maximize R_{total}, we incorporated a standard Kullback-Leibler (KL) divergence penalty into the PPO training (Ziegler et al., 2020). Specifically, we froze the original SFT model as a Reference Model. We calculated the distributional shift between the current policy π_θ and the reference policy π_{ref} during generation.

$$R_{\text{final}} = R_{\text{total}} - \beta \log \frac{\pi_\theta(y \mid x)}{\pi_{\text{ref}}(y \mid x)} \tag{2}$$

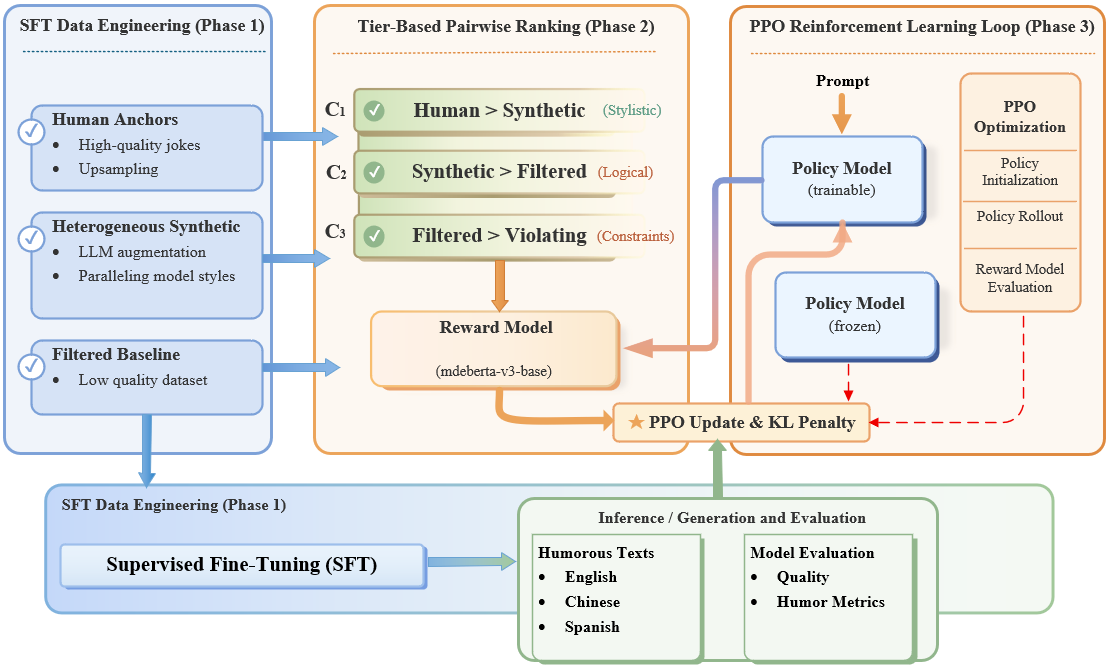

Figure 1: PPO-based Alignment Procedure for Humor Generation

where R_{final} refers to the ultimate reward signal used to update the weights of the policy model, R_{total} represents the previously defined hybrid score encompassing humor quality and rule constraints,, and β denotes the KL divergence penalty coefficient, which prevents the model from sacrificing fundamental multilingual fluency in pursuit of higher rewards. Additionally, $\pi_\theta(y \mid x)$ refers to the probability of the current policy model generating the humorous text y given the news headline or two unrelated words x. In contrast, $\pi_{\text{ref}}(y \mid x)$ represents the probability of the frozen reference model (fine-tuned on the V13 hybrid dataset) generating the same text.

By dynamically tuning the KL penalty coefficient β, we sought to balance exploration of humorous expressions with the linguistic priors acquired during SFT.

2.2 Method 2: Two-Stage Dynamic Few-Shot Generation via GPT-5.2

Whether it is the deadpan delivery typical of English dry humor or the vibrant situational irony in Spanish, eliciting a genuine laugh heavily relies on native-level cultural intuition, slang vocabulary, and punchline rhythm. Small to medium open-source models often struggle with a rigid translation feel when handling non-native humor, compounded by the convergence instability of reinforcement learning in practical deployment. Consequently, we adjusted our strategy. Recognizing that GPT-5.2 exhibits superior native linguistic intuition and cultural understanding in English and Spanish compared to models like Qwen or DeepSeek, we leveraged its powerful few-shot reasoning capabilities alongside a specialized prompting framework for the final generation.

Stage 1: We used GPT-5.2 to construct a "Golden Library" containing 120 top-tier exemplars for both English and Spanish. During construction, we strictly controlled the ratio of task types: each language pool consists of 60 humorous texts generated from news headlines and 60 from two unrelated keywords. This 1:1 balanced distribution ensures the model maintains robust generative capabilities for both tasks, preventing performance degradation or task bias during large-scale data processing. The instructions employ system prompts to mandate a sharp, satirical narrative tone, utilizing manually curated one-shot anchors as stylistic benchmarks. Through in-context learning, the model naturally adopts the narrative pacing of immediate scene conflict found in the exemplars, ensuring high-quality comedic tension and structural diversity.

Stage 2: Building upon the established golden library, we implemented a dynamic sampling strategy for full-scale inference with GPT-5.2. For each test instance, the program randomly samples three records from the corresponding language library to

serve as contextual demonstrations. This approach prevents the homogenization of output typically caused by fixed prompt templates. By exposing the model to diverse entry points and comedic structures in every cycle, we effectively maximize the diversity of the generated text while maintaining a consistent satirical tone.

3 Results and Analysis

3.1 Official and Automated Evaluation Framework

According to the official evaluation protocol for SemEval-2026 Task 1, final rankings are determined by Human Preference Judgments. This mechanism employs a Pairwise Comparison model, in which human evaluators compare two generated texts produced under identical constraints and select the more humorous entry. The systems are ranked using an Elo-based leaderboard. Given the highly subjective nature of human humor preferences and the limited human resources of a single participating team, it is difficult to organize large-scale manual evaluations frequently during the development phase. Therefore, we designed an automated proxy evaluation framework consisting of Hard Constraint Verification and LLM-as-a-Judge Blind Testing.

Constraint Satisfaction Rate (CSR): Serves as an objective baseline metric to evaluate the models' instruction-following capabilities; the calculation accounts for the heterogeneity of task inputs. Specifically, the metric primarily verifies whether the generated texts strictly adhere to the official language-specific length limits, namely 300 characters for Chinese and 900 for English and Spanish. Building on this, for the two-word task, the system employs regular expressions and lemmatization to ensure that the given unrelated words are fully included. Conversely, for the news headline task, since high-quality humor relies on the satirical extension of the background event rather than mechanical repetition of the prompt, the evaluation of topical deviation is decoupled from rigid string matching and delegated to the downstream LLM-as-a-judge for semantic relevance assessment.

LLM-as-a-Judge Blind: To simulate the official arena environment and mitigate the preference bias of any single model, we employed an ensemble of advanced large language models, including Qwen and DeepSeek, as independent judges for blind A/B testing. For identical inputs, the judges perform a forced choice (Win-Rate) between texts generated by different systems. Regarding the evaluation dimensions for the news headline task, the judges not only measure the co-medic punchline impact but also heavily weigh contextual integration. This ensures that the humor emerges naturally from the news event itself, rather than being an awkwardly stitched, unrelated joke. Furthermore, to eliminate position bias, all comparative pairs are bidirectionally randomized before being evaluated by the judge models (Zheng et al., 2023).

3.2 Constraint Satisfaction Rate Analysis

To quantify the instruction-following capabilities of different methods in a multilingual environment, we summarize the system performance in Table 2. This table illustrates the comprehensive performance of each approach across the English, Spanish, and Chinese tracks, distinguishing between the Length Compliance and the Keyword Inclusion Rate specifically for the two-word task.

The Baseline in this evaluation is established using the vanilla Qwen2.5-3B-Instruct model without any additional humor-specific tuning. This baseline represents the model's raw potential for instruction following. As evidenced by its perfect scores across all tracks, the base model possesses an innate ability to adhere to character limits and keyword constraints.

However, it is crucial to emphasize that the metrics in Table 2 represent purely objective hard constraints. High CSR scores indicate a model's obedience to the prompt's structural requirements but do not necessarily reflect the qualitative funniness or the comedic impact of the generated punchlines. A model might achieve a perfect score by mechanically inserting keywords into a dry sentence, thereby failing the ultimate goal of the humor generation task.

3.3 Analysis of Win-Rate Results

Figure 2 illustrates the Win-Rate performance across three language tracks for three representative system pairings.

Method 2 vs Method 1: The system driven by Method 2 consistently and significantly outperforms the system aligned via Method 1. Specifically, under the DeepSeek-V3.2 judge, the win rates for Method 2 are 0.9400 in English, 0.9650 in Spanish, and 0.9283 in Chinese. These results provide strong evidence that, when evaluated using the combined criteria of comedic impact and con-

	EN		ES		ZH		avg	
System	L	K	L	K	L	K	L	K
Baseline	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Method 1	100.00	88.00	100.00	68.00	100.00	44.00	100.00	66.67
Method 2	100.00	96.00	100.00	100.00	100.00	100.00	100.00	98.67

Table 2: Results of different methods on the humor generation task across different languages with Length Compliance (L) and Keyword Inclusion (K) scores. Language codes: English (EN), Spanish (ES), and Chinese (ZH).

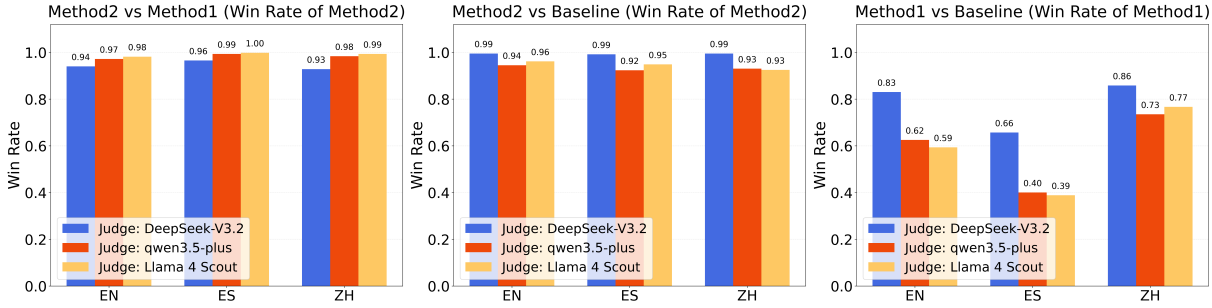

Figure 2: LLM-as-a-Judge Blind Win-Rate Comparison Across Three Languages and Three Judges

textual integration, outputs generated by Method 2 are widely recognized as both more humorous and more intrinsically linked to the input events. In contrast, outputs from Method 1 are frequently flagged for insufficient context fusion or weak comedic delivery.

Method 2 vs Baseline: The advantage of Method 2 remains robust when compared against the original Qwen2.5-3B-Instruct baseline model. Qwen3.5-Plus and Llama 4 Scout maintain a high and consistent preference for Method 2, with win rates ranging from 0.9233 to 0.9450 and 0.9250 to 0.9616, respectively. These findings reinforce the core conclusion presented in Section 3.2: while a baseline model can achieve a perfect constraint satisfaction rate by strictly following instructions, such obedience does not equate to producing high-quality humor in preference-based evaluations.

Method 1 vs Baseline: While Method 1 yields marginal gains over the baseline, it exhibits significant performance volatility across languages and evaluators. In the Spanish track, the win rate for Method 1 is the lowest across all evaluations, ranging from 0.38 to 0.65. Given that Method 1 was fine-tuned on the Qwen2.5-3B model, a small-parameter model is unlikely to simultaneously master sophisticated semantic expression and culture-specific humor logic across multiple languages with limited training resources. A 3B model may fail to produce high-quality humor in a non-native language like Spanish, even after extensive reinforcement learning, aligns with expectations about model capacity limits.

Furthermore, these results confirm that humor perception is highly subjective. To address this, we specifically employed three distinct large language models as judges to minimize the influence of individual model preferences and ensure the objectivity of our findings. This diversity in evaluation not only highlights the differing training backgrounds of the models but also reveals the persistent challenge of establishing a universal automated metric for humor in constrained generation tasks.

4 Conclusion

This paper develops a multi-stage humor generation system for SemEval-2026 Task 1, integrating both PPO-based reinforcement learning and dynamic ICL strategies. Experimental results demonstrate that our system maintains a superior CSR, with Method 2 consistently achieving an average win rate of over 94% against the baseline across all linguistic tracks and evaluators. While Method 1 exhibited some volatility in cross-lingual transfer due to model capacity limits, the overall framework proved highly effective and competitive. Future research will focus on enhancing the native humorous expression of small-parameter models in non-native contexts and refining automated evaluation metrics to better account for cross-cultural nuances.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (NSFC) under Grant Nos. 61966038 and 62266051. The authors would like to thank the anonymous reviewers for their constructive comments.

References

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, and 12 others. 2020. Language models are few-shot learners. *Preprint*, arXiv:2005.14165.

Santiago Castro, Luis Chiruzzo, Santiago Góngora, Salar Rahili, Naihao Deng, Ignacio Sastre, Victoria Amoroso, Guillermo Rey, Aiala Rosá, Guillermo Moncecchi, J. A. Meaney, Juan José Prada, and Rada Mihalcea. 2026. SemEval-2026 Task 1: MWAHAHA, Models Write Automatic Humor And Humans Annotate. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*.

Shen Chen, Jin Wang, and Xuejie Zhang. 2025. YNU-HPCC at SemEval-2025 task3: Leveraging zero-shot learning for halluciantion detection. In *Proceedings of the 19th International Workshop on Semantic Evaluation (SemEval-2025)*, pages 28–33, Vienna, Austria. Association for Computational Linguistics.

Ameet Deshpande, Vishvak Murahari, Tanmay Rajpurohit, Ashwin Kalyan, and Karthik Narasimhan. 2023. Toxicity in chatgpt: Analyzing persona-assigned language models. *Preprint*, arXiv:2304.05335.

Ari Holtzman, Jan Buys, Li Du, Maxwell Forbes, and Yejin Choi. 2020. The curious case of neural text degeneration. *Preprint*, arXiv:1904.09751.

Nabil Hossain, John Krumm, and Michael Gamon. 2019. "president vows to cut <taxes> hair": Dataset and analysis of creative text editing for humorous headlines. In *Proceedings of the 2019 Conference of the North*, pages 133–142. Association for Computational Linguistics.

Sunghwan Kim, Dongjin Kang, Taeyoon Kwon, Hyungjoo Chae, Dongha Lee, and Jinyoung Yeo. 2025. Rethinking reward model evaluation through the lens of reward overoptimization. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 13252–13280, Vienna, Austria. Association for Computational Linguistics.

Chong Li, Shaonan Wang, Jiajun Zhang, and Chengqing Zong. 2024. Improving in-context learning of multilingual generative language models with cross-lingual alignment. *Preprint*, arXiv:2311.08089.

Tyler Loakman, Aaron Maladry, and Chenghua Lin. 2023. The iron(ic) melting pot: Reviewing human evaluation in humour, irony and sarcasm generation. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 6676–6689, Singapore. Association for Computational Linguistics.

Rada Mihalcea and Carlo Strapparava. 2005. Making computers laugh: Investigations in automatic humor recognition. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 531–538, Vancouver, British Columbia, Canada. Association for Computational Linguistics.

Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. 2022. Training language models to follow instructions with human feedback. *Preprint*, arXiv:2203.02155.

Joseph Tomasulo, Jin Wang, and Xuejie Zhang. 2020. Ynu-hpcc at semeval-2020 task 7: Using an ensemble bigru model to evaluate the humor of edited news titles. In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pages 871–875. International Committee for Computational Linguistics.

Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, Dan Bikel, Lukas Blecher, Cristian Canton Ferrer, Moya Chen, Guillem Cucurull, David Esiobu, Jude Fernandes, Jeremy Fu, Wenyin Fu, and 49 others. 2023. Llama 2: Open foundation and fine-tuned chat models. *Preprint*, arXiv:2307.09288.

Lewis Tunstall, Edward Beeching, Nathan Lambert, Nazneen Rajani, Kashif Rasul, Younes Belkada, Shengyi Huang, Leandro von Werra, Clémentine Fourrier, Nathan Habib, Nathan Sarrazin, Omar Sanseviero, Alexander M. Rush, and Thomas Wolf. 2023. Zephyr: Direct distillation of lm alignment. *Preprint*, arXiv:2310.16944.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc Le, and Denny Zhou. 2023. Chain-of-thought prompting elicits reasoning in large language models. *Preprint*, arXiv:2201.11903.

Diyi Yang, Alon Lavie, Chris Dyer, and Eduard Hovy. 2015. Humor recognition and humor anchor extraction. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 2367–2376. Association for Computational Linguistics.

Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zi Lin, Zhuohan Li, Dacheng Li, Eric P. Xing, Hao Zhang, Joseph E. Gonzalez, and Ion Stoica. 2023. Judging llm-as-a-judge with mt-bench and chatbot arena. *Preprint*, arXiv:2306.05685.

Shanshan Zhong, Zhongzhan Huang, Shanghua Gao, Wushao Wen, Liang Lin, Marinka Zitnik, and Pan Zhou. 2024. Let's think outside the box: Exploring leap-of-thought in large language models with creative humor generation. *Preprint*, arXiv:2312.02439.

Chunting Zhou, Pengfei Liu, Puxin Xu, Srini Iyer, Jiao Sun, Yuning Mao, Xuezhe Ma, Avia Efrat, Ping Yu, Lili Yu, Susan Zhang, Gargi Ghosh, Mike Lewis, Luke Zettlemoyer, and Omer Levy. 2023. Lima: Less is more for alignment. *Preprint*, arXiv:2305.11206.

Daniel M. Ziegler, Nisan Stiennon, Jeffrey Wu, Tom B. Brown, Alec Radford, Dario Amodei, Paul Christiano, and Geoffrey Irving. 2020. Fine-tuning language models from human preferences. *Preprint*, arXiv:1909.08593.