ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 979-8-3503-6874-1/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICASSP49660.2025.10890771

HCoTT: Hierarchical Chain-of-Thought Distillation

Zhichang Wang*
SECE, Peking University
Beijing, China
wzcc@stu.pku.edu.cn

Xianwei Zhuang*
SECE, Peking University
Beijing, China
xwzhuang@stu.pku.edu.cn

Zhihong Zhu
SECE, Peking University
Beijing, China
zhihongzhu@stu.pku.edu.cn

Yuexian Zou†
SECE, Peking University
Beijing, China
zouyx@pku.edu.cn

***Abstract*—Chains of Thought (CoT) have shown potential in augmenting the reasoning capabilities of language models, yet their effectiveness is predominantly observed in large language models (LLMs). Recently, several attempts have been made to inject CoT into small language models (SLMs) using distillation and achieved promising results. However, current methods (1) ignore the rationality and hierarchical logic of reasoning when constructing CoT; (2) fail to inject hierarchical reasoning priors into SLMs. In this paper, we design a Hierarchical CoT distillation framework termed HCoTT, whose core component is a hierarchical recursive sampling module and a hierarchical learning module. Specifically, hierarchical recursive sampling utilizes a hierarchical logic process to generate more diverse explanations and a Hierarchical Chain of Thought (HCoT). Furthermore, hierarchical learning encompasses hierarchical supervision and representation learning, which is designed to augment learning and representation of implicit explanatory priors in HCoT for SLMs. Experimental results show that HCoTT can effectively improve the performance of SLMs on Faculty-Reasoning and Multiple-Choice QA tasks. More impressively, our method is model-independent and can consistently improve performance with existing language model fusions of different scales.**

***Index Terms*—Chain of Thought, Large Language Model, Question-Answering Task**

I. Introduction

Large Language Models (LLMs) exhibit robust capabilities across various downstream tasks such as language generation and question-answer reasoning. Prior studies [1], [2] have demonstrated that the generation of a Chain of Thought (CoT) [3] can markedly enhance the reasoning capabilities of LLMs. Nevertheless, the application of the CoT methodology [4], [5] to enhance reasoning in Small Language Models (SLMs) presents significant challenges. Some works focus on transferring reasoning abilities from large to small language models. West et al. [6] trained students for knowledge completion. Chan et al. [7] used principles to augment teacher models. Shridhar et al. [8] trained students to decompose questions. Li et al. [9] proposed joint answer-generation training. SCOTT [10] applied contrastive learning to enhance explanation consistency and counterfactual reasoning. Although their work has shown significant progress, we have identified two key issues: (1) The constructed CoT exhibits linearity and independence, lacking a hierarchical logical structure. (2) The employed learning objectives fail to equip SLMs with the capacity to grasp hierarchical reasoning semantics within CoT.

The idea of hierarchy learning is widely used in computer vision, such as text categorization [11], functional genomics [12], and image classification [13]. DHSS [14] applies hierarchical learning to semantic segmentation. In NLP, hierarchical concepts are used in text classification [15], [16] and QA tasks [17], [18], though these focus on fine-tuning SLMs rather than integrating CoT for small models. A hierarchical vectorization algorithm arranges data samples in a high-dimensional space, where proximity reflects semantic similarity [19]. Some methods parameterize the vector space hierarchically using models [20], [21], but this is computationally intensive. Others use hierarchy-aware metric learning [22], [23] to shape the vector space directly.

We propose **H**ierarchical **C**hain-**o**f-**T**hought Dis**T**illation (**HCoTT**), a novel distillation method that uses a recursive approach to generate hierarchical CoT (HCoT) with LLMs and trains on SLMs with a hierarchical loss. We conducted extensive experiments across multiple datasets focused on factual reasoning and multi-item question answering to substantiate the effectiveness and universality of HCoTT. Our contributions can be summarized as follows: (1) Introduction of a hierarchical recursive sampling method for constructing HCoT, facilitating the acquisition of randomly distributed hierarchical CoT information. (2) Proposal of the hierarchical consistent learning module, enabling the incorporation of semantic priors with hierarchical constraints into small models. This facilitates enhanced learning of implicit logical information from CoT pathways by SLMs. (3) Introduction of hierarchical thought contrastive learning, aimed at optimizing the representation of the chain of thought within the semantic space of small models. This allows SLMs to effectively represent CoT nodes across both similar and dissimilar pathways.

II. Conventional Approach

HCoTT is generally divided into two parts, namely the teacher network and the student network, where the teacher model is an LLM. For all tasks, input structured data can be formalized as $\{< q, a_t >| \ a_t \in A_c\}$, where, q represents a question statement, $A_c = \{a_1, .., a_c\}$ denotes a set of candidate answers and $a_t^* \in A_c$ is the correct answer among a series of candidate answers.

Teacher Network. We utilize LLMs as teacher networks to generate a series of explanations $v_t^i \in \mathbf{v}_t$ for each question and answer pair $< q, a_t >$. We further define all prefixes that generate the specific explanations v_t^i as state paths $\mathrm{P}(v_t^i) =$

[*] Equal contribution
[†] Corresponding author

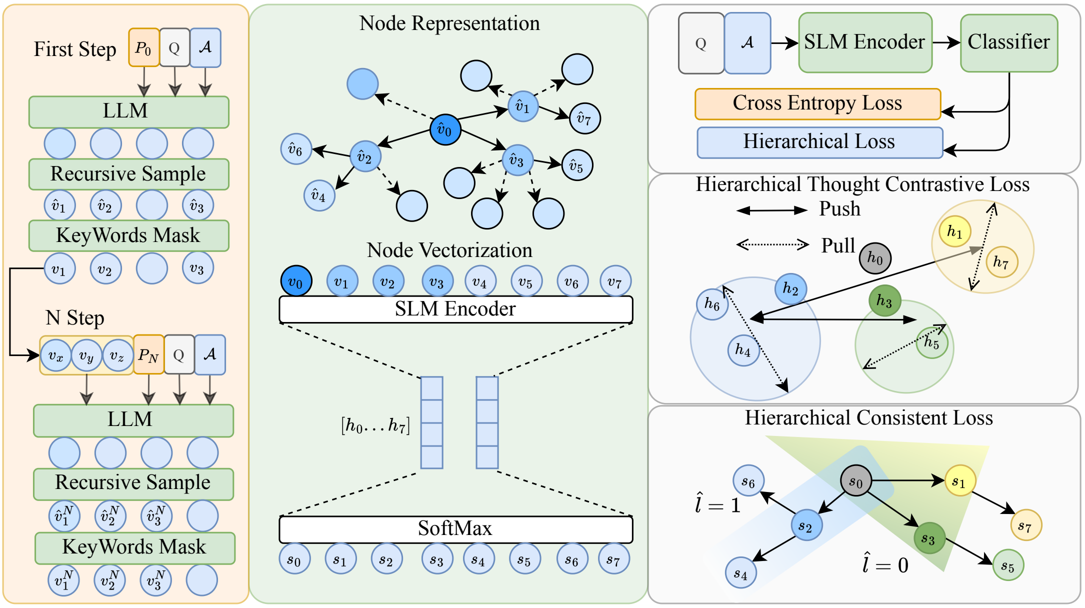

Fig. 1. The illustration depicts the HCoTT framework, comprising: (1) HCoT, which encompasses explanations structured with hierarchical logic achieved via N-step recursive construction; and (2) the effective fine-tuning of hierarchical logic tailored for SLMs facilitated by hierarchical consistent learning and hierarchical thought contrastive learning.

$[v_t^0 \to v_t^i]$. Therefore, each optimal explanatory node v_t^i in the teacher network will be generated in the following unified way:

$$v_t^{i^*} = \arg\max \log P\left(v_i \mid p, q, a_t^*, \mathrm{P}(v_t^i)\right), \qquad (1)$$

where p denotes an input prompt.

Student Network. The student network needs to enable the classifier [11] to fit the correct answers. SLMs will utilize explanatory sets $\mathbf{v}_t$ and questions q as inputs to predict answers y_t^*. Therefore, the category cross-entropy loss is optimized as:

$$\mathcal{L}_{cce}(y) = \mathrm{CE}(a_t, y_i), \qquad (2)$$

where, $\mathrm{CE}(\cdot)$ denotes the cross entropy function.

A. Construction of HCoT

Root of HCoT. For each question and correct answer pair $< q, a_t^* >$, we provide a constant prompt p_0. We will input $< q, a_t^*, p_0 >$ as an instruction into LLMs to obtain n output explanations, and further obtain m explanations $\hat{V}^1 = [\hat{v}_1^1, \hat{v}_2^1, \cdots, \hat{v}_m^1]$ using hierarchical sampling. The leftmost legend of Figure 1 shows the hierarchical sampling process.

Recursively Obtain the Next State. We further iterate through the explanations $\hat{v}_x^1 \in \hat{V}^1$, and concatenate $< q, a_t^*, p_0, \hat{v}_x^1 >$ as the instruction input for LLMs, and sample in the same way to obtain the next state $\hat{V}_2^2$. Through repeated iterations, we will obtain a complete hierarchical reasoning explanation $\hat{V}^1, \cdots, \hat{V}^m$. We further utilize the mask function to obtain the explanation $V^1, \cdots, V^m$ desensitized to the answer keywords:

$$V_j = \hat{V}_j \odot \prod_{k=1}^{c} (1 - \mathrm{MASK}(a_k)) \quad \text{for } j = 1, 2, \ldots, m \qquad (3)$$

Formulated HCoT. We formalize the thinking process of HCoT as $\mathcal{T} = (\mathcal{V}, \mathcal{E})$, where, each state $v \in \mathcal{V}$ represents an explanation of thought on HCoT, and $\mathcal{E}$ represents the parent dependency edge of each explanation. We define the question q as the root state. Except for the root node, each node is sampled through LLMs, and the distribution of each node satisfies: $v_i \sim p_{\mathrm{LLM}}(v_i | a, \mathrm{P}(v_i))$, where, $\mathrm{P}(v_i)$ is the set of all parent explanation states of v_i, and a represents the answer related to question q.

Each edge $(u, v) \in \mathcal{E}$ represents a relationship between two different levels of explanations. The parent state u is the inference process of the previous level, i.e., v is obtained from u through LLMs. The leaf nodes $\mathcal{V}_\chi$ are the final layer of reasoning, focusing more on knowledge at the level of details. For each explanation, we construct a discriminant function $\hat{\mathbf{l}} = [\hat{l}_0, \cdots, \hat{l}_{|\mathcal{V}|}]$ to determine whether it belongs to a node in $\mathcal{V}$, where $\hat{l}_i \in \hat{\mathbf{l}}$ represents a boolean value.

B. Hierarchical Consistent Learning

Refer to previous work [24], we input each explanation v_i in HCoT into the encoder of SLMs to obtain $\boldsymbol{h}_i \in \mathbb{R}^{dim}$: $h_i = \mathrm{Encoder}(v_i)$, where dim denotes the feature dimension. We further utilize the softmax function to obtain the score vector $\mathbf{s} \in \mathbb{R}^n$ for $\boldsymbol{h}_i$:

$$\mathbf{s} = \mathrm{softmax}(W^\top \boldsymbol{h}_i), \qquad (4)$$

where n denotes the number of nodes in HCoT, and $W \in \mathbb{R}^{dim \times n}$ is a matrix. $\mathbf{s} = [s_v]_{v \in \mathcal{V}}$ indicates the score that belongs to each node of HCoT in the hierarchical distribution.

For each explanation, the reasoning path it represents in HCoT $\mathcal{T}$ is formalized as:

$$\left\{v_1^*, \cdots, v_{|\mathcal{P}|}^*\right\} = \operatorname*{arg\,max}_{\mathcal{P} \subseteq \mathcal{T}} \sum_{v_p \in \mathcal{P}} s_{v_p}, \qquad (5)$$

TABLE I
A PERFORMANCE COMPARISON OF VARIOUS METHODS ON THE FACTUAL REASONING BENCHMARK

Base Model	Method	CREAK			CSQA2			StrategyQA		
		LAS	LAS_{f1}	LAS_{ins}	LAS	LAS_{f1}	LAS_{ins}	LAS	LAS_{f1}	LAS_{ins}
BERT-base	CoT	26.34	26.36	26.63	33.46	33.57	32.95	10.95	10.95	10.14
	SCOTT	26.44	26.46	26.62	33.43	33.54	33.33	10.35	10.35	10.09
	HCoTT	27.41	27.43	27.52	38.46	38.57	37.91	10.49	10.49	9.96
	HCoTT+L_h	**27.6**	**27.62**	**27.97**	**39.4**	**39.51**	**39.07**	**11.31**	**11.31**	**10.87**
RoBERTa-base	CoT	28.07	28.07	28.19	37.52	37.53	36.36	14.25	14.26	13.72
	SCOTT	28.84	28.84	28.81	38.71	38.72	37.51	14.33	14.34	13.98
	HCoTT	28.59	28.59	28.56	40.62	40.63	39.68	14.25	14.26	13.72
	HCoTT+L_h	**29.29**	**29.29**	**29.3**	**42.52**	**42.53**	**41.36**	**14.7**	**14.71**	**14.26**

where $\mathcal{P} = \left\{v_1, \cdots, v_{|\mathcal{P}|}\right\} \subseteq \mathcal{T}$ represents the reasoning path from the question to the leaf node, *i.e.*, $v_{|\mathcal{P}|} \in \mathcal{V}_\chi$.

Hierarchical Consistent Loss. To ensure the satisfaction of the two hierarchy constraints, we estimate a hierarchy-coherent score map $\mathbf{m}$ from $\mathbf{s}$. For explanation, the updated score vector $\mathbf{m} = [m_v]_{v\in\mathcal{V}}$ is formalized as:

$$\begin{cases} m_v = \min\limits_{u\in\mathcal{A}_v}(s_u) & \text{if } \hat{l}_v = 1 \\ 1 - m_v = 1 - \max\limits_{u\in\mathcal{C}_v}(s_u) & \text{if } \hat{l}_v = 0 \end{cases} \tag{6}$$

where $\mathcal{A}_v$ and $\mathcal{C}_v$ denote the superclass and subclass sets of explainations v in HCoT respectively, and $\mathbf{s} = [s_v]_{v\in\mathcal{V}}$ refers to the original score vector of explanation.

We thus build a Hierarchical Consistent Loss as:

$$\begin{aligned} \mathcal{L}_{hc}(\boldsymbol{m}) &= \sum_{v\in\mathcal{V}} -\hat{l}_v \log(m_v) - \left(1 - \hat{l}_v\right)\log(1 - m_v) \\ &= \sum_{v\in\mathcal{V}} -\hat{l}_v \log\left(\min_{u\in\mathcal{A}_v}(s_u)\right) - \\ &\quad \left(1 - \hat{l}_v\right)\log\left(1 - \max_{u\in\mathcal{C}_v}(s_u)\right) \end{aligned} \tag{7}$$

$\mathcal{L}_{hc}$ can better satisfy hierarchical constraints, allowing SLMs to learn the hierarchical logic provided by LLMs and gain diverse label understanding.

C. Hierarchical Thought Contrastive Learning

Positive and Negative Samples. We define the function $\phi(u, v)$ as the shortest path between two explanations u and v on HCoT, similar to the shortest path between two nodes in a tree structure. On our HCoT, $\phi(u, v)$ of u and v reflect their relative distance in the hierarchical representation space, which will serve as a semantic similarity metric. Our representation loss is optimized on a set of explanation triplets $\left\{v_i, v_i^+, v_i^-\right\}$, where v_i, v_i^+, v_i^- are anchor, positive and negative explanation samples, respectively. $\left\{v_i, v_i^+, v_i^-\right\}$ are sampled from the whole training batch, such that $\phi\left(v_i, v_i^+\right) < \phi\left(v_i, v_i^-\right)$ on HCoT. As such, in our representation loss, the positive samples are more semantically similar to the anchor explanations in HCoT, compared with the negative explanations.

Hierarchical Thought Contrastive Loss. With a valid explanation triplet $\left\{v_i, v_i^+, v_i^-\right\}$, our Hierarchical Thought Contrastive Loss is given as:

$$\begin{cases} \mathcal{L}_{ht}\left(\boldsymbol{d}_i\right) = \max\left\{\boldsymbol{d}_i + mg, 0\right\} \\ \boldsymbol{d}_i = \left\langle \boldsymbol{h}_i, \boldsymbol{h}_i^+ \right\rangle - \left\langle \boldsymbol{h}_i, \boldsymbol{h}_i^- \right\rangle \end{cases} \tag{8}$$

where $\boldsymbol{h}_i, \boldsymbol{h}_i^+, \boldsymbol{h}_i^- \in \mathbb{R}^{dim}$ are the embeddings of v_i, v_i^+, and v_i^-, respectively, obtained from the encoder of SLMs, and $\langle\cdot,\cdot\rangle$ is a cosine function to measure the similarity of two inputs. The margin mg forces the gap of $\left\langle \boldsymbol{h}_i, \boldsymbol{h}_i^- \right\rangle$ and $\left\langle \boldsymbol{h}_i, \boldsymbol{h}_i^+ \right\rangle$ larger than mg. The margin mg is determined as:

$$\begin{aligned} mg &= mg_\varepsilon + 0.5 mg_\tau \\ mg_\tau &= \left(\phi\left(v_i, v_i^-\right) - \phi\left(v_i, v_i^+\right)\right)/2D \end{aligned} \tag{9}$$

where m_ε is set as a constant for the tolerance of the intra-class variance, i.e., maximum intra-class distance, $mg_\tau \in [0, 1]$ is a dynamic violate margin, which is computed according to the semantic relationships among v_i, v_i^+, and v_i^- over HCoT, and D refers to the height of HCoT.

D. Overall Objective

Our complete learning objective consists of three parts: category cross-entropy loss $\mathcal{L}_{cce}$ in Eq. 2 for prediction, hierarchical consistency loss $\mathcal{L}_{hc}$ in Eq. 7 for hierarchical supervised learning, and hierarchical thought contrastive loss $\mathcal{L}_{ht}$ in Eq. 8 for hierarchical representation learning. We define the loss for hierarchical constraints as: $\mathcal{L}_h = \mathcal{L}_{hc} + \alpha\mathcal{L}_{ht}$, where α is a trade-off hyperparameter. Therefore, our complete training loss can be expressed as: $\mathcal{L}_{total} = \mathcal{L}_{cce} + \lambda\mathcal{L}_h$, where λ is a hyperparameter that compromises prediction and hierarchical constraints. The hyperparameter α and λ are set to 0.5 to achieve the optimal performance in experiments.

III. EXPERIMENTS

A. Experimental Setup

We chose three Faculty Reasoning datasets, namely CREAK [25], StrategyQA [26], and CSQA2 [27], and three Multiple-Choice datasets, namely CSQA [28], QASC [29], and OBQA [30], as the main datasets for the experiments.To measure how well the student network performs with and without explanations on the same QA pair, we use LAS [31] to evaluate HCoTT gains.

We use GPT-3.5-turbo-1106, a large language model developed by OpenAI, as the teacher network for the construction of CoT and HCoT. We use RoBERTa [32], BERT [33] as

TABLE II
A PERFORMANCE COMPARISON OF VARIOUS METHODS ON THE MULTIPLE-CHOICE QA BENCHMARK

Base Model	Method	CSQA			OBQA			QASC		
		LAS	LAS_{f1}	LAS_{ins}	LAS	LAS_{f1}	LAS_{ins}	LAS	LAS_{f1}	LAS_{ins}
BERT-base	CoT	7.89	14.67	27.22	8.15	13.23	21.20	6.61	26.20	46.32
	SCOTT	16.05	27.10	40.45	11.85	19.63	31.80	6.24	28.21	55.10
	HCoTT	15.94	26.90	41.30	15.45	25.95	41.40	8.74	33.56	62.76
	HCoTT+L_h	**16.21**	**27.35**	**41.47**	**15.70**	**26.08**	**41.20**	**11.49**	**40.20**	**74.65**
RoBERTa-base	CoT	8.84	14.23	22.80	10.30	14.10	20.80	4.94	17.26	26.53
	SCOTT	14.97	23.46	31.86	15.95	23.98	33.80	0.35	0.15	-1.40
	HCoTT	16.39	25.67	33.88	19.15	28.10	38.00	9.78	33.15	56.31
	HCoTT+L_h	**16.54**	**25.90**	**33.80**	**19.25**	**28.29**	**38.20**	**11.82**	**37.82**	**61.53**

base models for our student network. For these models, we tested the adoption of two hierarchical loss functions and the original cross-entropy loss in our experiments. The constant m_ε for the tolerance of the intra-class variance is set as 0.1.The hidden size for text is set to 768. We employ Adam as the optimizer with a weight decay of 0.01 and tune all models for 6 epochs. We set the learning rate of 3e-6 on all the datasets. All experiments are conducted on 8 RTX 3090 GPUs. We will publish the complete code after the paper is accepted.

B. Main Results

1) Evaluation on Faculty-Reasoning Tasks

We present experimental results on three Faculty-Reasoning datasets in Table I, using LAS, LAS_{f1}, and LAS_{ins} metrics. Key insights from Table I are: (1) All models show significant metric improvements with CoT explanations compared to using raw data, as explanations add prior knowledge that enhances model fitting. (2) HCoTT+L_h outperforms other methods in all metrics. HCoTT+L_h shows better performance than HCoTT, which is weaker than CoT and SCOTT in some indicators, indicating limitations in learning explanatory semantics with hierarchical recursive sampling alone. (3) `RoBERTa-base` shows more significant gains over `BERT-base`, with notable differences across datasets, highlighting that model performance depends on task characteristics and base model capabilities. Stronger models better leverage provided explanations to enhance reasoning ability.

2) Evaluation on Multiple-Choice QA Tasks

We present experimental results on three multiple-choice QA datasets in Table II, using LAS, LAS_{f1}, and LAS_{ins} metrics. Key insights from Table II are: (1) For complex tasks, the data gain varies significantly across methods. Not all methods effectively enhance model performance; for instance, SCOTT showed a 1.40% drop in QASC's LAS_{ins} due to ineffective prior knowledge or interference. (2) HCoTT+L_h outperforms other methods across all indicators, especially for more complex tasks. For example, in the same model condition (`RoBERTa-base`), HCoTT+L_h achieved better LAS_{ins} on CSQA, OBQA, and QASC compared to the Faculty-Reasoning datasets in Table II. (3) HCoTT's metrics on Multiple-Choice QA datasets surpass those of SCOTT and CoT, demonstrating that enhancing explained semantic information effectively improves reasoning for complex tasks.

TABLE III
A PERFORMANCE COMPARISON OF VARIOUS HIERARCHICAL LOSS FUNCTIONS ON THE MULTIPLE BENCHMARK.

Datasets	Methods	LAS	LAS_{f1}	LAS_{ins}
CREAK	HCoTT	28.59	28.59	28.56
	HCoTT+$\mathcal{L}_{hc}$	29.18	29.18	29.15
	HCoTT+$\mathcal{L}_{ht}$	28.96	28.96	28.93
	HCoTT+$\mathcal{L}_h$	29.29	29.29	29.30
CSQA2	HCoTT	40.62	40.63	39.68
	HCoTT+$\mathcal{L}_{hc}$	40.92	40.93	39.76
	HCoTT+$\mathcal{L}_{ht}$	40.94	40.75	38.88
	HCoTT+$\mathcal{L}_h$	42.52	42.53	41.36
QASC	HCoTT	9.78	33.15	56.31
	HCoTT+$\mathcal{L}_{hc}$	11.74	37.63	61.42
	HCoTT+$\mathcal{L}_{ht}$	11.69	37.52	61.09
	HCoTT+$\mathcal{L}_h$	11.82	37.82	61.53

C. Ablation Study

1) The Effect of the Hierarchical Loss Enhancement

We present experimental results on three dataset types in Table III. Using `RoBERTa-base` generally yields higher scores than `BERT-base`. Compared to HCoTT, $+\mathcal{L}hc$ and $+\mathcal{L}ht$ improve all metrics, highlighting the effectiveness of Hierarchical Consistent Loss and Hierarchical Thought Contrastive Loss. However, $+\mathcal{L}hc$ and $+\mathcal{L}ht$ perform lower than $+\mathcal{L}h$, suggesting that combining loss functions synergistically enhances the student network's ability. The ablation results confirm significant performance gains from $+\mathcal{L}hc$ and $+\mathcal{L}_{ht}$.

IV. CONCLUSION

In this paper, we propose HCoTT, a hierarchical CoT distillation framework with two main components: a hierarchical recursive sampling module for CoT construction and a hierarchical learning module for mapping explanations to a semantic space. Experimental results demonstrate that HCoTT significantly boosts small model performance on Faculty-Reasoning and Multiple-Choice QA tasks.

References

[1] M. Nye, A. J. Andreassen, G. Gur-Ari, H. Michalewski, J. Austin, D. Bieber, D. Dohan, A. Lewkowycz, M. Bosma, D. Luan *et al.*, "Show your work: Scratchpads for intermediate computation with language models," *arXiv preprint arXiv:2112.00114*, 2021.

[2] J. Wei, X. Wang, D. Schuurmans, M. Bosma, E. Chi, Q. Le, and D. Zhou, "Chain of thought prompting elicits reasoning in large language models," *arXiv preprint arXiv:2201.11903*, 2022.

[3] J. Wei, X. Wang, D. Schuurmans, M. Bosma, B. Ichter, F. Xia, E. Chi, Q. Le, and D. Zhou, "Chain-of-thought prompting elicits reasoning in large language models," 2023.

[4] X. Ye and G. Durrett, "The unreliability of explanations in few-shot in-context learning," *arXiv preprint arXiv:2205.03401*, 2022.

[5] P. Wang, A. Chan, F. Ilievski, M. Chen, and X. Ren, "Pinto: Faithful language reasoning using prompt-generated rationales," *arXiv preprint arXiv:2211.01562*, 2022.

[6] P. West, C. Bhagavatula, J. Hessel, J. D. Hwang, L. Jiang, R. L. Bras, X. Lu, S. Welleck, and Y. Choi, "Symbolic knowledge distillation: from general language models to commonsense models," *arXiv preprint arXiv:2110.07178*, 2021.

[7] A. Chan, Z. Zeng, W. Lake, B. Joshi, H. Chen, and X. Ren, "Knife: Knowledge distillation with free-text rationales," *arXiv preprint arXiv:2212.09721*, 2022.

[8] K. Shridhar, A. Stolfo, and M. Sachan, "Distilling multi-step reasoning capabilities of large language models into smaller models via semantic decompositions," *arXiv preprint arXiv:2212.00193*, 2022.

[9] S. Li, J. Chen, Y. Shen, Z. Chen, X. Zhang, Z. Li, H. Wang, J. Qian, B. Peng, Y. Mao *et al.*, "Explanations from large language models make small reasoners better," *arXiv preprint arXiv:2210.06726*, 2022.

[10] P. Wang, Z. Wang, Z. Li, Y. Gao, B. Yin, and X. Ren, "SCOTT: Self-consistent chain-of-thought distillation," in *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Toronto, Canada: Association for Computational Linguistics, Jul. 2023, pp. 5546–5558. [Online]. Available: https://aclanthology.org/2023.acl-long.304

[11] J. Rousu, C. Saunders, S. Szedmak, and J. Shawe-Taylor, "Kernel-based learning of hierarchical multilabel classification models," *Journal of Machine Learning Research*, vol. 7, pp. 1601–1626, 2006.

[12] Z. Barutcuoglu, R. E. Schapire, and O. G. Troyanskaya, "Hierarchical multi-label prediction of gene function," *Bioinformatics*, vol. 22, no. 7, pp. 830–836, 2006.

[13] S. Bengio, J. Weston, and D. Grangier, "Label embedding trees for large multi-class tasks," *Advances in neural information processing systems*, vol. 23, 2010.

[14] L. Li, T. Zhou, W. Wang, J. Li, and Y. Yang, "Deep hierarchical semantic segmentation," 2022.

[15] R. Balyan, K. S. McCarthy, and D. S. McNamara, "Applying natural language processing and hierarchical machine learning approaches to text difficulty classification," *International Journal of Artificial Intelligence in Education*, vol. 30, pp. 337–370, 2020.

[16] R. A. Stein, P. A. Jaques, and J. F. Valiati, "An analysis of hierarchical text classification using word embeddings," *Information Sciences*, vol. 471, pp. 216–232, 2019.

[17] L. Pang, Y. Lan, J. Guo, J. Xu, L. Su, and X. Cheng, "Has-qa: Hierarchical answer spans model for open-domain question answering," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 6875–6882.

[18] Y. Fang, S. Sun, Z. Gan, R. Pillai, S. Wang, and J. Liu, "Hierarchical graph network for multi-hop question answering," *arXiv preprint arXiv:1911.03631*, 2019.

[19] M. Nickel and D. Kiela, "Poincaré embeddings for learning hierarchical representations," 2017.

[20] T. Chen, W. Wu, Y. Gao, L. Dong, X. Luo, and L. Lin, "Fine-grained representation learning and recognition by exploiting hierarchical semantic embedding," in *ACMMM*, 2018.

[21] S. F. Mousavi, M. Safayani, A. Mirzaei, and H. Bahonar, "Hierarchical graph embedding in vector space by graph pyramid," *PR*, 2017.

[22] T. Kerola, J. Li, A. Kanehira, Y. Kudo, A. Vallet, and A. Gaidon, "Hierarchical lovász embeddings for proposal-free panoptic segmentation," in *CVPR*, 2021.

[23] S. Yang, W. Yu, Y. Zheng, H. Yao, and T. Mei, "Adaptive semantic-visual tree for hierarchical embeddings," in *ACMMM*, 2019.

[24] E. Amigó and A. Delgado, "Evaluating extreme hierarchical multi-label classification," in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2022, pp. 5809–5819.

[25] Y. Onoe, M. J. Zhang, E. Choi, and G. Durrett, "Creak: A dataset for commonsense reasoning over entity knowledge," *OpenReview*, 2021.

[26] M. Geva, D. Khashabi, E. Segal, T. Khot, D. Roth, and J. Berant, "Did aristotle use a laptop? a question answering benchmark with implicit reasoning strategies," *Transactions of the Association for Computational Linguistics*, vol. 9, pp. 346–361, 2021.

[27] A. Talmor, O. Yoran, R. L. Bras, C. Bhagavatula, Y. Goldberg, Y. Choi, and J. Berant, "CommonsenseQA 2.0: Exposing the limits of AI through gamification," in *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 1)*, 2021. [Online]. Available: https://openreview.net/forum?id=qF7FlUT5dxa

[28] A. Talmor, J. Herzig, N. Lourie, and J. Berant, "Commonsenseqa: A question answering challenge targeting commonsense knowledge," *arXiv preprint arXiv:1811.00937*, 2018.

[29] T. Khot, P. Clark, M. Guerquin, P. Jansen, and A. Sabharwal, "QASC: A dataset for question answering via sentence composition," in *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*. AAAI Press, 2020, pp. 8082–8090.

[30] T. Mihaylov, P. Clark, T. Khot, and A. Sabharwal, "Can a suit of armor conduct electricity? a new dataset for open book question answering," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Brussels, Belgium: Association for Computational Linguistics, Oct.-Nov. 2018, pp. 2381–2391. [Online]. Available: https://aclanthology.org/D18-1260

[31] P. Hase, S. Zhang, H. Xie, and M. Bansal, "Leakage-adjusted simulatability: Can models generate non-trivial explanations of their behavior in natural language?" *arXiv preprint arXiv:2010.04119*, 2020.

[32] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "RoBERTa: A robustly optimized bert pretraining approach," *ArXiv*, vol. abs/1907.11692, 2019.

[33] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *NAACL*, 2019.