CDFG: ENHANCING CHAIN-OF-THOUGHT DISTILLATION WITH FEEDBACK

Lingzhi Gao [1] Xuan Wang [2] Tianrun Cai [3] Xiunao Lin [4] and Chao Wu [1]

[1] Zhejiang University, Hangzhou, China
[2] CHINA MEDIA GROUP, Beijing, China
[3] University of Manchester, Manchester, UK
[4] Zhejiang Post & Telecommunication Construction Co. Ltd, Hangzhou, China

ABSTRACT

Chain-of-thought (CoT) prompting has shown great potential in enhancing the reasoning capabilities of large language models (LLMs), and recent studies have explored distilling this ability into smaller models. However, existing CoT distillation methods often overlook student model errors as valuable learning signals. In this paper, we propose CDFG, a two-stage distillation framework that treats model errors as opportunities for improvement. After an initial imitation-based training phase, the teacher model analyzes the student's incorrect outputs and generates natural language feedback that highlights reasoning flaws and suggests correction strategies. The student model is then retrained using this guided input. Experiments on several mathematical reasoning benchmarks demonstrate that CDFG consistently improves student model performance. Our results show that incorporating feedback-driven learning into CoT distillation can enhance reasoning accuracy.

KEYWORDS

Chain-of-thought distillation, Large language model, Reasoning

1. INTRODUCTION

Large language models (LLMs) [1, 2] have demonstrated extraordinary performance across a wide array of natural language processing (NLP) tasks [3, 4]. Trained on massive corpora of internet-scale text, these models exhibit remarkable capabilities in language understanding, generation, and even multi-step reasoning. Their success has fundamentally changed how researchers and practitioners approach tasks ranging from machine translation to mathematical problem solving [5]. However, their practical deployment is often constrained by substantial computational costs, memory requirements, and latency, particularly in resource-limited or real-time scenarios. This limitation has motivated efforts to distil the reasoning capabilities of large teacher models into smaller and more efficient student language models. This process is commonly referred to as knowledge distillation [6, 7].

Among the various prompting strategies developed for LLMs, Chain-of-Thought (CoT) [8] prompting has emerged as a particularly effective tool for enhancing reasoning performance. By encouraging the model to articulate intermediate steps before arriving at a final answer, CoT prompting improves both accuracy and interpretability. It helps expose the model's underlying logic, making the inference process more transparent and often more reliable [9, 10]. Encouraged by the success of CoT prompting, a growing body of work has explored CoT distillation—that is, training smaller models to imitate the reasoning traces generated by larger LLMs. This technique

David C. Wyld et al. (Eds): CSITEC, NLPI, CMCA, NECOM, DMML, SP, ADCO – 2026
pp. 139-149, 2026. CS & IT - CSCP 2026 DOI: 10.5121/csit.2026.160812

enables compact models to perform multi-step reasoning without needing the full capacity of large models during inference [11, 12].

Despite its promise, existing CoT distillation approaches face several notable limitations. A common assumption in these methods is that directly mimicking the teacher's reasoning traces is sufficient for the student model to internalize complex logical patterns. However, in practice, student models often capture only the surface patterns of CoT outputs, such as step formatting or sequence structure, without fully understanding the underlying reasoning logic [13, 11]. More importantly, current CoT distillation methods treat all examples equally, regardless of whether the student model finds them easy or difficult. They lack mechanisms for leveraging the student model's mistakes as learning opportunities. In human learning, mistakes serve not only as failures but also as valuable diagnostic signals. Experienced educators analyze students' errors to identify misunderstandings, provide corrective feedback, and suggest strategies for improvement. This cycle of attempt, feedback, and revision plays a critical role in fostering deep and transferable understanding. However, such dynamics are largely absent from existing distillation pipelines, which typically involve a single-pass imitation process without explicit reflection or revision.

In response to these limitations, we propose a novel approach called CDFG (Chain-of-Thought Distillation with Feedback Guidance), which introduces a human-inspired, feedback-driven learning paradigm into the CoT distillation process. Rather than relying solely on direct imitation of teacher outputs, CDFG integrates reflective learning into the training loop by drawing on the model's own errors. The core intuition is simple: by identifying where a student model goes wrong, and then using a teacher model to analyze those errors and offer targeted advice, we can help the student model not just learn what is correct, but why it is correct, and how to avoid similar mistakes in the future.

The CDFG framework operates under the assumption that LLMs possess not only strong generative abilities but also sufficient language and reasoning capabilities to provide useful feedback on incorrect inferences. It also assumes that student language models have a basic level of language understanding. Our method leverages the teacher model's explanatory capacity to transform the student model's incorrect predictions into rich, context-specific suggestions. These suggestions are not fixed templates or symbolic rules, but dynamically generated natural language strategies tailored to each specific error case. This enables the feedback to capture reasoning flaws and offer more flexible and comprehensible guidance, helping the student model better understand the reasoning process.

To sum up, our main contributions are:

- We propose CDFG, a two-stage distillation method that mimics the human teaching paradigm of learning from mistakes. It first generates CoT demonstrations and then provides targeted feedback based on the student's errors, enabling more effective retraining and improved reasoning.

- Our CDFG employs a feedback-guided input augmentation strategy. The teacher's suggestions are combined with the original question, guiding the student model toward more accurate reasoning.

- We evaluate the effectiveness of our method in improving reasoning performance of small language models through distillation on a mathematical benchmark dataset.

2. RELATED WORK

2.1. Chain-of-Thought Prompting

Chain-of-Thought (CoT) prompting was introduced as a method to explicitly elicit multi-step reasoning from large language models [8]. Unlike direct answer generation, CoT encourages the model to generate intermediate steps, leading to better performance on tasks that require logical deduction, arithmetic computation, or commonsense reasoning. This approach mitigates issues such as reasoning shortcuts and hallucinated answers, making the model's decision-making process more interpretable and reliable. Following this insight, numerous extensions have emerged. For large language models, even in the zero-shot setting, simple natural language prompts like "Let's think step by step" can significantly improve reasoning quality [14]. A Self-Consistency strategy that generates multiple reasoning paths and selects the final answer via voting, increasing accuracy and robustness [15]. Moreover, CoT has been successfully applied to other domains, including program synthesis and task planning [10].

Further developments have diversified the CoT prompting paradigm. Least-to-Most prompting breaks down complex tasks into simpler subproblems, facilitating gradual understanding. Tree-of-Thoughts [16] implements a tree-based search over multiple reasoning trajectories, enabling more strategic and diverse exploration of the solution space. Auto-CoT [17] automates the generation of CoT examples using self-constructed prompts, reducing reliance on manual labeling. These advancements highlight CoT's growing influence as a foundation for improving LLM reasoning across domains.

2.2. Cot Distillation

Despite its effectiveness, CoT prompting often relies on large, resource-intensive language models. This limits its usability in settings with constrained computational resources. To bridge this gap, recent works have explored CoT distillation, aiming to transfer the reasoning ability of large teacher models to smaller, more efficient student models through supervised learning on generated CoT traces [6].

Hsieh et al. [12] proposed a multi-task distillation framework where the student learns from both final answers and intermediate rationales generated by the teacher model. The method achieves strong performance even with limited data and outperforms larger models in some tasks. KPDD [11] introduced a key-point-driven CoT distillation method that separates reasoning into semantic key points and computation steps, using targeted loss functions to address different types of reasoning errors. DCD [18] proposed Distillation Contrastive Decoding, which integrates contrastive CoT prompting with distillation techniques, enabling improved inference without requiring a separate amateur model. Wadhwa et al. [19] analyzed the mechanics of CoT-augmented distillation and found that placing CoT sequences after target labels leads to better downstream performance—even when these rationales are shuffled or partially removed—thus proposing a more efficient rationale injection method. Li et al. [20] introduced a Mixed Distillation framework that jointly distills both Chain-of-Thought and Program-of-Thought signals, using multi-task learning to significantly boost small model performance on complex multi-path reasoning tasks. KARD [21] incorporated external knowledge into the distillation process for knowledge-intensive tasks, enhancing the quality of student-generated reasoning paths through structured knowledge guidance. The success of this approach also shows that incorporating additional information during CoT distillation is a feasible strategy.

Existing CoT distillation methods have several limitations. They usually rely on a single-pass training process, which overlooks the valuable feedback that can be gained from student errors. Additionally, these methods often fail to include mechanisms for error-based correction or difficulty-aware training. Most importantly, they tend to emphasize output imitation rather than promoting a deeper understanding of the reasoning process. To address this gap, we propose CDFG, a feedback-driven, two-stage distillation framework. By leveraging student errors and incorporating teacher-provided suggestions, CDFG enables more targeted and effective reasoning transfer, offering a promising direction for future research in CoT distillation.

3. METHOD

3.1. Stage One: Basic Chain-of-Thought Distillation

In the first stage of the proposed CDFG framework, conventional CoT distillation is performed using a teacher–student paradigm. Given a training dataset $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{N}$, where x_i denotes a natural language input (e.g., a question) and y_i the corresponding ground-truth answer, the goal is to train a student language model S to replicate the reasoning capabilities of a large teacher model T.

For each training instance, the teacher model generates a detailed CoT reasoning sequence $\$c_i = T(x_i)$, which includes both intermediate reasoning steps and the final answer y_i. These CoT sequences are collected to form an augmented dataset:

$$\mathcal{D}_{\text{cot}} = \{(x_i, c_i)\}_{i=1}^{N} \tag{1}$$

The student model is then trained on$\mathcal{D}_{\text{cot}}$using teacher forcing and a standard sequence-level cross-entropy loss. Specifically, the objective is to minimize the discrepancy between the student's output $S(x_i)$ and the target CoT sequence c_i:

$$\mathcal{L}_1 = \frac{1}{N}\sum_{i=1}^{N} \text{CE}(S(x_i), c_i) \tag{2}$$

where \mathrm{CE} $\vdash(\cdot \dashv$) denotes the token-level cross-entropy loss. Upon completion of this stage, the student model acquires an initial ability to mimic the teacher's reasoning process, establishing a basis for subsequent refinement.

3.2. Stage Two: Feedback-Guided Retraining

While initial CoT distillation enables the student model to mimic structured reasoning, it often results in superficial pattern matching rather than genuine reasoning ability. To address this limitation, the second stage of CDFG introduces a feedback-guided training mechanism that explicitly analyzes the student's failure cases and uses them as an additional learning signal.

After completing the first-stage training, the student model is applied to the original dataset $\mathcal{D}ct$for inference. For each inputx_i, it generates a reasoning chain $\widehat{c_i} = S(x_i)$ and produces a final answer

$\widehat{y_i}$. If the predicted answer is incorrect, i.e., $\widehat{y_i} \neq y_i$, the sample is identified as a failure case and included in an error-specific dataset:

$$\mathcal{D}_{\text{err}} = \{(x_i, \widehat{c_i}, \widehat{y_i}) \mid \widehat{y_i} \neq y_i\} \quad (3)$$

For each failure case, both the original input x_i and the erroneous reasoning $\widehat{c_i}$ are passed to the teacher model, which is prompted to analyze the mistake and generate a corrective explanation or suggestion. This results in a feedback message $a_i$defined as:

$$a_i = T_{\text{feedback}}(x_i, \widehat{c_i}) \quad (4)$$

This feedback will often highlight specific error types, such as logical flaws, arithmetic errors, or missing assumptions, and include ideas for resolving them. To generalize the corrective signals and avoid overfitting to individual instances, we prompt the teacher model to summarize each feedback message into a more broadly applicable suggestion. Specifically, for all error cases, the teacher generates a distilled suggestion $\tilde{a}$ that encapsulates the key reasoning flaw identified in a_i:

$$\$\tilde{a} = T_{\{\text{summarize}} \left(\{a_i\}_{i \in |\mathcal{D}_{\{\text{err}}|}\right)\$ \quad (5)$$

This yields concise and instructional guidance tailored to the student's error, helping the model internalize general reasoning principles rather than memorizing specific fixes. The student model is then retrained exclusively on the error-specific dataset. For each (x_i, c_i) in $\mathcal{D}_{\text{err}}$ a new input is constructed by concatenating the original query with the generalized feedback:

$$\widetilde{x_i} = \text{Concat}(x_i, \tilde{a}) \quad (6)$$

and the student model is trained to regenerate the original teacher-produced reasoning c_i. The corresponding loss function is:

$$\mathcal{L}_2 = \frac{1}{|\mathcal{D}_{\text{err}}|} \sum_i \text{CE}(S(\widetilde{x_i}), c_i) \quad (7)$$

To form a unified training objective, the total loss combines the initial CoT distillation loss and the feedback-guided correction loss, weighted by a hyperparameter λ:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_1 + \lambda \cdot \mathcal{L}_2. \tag{8}$$

The training procedure is shown in Figure 1 and Figure 2.

Algorithm 1 CDFG

Require: Training dataset $\mathcal{D} = \{(x_i, y_i)\}$, teacher language model T, student language model S

Stage 1: CoT Distillation
for each (x_i, y_i) in $\mathcal{D}$ **do**
  Generate CoT reasoning $c_i \leftarrow T(x_i)$
end for
Construct CoT dataset $\mathcal{D}_{\text{cot}} = \{(x_i, c_i)\}$
Train student model S on $\mathcal{D}_{\text{cot}}$ using cross-entropy loss $\mathcal{L}_1$
Stage 2: Feedback-Guided Fine-Tuning
for each (x_i, y_i) in $\mathcal{D}$ **do**
  Predict $\hat{c}_i \leftarrow S(x_i)$ and extract $\hat{y}_i$ from $\hat{c}_i$
  if $\hat{y}_i \neq y_i$ **then**
    Generate feedback $a_i \leftarrow T_{\text{feedback}}(x_i, \hat{c}_i)$
    Summarize to obtain abstract suggestion $\tilde{a} \leftarrow T_{\text{summarize}}(a_i)$
    Create augmented input $\tilde{x}_i \leftarrow \text{Concat}(x_i, \tilde{a})$
    Store $(\tilde{x}_i, c_i)$ in $\mathcal{D}_{\text{err}}$
  end if
end for
Retrain S on $\mathcal{D}_{\text{err}}$ with loss $\mathcal{L}_{\text{total}}$
return Fine-tuned student language model.

Figure 1. CDFG Algorithm

3.3. Inference

During inference, the student model only needs to take the input question to generate the corresponding reasoning and final answer. The suggestion texts produced during the second stage are designed to assist the student model in understanding the reasoning process during training. Specifically, the student model learns the relationship between the suggestion and the reasoning chains in the second stage. As a result, the model can perform accurate and coherent inference solely based on the input question, without requiring additional context or suggestions at test time. This ensures that the inference cost remains unchanged while still benefiting from the improvements gained through suggestion-based training.

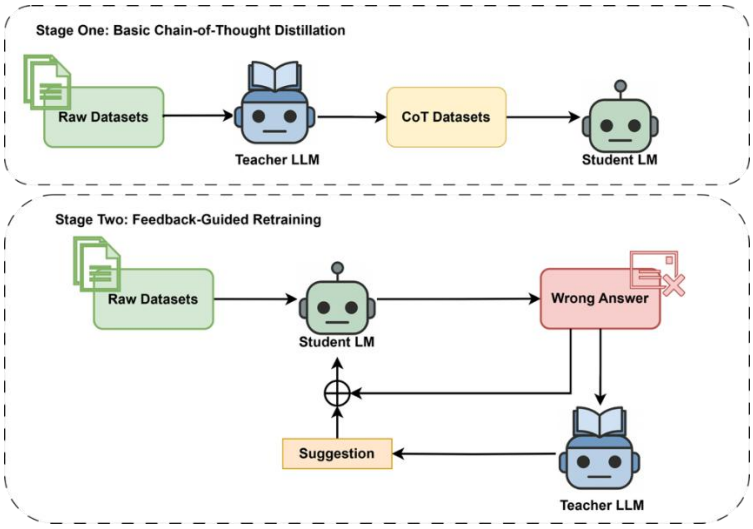

Figure 2. An illustration of CDFG

4. EXPERIMENTS

4.1. Dataset

ASDiv dataset [22] contains 2,305 English math word problems (MWPs) and is primarily used for evaluating and developing MWP solvers. It enables researchers and developers to test and improve models' ability to solve mathematical word problems. GSM8K [23] is a high-quality dataset of 8,500 grade school-level mathematical word problems curated to evaluate the arithmetic reasoning ability of language models. Each question requires a series of simple, interpretable steps to reach the answer, making it a standard benchmark for assessing chain-of-thought and numerical reasoning capabilities. SVAMP [24] consists of 1,000 elementary math word problems and serves as a benchmark dataset for evaluating the robustness of deep learning models to simple variations in mathematical word problems.

4.2. Baselines

In the Zero-shot [25], the student model generates answers directly from the input without any intermediate reasoning steps, reflecting its raw, unassisted reasoning ability. Zero-shot+CoT [14] enhances this by incorporating chain-of-thought (CoT) prompts into the input, encouraging more structured reasoning. CoT-KD (Chain-of-Thought Knowledge Distillation) [6] supervises the student model using full reasoning chains produced by a teacher model, enabling the student to mimic coherent multi-step thinking. SFT (Supervised Fine-Tuning) trains the model solely on final answers, without exposing it to intermediate reasoning, which may limit its ability to generalize. Lastly, Step-by-step [12] adopts a multi-task learning framework, training the model to generate both the reasoning process and the final answer, promoting a more holistic understanding of the task.

4.3. Implementations

For all experiments, we use Mistral-7B [26] and LLaMA2-7B [27] as student models. We employ GPT-4o as the teacher model, which is responsible for generating the initial Chain-of-Thought (CoT) data, as well as conducting subsequent error analysis and suggestion extraction. During the initial construction of the CoT dataset, we only select samples where the teacher model provides correct answers to train the student models. This precaution ensures that erroneous information is not propagated to the student models during the early stages of training, thereby maintaining the quality and reliability of the distilled reasoning knowledge. We adopt LoRA [28] to efficiently fine-tune the parameters of the student models. All experiments are conducted on four NVIDIA V100 GPUs, each equipped with 32GB of memory.

4.4. Main Results

As shown in Table 1, the proposed CDFG algorithm consistently achieves strong performance across all math reasoning benchmarks. Under the initial Zero-shot setting, both student models exhibit limited reasoning ability. Comparing Zero-shot with Zero-shot+CoT, we observe that CoT prompts provide valuable guidance, helping the models better understand and analyze the problems. When comparing the two student models, Mistral-7B consistently outperforms LLaMA2-7B, suggesting its stronger base reasoning capacity. After fine-tuning with different strategies, both models show substantial improvements in reasoning accuracy. Notably, even

standard CoT-KD brings significant gains, highlighting the effectiveness of distilling structured reasoning traces. Building upon this foundation, our CDFG method further enhances performance by introducing feedback-driven refinement, ultimately achieving the best results across all datasets. These findings demonstrate the effectiveness of our approach. Specifically, after the initial CoT distillation, we identify failure cases where the student model produces incorrect reasoning. For these cases, the teacher model generates targeted feedback based on the student's reasoning path. Leveraging the language model's natural ability to understand and incorporate textual suggestions, the student learns not only from the input problem but also from the teacher's corrections. This feedback-guided learning encourages deeper comprehension of reasoning patterns and further improves the model's inference capability.

Dataset	ASDiv	GSM8k	SVAMP	Avg
Mistral-7B				
Zero-shot	12.10	4.28	24.00	13.46
Zero-shot+CoT	24.20	15.63	29.33	23.06
CoT-KD	79.93	72.83	85.33	79.37
SFT	60.82	12.59	67.33	46.92
Step-by-step	60.50	13.65	70.66	48.72
CDFG	**83.12**	**73.14**	**88.00**	**81.42**
LLaMA2-7B				
Zero-shot	11.29	2.31	12.08	8.56
Zero-shot+CoT	13.29	4.59	11.56	9.82
CoT-KD	63.69	46.43	64.67	58.26
SFT	57.64	11.08	58.00	42.24
Step-by-step	46.26	30.43	44.78	40.49
CDFG	**65.92**	**47.19**	**69.33**	**60.81**

Table 1. Experimental results (%) of Mistral-7B [26] and LLaMA2-7B [27] on math reasoning benchmarks using different training strategies. The best result in each setting is highlighted in bold.

4.5. Ablation Study

To further evaluate the effectiveness of the CDFG framework, we conducted ablation studies focusing on the second-stage training process. In this variant, we removed the suggestion generation component and applied basic CoT distillation only to the error cases identified from the student model's predictions.

The results show that simply re-training on the student's incorrect samples without incorporating feedback not only fails to improve performance but may even cause a slight decline. This highlights the critical role of feedback generation in our framework. By offering explicit, error-specific guidance, the teacher model enables the student to better understand where its reasoning failed and how to correct it. This targeted supervision leads to more effective learning and ultimately enhances the student's reasoning capabilities.

Dataset	ASDiv	GSM8k	SVAMP	Avg
Mistral-7B				
CoT-KD	79.93	72.83	85.33	79.37
w/o suggestion $\tilde{a}$	82.17	68.13	85.33	78.54
CDFG	**83.12**	**73.14**	**88.00**	**81.42**
LLaMA2-7B				
CoT-KD	63.69	46.43	64.67	58.26
w/o suggestion $\tilde{a}$	63.38	43.70	67.33	58.14
CDFG	**65.92**	**47.19**	**69.33**	**60.81**

Table 2. Ablation Study of CDFG

5. Conclusions and Future Work

In this paper, we present CDFG, a two-stage CoT distillation framework for enhancing the reasoning ability of student models. The first stage performs standard CoT distillation using teacher-generated reasoning traces. In the second stage, the teacher analyzes failure cases from the student and provides targeted suggestions to guide further learning. By incorporating these suggestions, the student revisits its errors and develops a deeper understanding of the reasoning process. Experimental results demonstrate that CDFG significantly improves student performance across multiple benchmarks.

For future work, more advanced distillation strategies could be incorporated in the first stage to further enhance student model performance. Moreover, as our framework involves multiple queries to the teacher model, which may increase computational and financial overhead, future efforts will focus on optimizing the distillation process to improve overall efficiency.

Acknowledgements

This work was supported by Zhejiang Provincial Key Research and Development Project (2023C01043), Zhejiang Province Leading Geese Plan(2025C02025), Academy Of Social Governance Zhejiang University, and Inclusive and Smart Urban-Rural Governance Lab, Zhejiang University.

References

[1] A. Chowdhery, S. Narang, J. Devlin, M. Bosma, G. Mishra, A. Roberts, P. Barham, H. W. Chung, C. Sutton, S. Gehrmann et al., "Palm: Scaling language modeling with pathways," Journal of Machine Learning Research, vol. 24, no. 240, pp. 1–113, 2023.

[2] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray et al., "Training language models to follow instructions with human feedback," Advances in neural information processing systems, vol. 35, pp. 27730–27744, 2022.

[3] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell et al., "Language models are few-shot learners," Advances in neural information processing systems, vol. 33, pp. 1877–1901, 2020.

[4] H. Liao, S. He, Y. Hao, X. Li, Y. Zhang, J. Zhao, and K. Liu, "Skintern: Internalizing symbolic knowledge for distilling better cot capabilities into small language models," in Proceedings of the 31st International Conference on Computational Linguistics, 2025, pp. 3203–3221.

[5] W. X. Zhao, K. Zhou, J. Li, T. Tang, X. Wang, Y. Hou, Y. Min, B. Zhang, J. Zhang, Z. Dong et al., "A survey of large language models," arXiv preprint arXiv:2303.18223, 2023.

[6] L. C. Magister, J. Mallinson, J. Adamek, E. Malmi, and A. Severyn, "Teaching small language models to reason," arXiv preprint arXiv:2212.08410, 2022.
[7] C.Dai, K. Li, W. Zhou, and S. Hu, "Beyond imitation: Learning key reasoning steps from dual chain-of-thoughts in reasoning distillation," arXiv preprint arXiv:2405.19737, 2024.
[8] J. Wei, X. Wang, D. Schuurmans, M. Bosma, F. Xia, E. Chi, Q. V. Le, D. Zhou et al., "Chain-of-thought prompting elicits reasoning in large language models," Advances in neural information processing systems, vol. 35, pp. 24824–24837, 2022.
[9] S. Nayab, G. Rossolini, M. Simoni, A. Saracino, G. Buttazzo, N. Manes, and F. Giacomelli, "Concise thoughts: Impact of output length on llm reasoning and cost," arXiv preprint arXiv:2407.19825, 2024.
[10] Z. Yu, L. He, Z. Wu, X. Dai, and J. Chen, "Towards better chain-of-thought prompting strategies: A survey," arXiv preprint arXiv:2310.04959, 2023.
[11] X. Zhu, J. Li, C. Ma, and W. Wang, "Key-point-driven mathematical reasoning distillation of large language model," arXiv preprint arXiv:2407.10167, 2024.
[12] C.-Y. Hsieh, C.-L. Li, C.-K. Yeh, H. Nakhost, Y. Fujii, A. Ratner, R. Krishna, C.-Y. Lee, and T. Pfister, "Distilling step-by-step! outperforming larger language models with less training data and smaller model sizes," arXiv preprint arXiv:2305.02301, 2023.
[13] H. Chen, S. Wu, X. Quan, R. Wang, M. Yan, and J. Zhang, "Mcc-kd: Multi-cot consistent knowledge distillation," arXiv preprint arXiv:2310.14747, 2023.
[14] T. Kojima, S. S. Gu, M. Reid, Y. Matsuo, and Y. Iwasawa, "Large language models are zero-shot reasoners," Advances in neural information processing systems, vol. 35, pp. 22199–22213, 2022.
[15] X. Wang, J. Wei, D. Schuurmans, Q. Le, E. Chi, S. Narang, A. Chowdhery, and D. Zhou, "Self-consistency improves chain of thought reasoning in language models," arXiv preprint arXiv:2203.11171, 2022.
[16] S. Yao, D. Yu, J. Zhao, I. Shafran, T. Griffiths, Y. Cao, and K. Narasimhan, "Tree of thoughts: Deliberate problem solving with large language models," Advances in neural information processing systems, vol. 36, pp. 11809–11822, 2023.
[17] Z. Zhang, A. Zhang, M. Li, and A. Smola, "Automatic chain of thought prompting in large language models," arXiv preprint arXiv:2210.03493, 2022.
[18] P. Phan, H. Tran, and L. Phan, "Distillation contrastive decoding: Improving llms reasoning with contrastive decoding and distillation," arXiv preprint arXiv:2402.14874, 2024.
[19] S. Wadhwa, S. Amir, and B. C. Wallace, "Investigating mysteries of cot-augmented distillation," arXiv preprint arXiv:2406.14511, 2024.
[20] C. Li, Q. Chen, L. Li, C. Wang, Y. Li, Z. Chen, and Y. Zhang, "Mixed distillation helps smaller language model better reasoning," arXiv preprint arXiv:2312.10730, 2023.
[21] M. Kang, S. Lee, J. Baek, K. Kawaguchi, and S. J. Hwang, "Knowledge-augmented reasoning distillation for small language models in knowledge-intensive tasks," Advances in Neural Information Processing Systems, vol. 36, pp. 48573–48602, 2023.
[22] S.-y. Miao, C.-C. Liang, and K.-Y. Su, "A diverse corpus for evaluating and developing english math word problem solvers," in Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020, pp. 975–984.
[23] K. Cobbe, V. Kosaraju, M. Bavarian, M. Chen, H. Jun, L. Kaiser, M. Plappert, J. Tworek, J. Hilton, R. Nakano et al., "Training verifiers to solve math word problems," arXiv preprint arXiv:2110.14168, 2021.
[24] A. Patel, S. Bhattamishra, and N. Goyal, "Are nlp models really able to solve simple math word problems?" arXiv preprint arXiv:2103.07191, 2021.
[25] Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever et al., "Language models are unsupervised multitask learners," OpenAI blog, vol. 1, no. 8, p. 9, 2019.
[26] Q. Jiang, A. Sablayrolles, A. Mensch, C. Bamford, D. S. Chaplot, D. de las Casas, F. Bressand, G. Lengyel, G. Lample, L. Saulnier, L. R. Lavaud, M.-A. Lachaux, P. Stock, T. L. Scao, T. Lavril, T. Wang, T. Lacroix, and W. E. Sayed, "Mistral 7b," 2023. [Online]. Available: https://arxiv.org/abs/2310.06825
[27] H. Touvron, L. Martin, K. Stone, P. Albert, A. Almahairi, Y. Babaei, N. Bashlykov, S. Batra, P. Bhargava, S. Bhosale et al., "Llama 2: Open foundation and fine-tuned chat models," arXiv preprint arXiv:2307.09288, 2023.
[28] E.J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, W. Chen et al., "Lora: Low-rank adaptation of large language models." ICLR, vol. 1, no. 2, p. 3, 2022.

AUTHORS

Lingzhi Gao received his Master of Science in Software Engineering from Zhejiang University. His research mainly focuses on large language models.

Xuan Wang is a Senior Staff Member at China Media Group.

Tianrun Cai is a PhD student at Alliance Manchester Business School, The University of Manchester. His research focuses on federated learning, large language models, and computational social science.

Xiunao Lin is a Senior Software Engineer at Zhejiang Post & Telecommunication Construction Co., Ltd.

Chao Wu is an Associate Professor at the School of Public Affairs, Zhejiang University. His research mainly focuses on large language models and computational social science.