"The Whole Is Greater Than the Sum of Its Parts": A Compatibility-Aware Multi-Teacher CoT Distillation Framework

Jin Cui[*1], **Jiaqi Guo**[*2], **Ruixuan Yang**[1], **Jiayi Lu**[1], **Jiepeng Zhou**[3], **Jiajun Xu**[4], **Jiangcheng Song**[1], **Boran Zhao**[†4] and **Pengju Ren**[1]

[1]State Key Laboratory of Human-Machine Hybrid Augmented Intelligence, Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University
[2]School of Mathematical Sciences, Nankai University,
[3]The Hong Kong University of Science and Technology(Guangzhou)
[4]School of Software Engineering, State Key Laboratory of Human-Machine Hybrid Augmented Intelligence, Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University
andycui@stu.xjtu.edu.cn, {boranzhao, pengjuren}@xjtu.edu.cn

Abstract

Chain-of-Thought (CoT) reasoning empowers Large Language Models (LLMs) with remarkable capabilities, but typically requires enormous parameter scales. CoT distillation has emerged as a promising paradigm to transfer reasoning prowess into compact student models (SLMs), yet existing approaches often rely on a solitary teacher, capping the student's potential since individual LLMs often exhibit distinct capability biases and may suffer from catastrophic forgetting. While leveraging diverse teachers seems appealing, effectively fusing their supervision remains challenging: teacher-student incompatibility risks amplifying hallucinations, and passive supervision fails to ensure genuine logic internalization. To address this, we introduce *COMPACT*, a framework that adaptively fuses supervisions from different teachers by dynamically weighting teacher gradients based on the student's real-time compatibility evaluated by a multi-dimensional metric: (1) *Graph-based Consensus* to filter misleading rationales by identifying mainstream reasoning paths; (2) *Mutual-Information-based Adaptability* to detect "epiphany moments" for genuinely understanding the reasoning process rather than merely imitating; and (3) *Loss-based Difficulty* to assess student receptivity to the teacher's guidance and prevent negative transfer. Our experiments and latent space analysis demonstrate that *COMPACT* effectively integrates diverse reasoning capabilities without damaging the model's original knowledge structure, achieving state-of-the-art performance on various benchmarks while effectively mitigating catastrophic forgetting. Code available: https://github.com/CAG-Research/COMPACT.git.

1 Introduction

Large Language Models (LLMs) have demonstrated remarkable capabilities in complex reasoning tasks, particularly when empowered by Chain-of-Thought (CoT) reasoning [Wei *et al.*, 2022; Kojima *et al.*, 2022]. However, this emergent ability typically requires enormous parameter scales, hindering its practical deployment in resource-constrained environments. Consequently, CoT distillation offers a promising solution by transferring the reasoning strategies of LLMs into compact Student Models (SLMs) using teacher rationales as supervision [Ho *et al.*, 2022; Magister *et al.*, 2023; Hsieh *et al.*, 2023].

However, existing distillation approaches typically constrain SLMs to fit a single teacher-generated rationale. Ignoring reasoning path diversity that often leads SLMs to merely mimic the teacher's superficial style rather than internalizing the underlying logic [Gudibande *et al.*, 2023], thereby limiting generalization capabilities. While subsequent methods attempt to mitigate this by using multiple rationales to enforce answer consistency [Chen *et al.*, 2023] or employing decoupled LoRA experts for different reasoning tasks. However, these methods are still limited by their dependence on a single teacher model. This poses a significant limitation: as LLMs often have specializations on specific domains, they inevitably exhibit capability biases and susceptibility to catastrophic forgetting [Kirkpatrick *et al.*, 2017; Liu *et al.*, 2024]. Consequently, no single teacher dominates across all domains. Instead, each model offers unique strengths and specialized knowledge [Jiang *et al.*, 2023; Wang *et al.*, 2024]. This necessitates a multi-teacher distillation framework that cultivates a more robust student.

Despite the appeal of collaborative distillation, effectively synthesizing the reasoning capabilities of different teachers remains challenging. First, a fundamental incompatibility exists between teacher distinctiveness and student adaptability. Recent studies indicate that teachers imprint recognizable "reasoning signatures" [Chen *et al.*, 2025; Gudibande *et al.*, 2023]; mismatches in these distributions can severely disrupt the transfer of long-horizon reasoning chains [Wu *et al.*,

2025; Magister *et al.*, 2023]. This issue is exacerbated by inevitable noise in teacher-generated rationales that makes mixing diverse reasoning paths without proper filtering risky. It could amplify hallucinations and degrade the reasoning precision of the distilled models. Second, the definition of an "optimal teacher" changes during distillation, depending on the dataset and the student's current learning progress. Existing methods largely remain agnostic to this adaptability. They treat SLMs as passive imitators, lacking mechanisms to ensure genuine logic comprehension rather than superficial pattern mimicry, thus failing to align with the student's evolving receptive capacity. Third, determining the optimal fusion paradigm, specifically between supervision-level integration and parameter-level merging, remains an unresolved dilemma. Supervision-level approaches are intrinsically constrained by the difficulty of quantifying teacher-student compatibility in advance. Conversely, while model merging has been explored as a promising alternative [Ilharco *et al.*, 2023; Matena and Raffel, 2022; Mitchell *et al.*, 2022], static parameter addition or averaging (e.g., [Shen *et al.*, 2025]) often induces conflicts within the parameter space in distillation tasks. This leads to destructive interference or "cancellation effects" among diverse supervision signals, resulting in suboptimal performance. These challenges necessitate a more adaptive, student-centric fusion mechanism.

To address these challenges, we propose **COMP**atibility-**A**ware Multi-teacher **C**oT Distilla**T**ion (*COMPACT*). Unlike static parameter merging that often incurs parameter conflicts, *COMPACT* employs a dynamic weighting mechanism that treats teacher-specific updates as independent task vectors to avoid interferences. This approach allows the student to selectively synthesize gradients from the most compatible and high-quality teachers at the instance level, effectively preventing the "cancellation effects" in post-hoc merging. Central to this framework is a **Multi-Dimensional Evaluation** mechanism designed to quantify teacher-student compatibility through three distinct metrics: (1) Graph-based Consensus, which constructs a semantic similarity graph of teacher outputs to filter isolated noise and identify mainstream reasoning paths; (2) Mutual-Information-based Adaptability, an information-theoretic probe that detects "epiphany moments" in the student's learning process to distinguish genuine logic comprehension from rote memorization; and (3) Loss-based Difficulty, which assesses the student's receptivity to a teacher's guidance to prevent negative transfer. Furthermore, to ensure robustness across diverse reasoning paths, we impose a consistency constraint that encourages the student to produce stable predictions regardless of different reasoning patterns.

Extensive evaluations across In-Distribution (ID) and Out-of-Distribution (OOD) benchmarks demonstrate that *COMPACT* achieves state-of-the-art performance, significantly enhancing the multi-dimensional capabilities of distilled SLMs. Furthermore, latent space analysis reveals that our method enables the internalization of teacher knowledge without disrupting the student's original representation space, confirming the model's superior ability to mitigate catastrophic forgetting while absorbing diverse reasoning patterns. Our main contributions are summarized as follows:

1. We identify the limitations of single teacher supervision and static parameter merging in CoT distillation, specifically the issues of parameter conflict, and the neglect of students' evolving adaptability.
2. We propose an effective multi-teacher distillation framework that leverages student-centroid multi-dimensional evaluation to adaptively internalize teacher capabilities while effectively mitigating catastrophic forgetting.
3. We creatively introduce Mutual Information (MI) to detect "epiphany moments" during distillation, enabling the distinction between genuine logic internalization and superficial pattern mimicry to enhance generalization.
4. Extensive experiments and latent space analysis confirm that *COMPACT* effectively synthesizes the strengths of diverse teachers, achieving the trade-off between superior performance and stability in preventing forgetting.

2 Related Work

2.1 Chain-of-Thought Distillation from LLMs

While Chain-of-Thought (CoT) reasoning has empowered Large Language Models (LLMs) to tackle tasks that previously seemed impossible, these emergent capabilities are heavily dependent on massive parameter scales. Small Language Models (SLMs) often struggle to actively generate coherent reasoning chains, driving efforts to distill the capabilities of LLMs into compact students [Ho *et al.*, 2022; Magister *et al.*, 2023; Hsieh *et al.*, 2023]. To prevent students from learning superficial shortcuts, methods like SCOTT [Wang *et al.*, 2023] enforce rationale-prediction consistency by contrastive decoding, while [Chen *et al.*, 2024] proposes maximizing mutual information between prediction and explanation tasks. Beyond single-path imitation, MCC-KD [Chen *et al.*, 2023] leverages diverse rationales to regularize predictions, and UniCoTT [Zhuang *et al.*, 2025] proposes a unified framework to transfer knowledge from diverse structural CoTs. Recently, EDIT [Dai *et al.*, 2025] utilizes a mistake-driven mechanism to assist students in distinguishing reasoning steps. However, despite these sophisticated mechanisms, existing paradigms predominantly rely on a single teacher model. This reliance confines the student's potential to the boundaries of a specific teacher's knowledge distribution and reasoning biases, necessitating collaborative multi-teacher distillation that leverages complementary expertise.

2.2 Model Combining and Merging in LLMs

To break the capability bounds of individual models, recent approaches dynamically combine multiple LLMs during inference. Methods like Mixture-of-Agents [Wang *et al.*, 2024] fuse intermediate activations, while LLM-Blender [Jiang *et al.*, 2023] ranks and ensembles final outputs. However, runtime ensembles inevitably incur inference overhead and high latency (Time to First Token). Model merging [Yang *et al.*, 2024] provides a more efficient solution by combining parameters into a single model. Advanced techniques surpass naive averaging by preserving functional capabilities; for instance, Fisher-Weighted Averaging [Matena and Raffel, 2022] utilizes diagonal Fisher information to approximate the posterior distribution, outperforming direct averaging. Similarly,

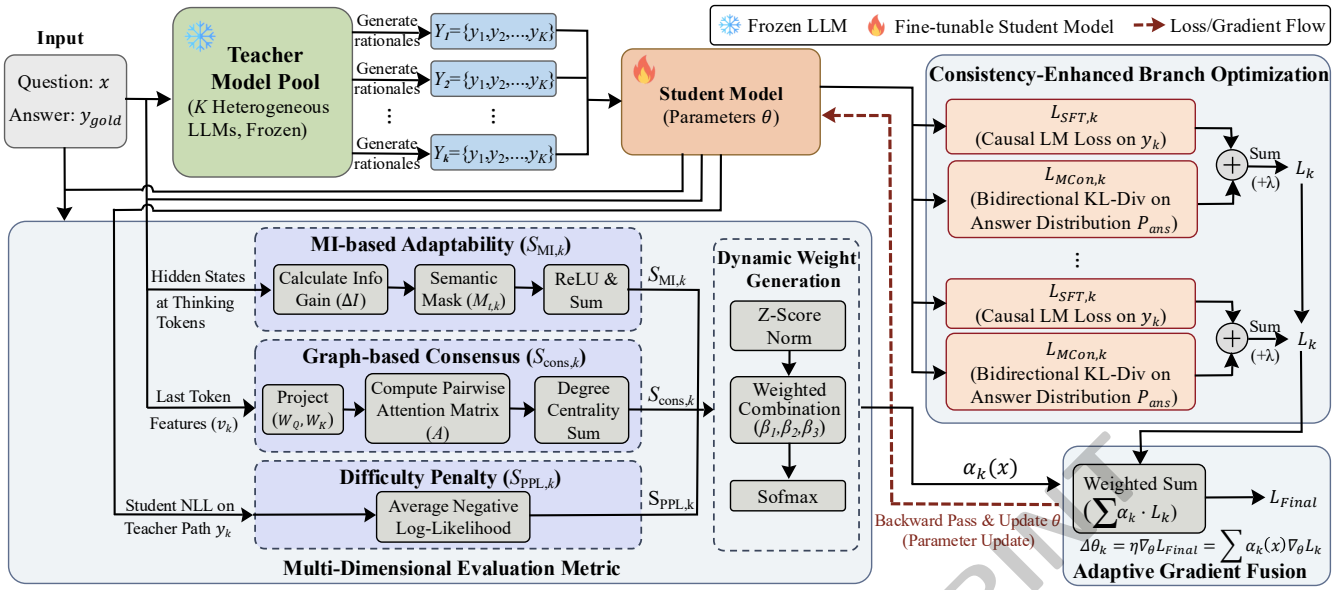

Figure 1: Overview of our *COMPACT* framework. The framework consists of three main components: (1) A frozen Teacher Model Pool generating diverse rationales; (2) A Multi-Dimensional Evaluation Metric that dynamically computes Adaptability (S_{MI}), Consensus (S_{cons}), and Difficulty (S_{PPL}) scores; and (3) An Adaptive Gradient Fusion mechanism that updates the student model using compatibility-aware dynamic weights $\alpha_k(x)$ derived from these scores.

Task Arithmetic [Ilharco *et al.*, 2023] demonstrates that task-specific vectors can be algebraically manipulated to change model behavior, enabling the modular addition or removal of capabilities without retraining. Other works extend these concepts to patch specific model failures [Ilharco *et al.*, 2022; Mitchell *et al.*, 2022] or disentangle weight magnitude from direction to adapt pre-trained models [Yu *et al.*, 2024]. However, these merging frameworks usually assume that tasks are orthogonal in parameter space[Ilharco *et al.*, 2023]. This assumption fails in multi-teacher distillation, where different teachers could provide conflicting reasoning paths for the same problem that leads to interference. In contrast, our approach adapts parameter fusion to the distillation setting, aiming to unify diverse reasoning patterns while dynamically resolving interferences between different teacher distributions.

3 Method

3.1 Preliminaries and Problem Formulation

Let $\mathcal{M}_\theta$ denote the student model parameterized by θ. Given an input question x and a gold answer y_{gold}, we assume access to a set of K heterogeneous teacher models. For each input, these teachers generate a set of diverse rationales (CoTs), denoted as $\mathcal{Y} = \{y_1, y_2, \ldots, y_K\}$. Our goal is to update θ by synthesizing the supervision signals from $\mathcal{Y}$. We formulate the parameter update rule based on the concept of Task Arithmetic [Ilharco *et al.*, 2023]. We treat the gradient from each teacher k as a specific "task vector" $\Delta\theta_k$. The global update is defined as a dynamic weighted summation:

$$\theta_{new} = \theta_{old} + \sum_{k=1}^{K} \alpha_k(x) \cdot \Delta\theta_k \tag{1}$$

where $\alpha_k(x)$ is a dynamic weighting coefficient conditioned on the student's real-time adaptability to teacher k.

3.2 Multi-Dimensional Dynamic Evaluation

To determine the optimal weighting coefficients $\alpha_k(x)$, we introduce a multi-dimensional evaluation mechanism, which assesses each teacher's rationale quality from the student's perspective across three dimensions: adaptability, consensus, and difficulty.

MI-based Adaptability Score (S_{MI}). A core challenge in CoT distillation is determining whether a student is genuinely internalizing the reasoning logic or merely mimicking surface textual patterns. Recent findings in information-theoretic analysis of LLMs [Qian *et al.*, 2025] reveal that the genuine reasoning manifests as "Information Peaks" (sudden reductions in the uncertainty of the final answer), occurring specifically at critical "Thinking Tokens" (e.g., Therefore, Because, Wait). These tokens act as indicators where the model consolidates prior context to deduce the next logical step.

Based on this, we design S_{MI} to measure the information gain triggered by the teacher's rationale at these specific pivotal moments. We monitor the "epiphany moments" where the student's hidden states significantly reduce uncertainty about the final answer y_{gold}. For each token t in the reasoning chain y_k, we compute the log-likelihood of generating the correct answer using the student's current hidden state $h_{t,k}$:

$$I_{proxy}(t,k) = \frac{1}{|y_{gold}|} \sum_{j=1}^{|y_{gold}|} \log P_\theta(y_{gold}^{(j)} | h_{t,k}) \tag{2}$$

We define the information gain as $\Delta I_{t,k} = I_{proxy}(t,k) - I_{proxy}(t-1,k)$. To focus on reasoning steps rather than trivial non-reasoning tokens, we apply a semantic mask $M_{t,k}$ which

assigns higher weights to predefined "Thinking Tokens" (e.g., "Therefore", "So") and a small weight ϵ otherwise. We calculate the final adaptability score by aggregating the positive information gains:

$$S_{MI,k} = \sum_{t=1}^{T_k} \text{ReLU}(\Delta I_{t,k}) \cdot M_{t,k} \tag{3}$$

Where T_k is the length of the reasoning chain. A high $S_{MI,k}$ indicates that the reasoning path provided by teacher k successfully triggers "epiphany moments" in the student's latent space, signaling effective logical transfer.

Graph-based Consensus Score (S_{cons}). Since individual teachers may hallucinate convincing but incorrect reasoning paths, we introduce a consensus score to identify the mainstream logic to filter noises. Unlike external evaluation models, we construct a semantic consensus graph directly within the student's own representation space. This ensures that the selected teacher is not only correct but also aligned with the student's current semantic understanding.

We treat the K reasoning paths as nodes in a fully connected graph. For each path y_k, we extract the feature vector v_k from the student's hidden state at the final valid token (EOS). To measure semantic affinity, we project these vectors using the student model's own attention projection matrices (W_Q and W_K from the final transformer layer), which preserves the model's native similarity metric. We compute the pairwise attention matrix $A \in \mathbb{R}^{K \times K}$:

$$A_{ij} = \text{Softmax}\left(\frac{(Q_{graph})_i \cdot (K_{graph})_j^T}{\sqrt{d}}\right) \tag{4}$$

Where $Q_{graph} = W_Q v, \quad K_{graph} = W_K v$. The consensus score is derived from the degree Centrality of the graph, defined as the sum of incoming attention weights:

$$S_{cons,k} = \sum_{j \neq k} A_{jk} \tag{5}$$

Paths with high centrality represent the mode of the distribution in the student's semantic space, effectively filtering out outlier hallucinations and providing reliable supervision.

Difficulty Penalty (S_{PPL}). While S_{MI} (although related to student's capability) and S_{cons} evaluate the intrinsic quality of the rationales, they ignore the student's receptivity. A rationale may be logically sound but expressed with a complexity or distribution shift that far exceeds the student's current capacity, making it difficult to absorb effectively. To mitigate this, we introduce a difficulty probe to measure the learnability of each reasoning path. We quantify this using the Negative Log-Likelihood (NLL) of the teacher's rationale y_k evaluated by the student model itself. This serves as a direct proxy for the distributional gap between the teacher's supervision and the student's current knowledge state:

$$S_{PPL,k} = -\frac{1}{T_k} \sum_{t=1}^{T_k} \log P_\theta(y_{k,t} | x, y_{k,<t}) \tag{6}$$

A lower $S_{PPL,k}$ indicates the reasoning path lies within the student's "Zone of Proximal Development" [Chen *et al.*, 2025], accessible enough to ensure stable gradient updates while preventing the student from being overwhelmed by excessively perplexing supervision. By integrating this penalty, the framework dynamically lowers the weights of rationales that are incompatible with the student, prioritizing effective knowledge transfer over exposure to complex reasoning.

Dynamic Weight Generation. The final fusion weight α_k is obtained by normalizing the fusion of these three scores:

$$\begin{aligned} Score_k &= \beta_1 \cdot \mathcal{N}(S_{MI,k}) + \beta_2 \cdot \mathcal{N}(S_{cons,k}) - \beta_3 \cdot \mathcal{N}(S_{PPL,k}) \\ \alpha_k &= \frac{\exp(Score_k/\tau)}{\sum_{j=1}^{K} \exp(Score_j/\tau)} \end{aligned} \tag{7}$$

where $\mathcal{N}(\cdot)$ denotes Z-Score normalization among teachers.

3.3 Consistency-Enhanced Branch Optimization

To prevent the student from overfitting to specific stylistic patterns of any single teacher, we design a dual-objective loss function for each teacher branch k.

Supervised Fine-Tuning (SFT). The primary objective is the standard causal language modeling loss on the teacher's path:

$$L_{SFT,k} = -\sum_t \log P(y_{k,t} | x, y_{k,<t}) \tag{8}$$

Multi-CoT Consistency (MCon). We enforce a consistency constraint on the answer span. Even if using different reasoning paths, the probability distribution over the final answer P_a should remain consistent across high-quality teacher paths. We minimize the Bidirectional KL-Divergence between the current branch k and other branches j:

$$L_{MCon,k} = \sum_{j \neq k} \alpha_j \cdot D_{KL}\left(P_a(\cdot|x, y_k) || (P_a(\cdot|x, y_j))\right) \tag{9}$$

This term encourages the student to achieve outcome-consistency regardless of the reasoning topology. The total loss for branch k is:

$$\mathcal{L}_k = L_{SFT,k} + \lambda L_{MCon,k} \tag{10}$$

3.4 Adaptive Parameter Fusion

Finally, we execute the parameter update. While theoretically formulated as the fusion of task vectors in the parameter space (where $\Delta\theta_k$ is equivalent to $\eta \nabla \mathcal{L}_k$), we implement this efficiently by automatic differentiation on the dynamically weighted objective:

$$\mathcal{L}_{Final} = \sum_{k=1}^{K} \alpha_k(x) \cdot \mathcal{L}_k \tag{11}$$

This generate equivalent gradient aggregation $\nabla_\theta \mathcal{L}_{Final} = \sum_{k=1}^{K} \alpha_k(x) \nabla_\theta \mathcal{L}_k$ without the overhead of per-branch gradient storage. By optimizing $\mathcal{L}_{Final}$, the student effectively updates its parameters in the direction of interpolated task vectors. This allows the student to prioritize the most instructive and robust reasoning paths for each specific instance.

Method	In-Distribution				Out-Of-Distribution			
	MATH500	GSM8K	SVAMP	Avg. gain	CSQA	StrategyQA	GPQA-D	Avg. gain
Qwen2.5-7B								
Base (Qwen2.5-7B-Instruct)	77.20	92.36	90.33	↑4.14	83.45	68.68	30.30	↑6.69
SBS [Hsieh *et al.*, 2023]	77.40↑ 0.20	94.77↑ 2.41	93.00↑ 2.67	↑2.38	83.20↓ 0.25	67.25↓ 1.43	27.46↓ 2.84	↑8.20
MCC [Chen *et al.*, 2023]	82.20↑ 5.00	90.52↓ 1.84	91.00↑ 0.67	↑2.86	81.72↓ 1.73	67.03↓ 1.65	26.77↓ 3.53	↑9.00
Committee [Li *et al.*, 2025]	77.25↑ 0.05	88.34↓ 4.02	91.51↑ 1.18	↑5.07	80.52↓ 2.93	64.19↓ 4.49	36.6↑ 6.3	↑7.20
MoT[Shen *et al.*, 2025]	81.60↑ 4.40	94.37↑ 2.01	93.00↑ 2.67	↑1.14	83.20↓ 0.25	66.81↓ 1.87	37.88↑ 7.58	↑5.01
Ours w/ fusion	**82.70**↑ 5.50	**95.30**↑ 2.94	**94.31**↑ 3.98	–	**83.59**↑ 0.14	**72.93**↑ 4.25	**46.39**↑ 16.09	–
Llama3.1-8B								
Base (Llama3.1-8B-Instruct)	46.00	83.89	87.00	↑4.28	74.77	70.74	21.71	↑2.68
SBS [Hsieh *et al.*, 2023]	52.60↑ 6.60	86.96↑ 3.07	87.67↑ 0.67	↑0.84	75.02↑ 0.25	71.05↑ 0.31	19.70↓ 2.01	↑3.16
MCC [Chen *et al.*, 2023]	53.00↑ 7.00	85.01↑ 1.12	83.33↓ 3.67	↑2.80	73.33↓ 1.44	67.99↓ 2.75	18.53↓ 3.18	↑5.13
Committee [Li *et al.*, 2025]	51.93↑ 5.93	81.83↓ 2.06	87.84↑ 0.84	↑2.71	74.95↑ 0.18	70.87↑ 0.13	18.47↓ 3.24	↑3.65
MoT[Shen *et al.*, 2025]	50.80↑ 4.80	86.73↑ 2.83	85.00↓ 2.00	↑2.40	74.94↑ 0.17	67.25↓ 3.49	23.74↑ 2.03	↑3.11
Ours w/ fusion	**53.56**↑ 7.56	**88.25**↑ 4.36	**87.93**↑ 0.93	–	**76.26**↑ 1.49	**71.72**↑ 0.98	**27.27**↑ 5.56	–
Qwen2.5-1.5B								
Base (Qwen2.5-1.5B-Instruct)	50.40	72.81	79.33	↑3.25	74.20	58.07	24.75	↑4.14
SBS [Hsieh *et al.*, 2023]	49.80↓ 0.60	73.91↑ 1.10	78.52↓ 0.81	↑3.35	73.46↓ 0.74	56.55↓ 1.52	22.99↓ 1.76	↑5.48
MCC [Chen *et al.*, 2023]	53.60↑ 3.20	70.54↓ 2.27	79.80↑ 0.47	↑2.78	71.32↓ 2.88	56.99↓ 1.08	20.71↓ 4.04	↑6.80
Committee [Li *et al.*, 2025]	58.10↑ 7.70	73.69↑ 0.88	77.98↓ 1.35	↑0.84	71.21↓ 2.99	59.43↑ 1.36	27.07↑ 2.32	↑3.91
MoT[Shen *et al.*, 2025]	54.00↑ 3.60	74.45↑ 1.64	75.25↓ 4.08	↑2.86	71.79↓ 2.41	58.08↑ 0.01	28.71↑ 3.96	↑3.62
Ours w/ fusion	**56.20**↑ 5.80	**75.08**↑ 2.27	**81.00**↑ 1.67	–	**75.37**↑ 1.17	**58.73**↑ 0.66	**35.33**↑ 10.58	–

Table 1: Performance comparison of COMPACT and other methods. ↑ and ↓ indicate the relative improvement/decrease to the base model. Avg. gain calculates the average relative gain of our method toward baselines. All baselines are trained on their original settings.

4 Experiments

4.1 Experimental Setup

Datasets. We evaluate *COMPACT* on several benchmarks to verify performance and generalization: (1) In-Distribution mathematical reasoning benchmarks: MATH500, GSM8K, SVAMP. (2) Out-Of-Distribution commonsense benchmarks: CSQA, StrategyQA, GPQA-Diamond.

Implementation Details. We employ DeepSeek-R1-Llama-70B, Qwen2.5-70B, Llama-3.1-70B and QWQ-32B as teacher models. For students, we select Qwen2.5-1.5B, Qwen2.5-7B, and Llama3.1-8B to evaluate scalability and architectural universality. We train models for 8 epochs(1.5B) and 6 epochs(7B/8B) with a learning rate of 1×10^{-4} for 7B/8B models and 5×10^{-5} for the 1.5B model.

Baselines. We evaluate *COMPACT* against diverse baselines: (1) Zero-shot base models, specifically Qwen2.5-1.5B-Instruct, Qwen2.5-7B-Instruct, and Llama3.1-8B-Instruct; (2) SbS-KD [Hsieh *et al.*, 2023], representing vanilla CoT distillation; (3) MCC-KD [Chen *et al.*, 2023], a multi-path approach enforcing consistency; (4) MoT [Shen *et al.*, 2025], baseline method that averages parameters over fine-tuned students; (5) Committee [Li *et al.*, 2025], a mistakes-aware method that simulates a peer-review among teachers.

4.2 Main Results

Table 1 presents the performance of our method against baselines across three student architectures. *COMPACT* consistently outperforms all baselines on In-Distribution (ID) reasoning tasks. Compared to multi-teacher baselines like MoT and Committee, *COMPACT* demonstrates superior performance and robustness. The performance gains are consistent across diverse model families and scales. Notably, the 1.5B student distilled via *COMPACT* even surpasses the 7B and 8B Base model on GPQA-D, highlighting the method's ability to compress reasoning capabilities into lightweight models efficiently. Crucially, *COMPACT* excels in preserving general capabilities. While baselines often suffer from catastrophic forgetting, our method achieves improvement on evaluated OOD tasks. This confirms that our compatibility-aware weighting effectively isolates reasoning enhancement from general knowledge retention, preventing the tax on versatility often paid by standard distillation. Figure 3 shows the dynamic weighting scores for different teachers during training on MATH dataset. Detailed analysis see Appendix C.

4.3 Latent Representation Stability Analysis

To investigate the internal adaptation of the student model during distillation, we conducted a PCA analysis on the latent representations of the layers in Qwen2.5-1.5B distilled by different methods. Following the framework proposed by [Xu *et al.*, 2025] and [Wu *et al.*, 2025], we project the activations of the distilled students onto the principal component basis constructed from the original, undistilled student.

Methodology and Rationale. The PCA shift (Δ) measures the Euclidean distance between the centroids of latent representations before and after distillation. A large shift indicates a drastic reorganization of the student's feature space, often signaling *structural degradation* or *low corpus affinity*. An ideal distillation should achieve *knowledge absorption without structural destruction*. This requires the student to exhibit controlled adaptation on in-distribution tasks (learning reasoning patterns) while maintaining near-zero shift on out-of-distribution tasks (preserving general capabilities).

Analysis on In-Distribution (ID) Reasoning Tasks. Figure

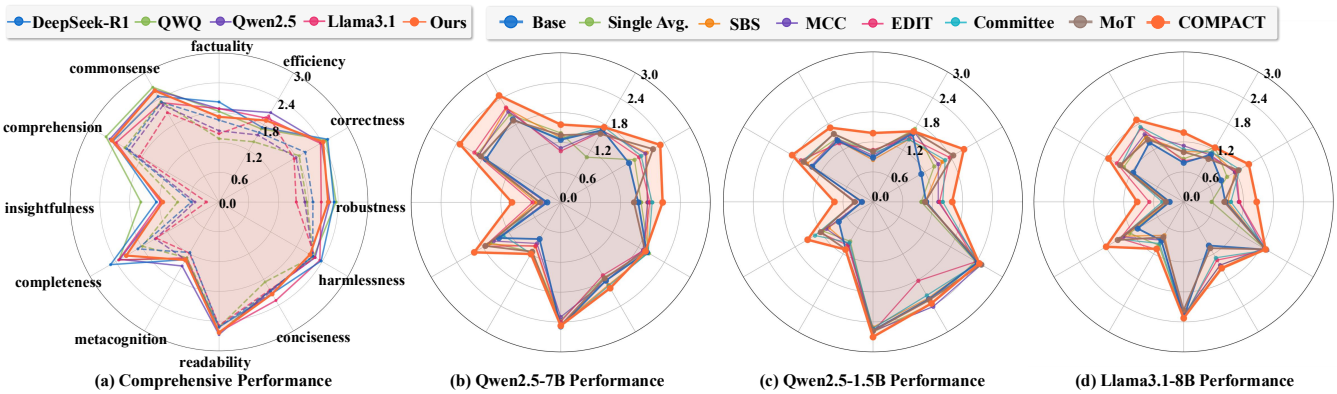

Figure 2: Fine-grained capability evaluation on FLASK framework. (a) compares teachers (DeepSeek-R1-Distill-Llama-70B, QWQ-32B, Qwen2.5-72B, Llama3.1-70B)(solid lines) and their corresponding distilled students (dashed lines), alongside *COMPACT*. (b)(c)(d) show the performance of different distillation methods in Table 1 across three student architectures. All results are scaled for better illustration.

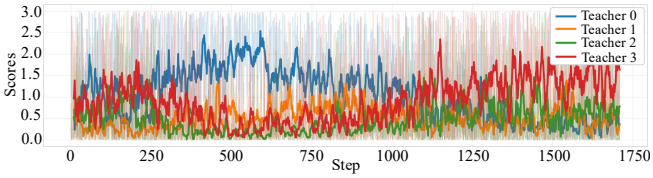

Figure 3: Evolution of dynamic weighting scores for different teachers during training. Teacher 0-3 corresponds to Qwen2.5-72B, Llama3.1-70B, DeepSeek-R1-Llama-70B, and QWQ-32B.

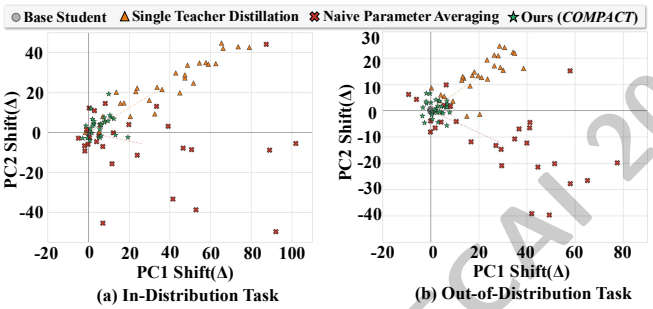

Figure 4: Layer-wise PCA shift of latent representation in Qwen2.5-1.5B distilled by various methods. *COMPACT* exhibits minimal shift, confirming its stability in preserving the general capabilities.

4(a) visualizes the representational shift on the ID reasoning dataset (MATH500). *Naive Parameter Averaging* exhibits the largest dispersion (high shift), confirming our hypothesis that static averaging induces gradient conflicts among heterogeneous teachers. *Single Teacher Distillation* shows a moderate and directional shift, reflecting the *Style Signature* effect, where the student is forced to reorganize its cognitive structure to mimic a specific teacher's distribution, potentially at the cost of its original structural stability. In contrast, *COMPACT* maintains a representation centroid tightly aligned with the original student with minimal shift. This suggests that *COMPACT* successfully identifies the most compatible paths, embedding new reasoning capabilities smoothly into the student's existing manifold without destructive reorganization.

Analysis on Out-of-Distribution (OOD) Tasks. Figure 4(b) illustrates the shift on OOD datasets (CSQA), serving as a proxy for catastrophic forgetting. Ideally, the latent representation of OOD tasks should remain invariant to preserve pre-existing knowledge. However, baselines exhibit substantial distributional shifts, indicating their indiscriminate supervision damages the student's original knowledge space. Conversely, *COMPACT* achieves near-zero shift. This confirms that our compatibility-aware weighting successfully disentangles reasoning internalization from general knowledge retention, thereby preserving the student's general capabilities.

4.4 Fine-grained Capability Analysis

To conduct a holistic assessment beyond simple accuracy, we employ the FLASK evaluation framework [Ye *et al.*, 2023], which decomposes model capabilities into four dimensions comprising 12 fine-grained skills across four primary abilities: Logical Thinking (*Correctness, Robustness, Efficiency*), Background Knowledge (*Factuality, Commonsense, Metacognition*), Problem Handling (*Comprehension, Completeness, Insightfulness*), and User Alignment (*Readability, Conciseness, Harmlessness*). Results are shown in Figure 2.

Breaking Capability Boundaries through Adaptive Fusion. Figure 2(a) confirms that different teachers excel in specific domains and the *style signature* phenomenon. While the naive parameter averaging (MoT) attempts to combine these abilities and shows gains in *Logical Correctness* and *Completeness*, it suffers a critical failure in *Logical Robustness*, accompanied by increased hallucinations (lower *Factuality* and *Commonsense Understanding*). This confirms that conflicting gradient directions neutralize the specialized expertise of individual teachers while disrupting the student's original knowledge structure, ultimately leading to mediocrity. In contrast, leveraging a dynamic weighting mechanism, the *COMPACT*-distilled student effectively breaks these capability boundaries, synthesizing the complementary strengths of diverse teachers to achieve superior performance.

Robustness and Genuine Comprehension. Specific capability gains directly validate our multi-dimensional evaluation metrics. (1) Robustness via Consensus: *COMPACT* holds a significant lead in Factuality and Commonsense Understanding. We attribute this to the Graph-based Consensus (S_{cons}) metric, which effectively filters out hallucinations from in-

dividual teachers, ensuring the student internalizes only the mainstream correct logic. (2) Insightfulness via Information Gain: The superior performance in *Insightfulness* and *Completeness* validates the MI-based Adaptability (S_{MI}) metric. Unlike baselines that rely on rote memorization of surface patterns, the high mutual information guidance fosters genuine logic comprehension, empowering the student to generate more complete and novel reasoning paths.

4.5 Ablation Study

To rigorously verify the contribution of each component, we conducted ablation studies on the Qwen2.5-1.5B student model. The results are summarized in Figure 5. We calculated the average relative degradation rate (%) and the shift in stability (Standard Error of the Mean, SEM) across all tasks.

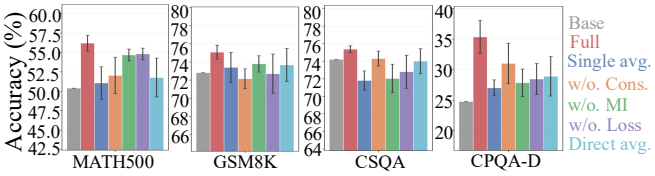

Figure 5: Comprehensive ablation study on model performance.

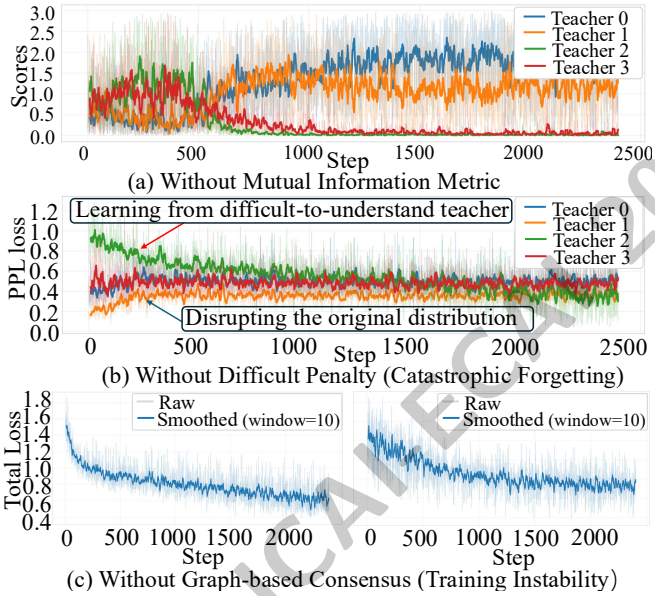

Figure 6: Training dynamics under component ablation.

Superiority of Dynamic Weighting over Static Merging. The *"Direct Avg."* baseline, while improving upon the base model, consistently underperforms the full *COMPACT* framework. Notably, *"Direct Avg."* baseline exhibits significantly higher variance (e.g., SEM of 1.85 on average). This instability suggests that without compatibility-aware weighting, the student is subject to destructive gradient interference, whereas *COMPACT*'s dynamic weighting stabilizes the learning trajectory by selecting the most compatible paths.

Effectiveness of Graph-based Consensus (S_{cons}). Removing S_{cons} results in a 4.60% performance drop on ID reasoning tasks and a 5.51% drop on OOD tasks. More critically, the lack of consensus filtering significantly amplifies the variance of the student's performance: the average SEM doubles ($2.08\times$) on ID tasks and increases by $1.19\times$ on OOD tasks. As illustrated in Figure 6(c), the removal introduces significant stochasticity into the optimization landscape that significantly hampers convergence efficiency and leads to an elevated final loss plateau. This validates S_{cons} is essential for mitigating gradient conflicts and ensuring stable convergence.

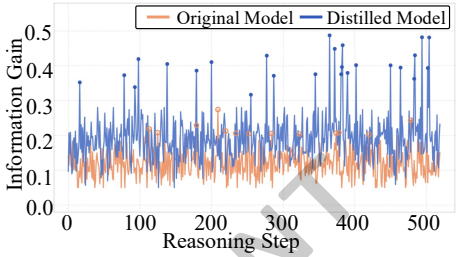

Figure 7: Comparison of mutual information trajectories.

Effectiveness of MI-based Adaptability (S_{MI}). The removal of the mutual information regularization causes the steepest decline in generalization capabilities, with a significant 9.65% average drop across OOD tasks and $1.11\times$ SEM increase on ID stability. Figure 6(a) also evidences a "collapse to mediocrity", where the student model exclusively exploits the low-entropy supervision of less capable generalist teachers (Qwen2.5/Llama3.1) while discarding the high-fidelity but complex reasoning of experts (QWQ/DeepSeek-R1). Figure 7 confirms our *COMPACT* framework successfully transfers the "epiphany" capabilities of strong teachers to the student. The distilled student model demonstrates more and sharper peaks while the original student model exhibits a relatively flat and low-entropy trajectory.

Effectiveness of Loss-based Difficulty (S_{PPL}). Excluding the difficulty penalty (S_{PPL}) leads to comprehensive degradation, with an average drop of 2.80% on ID and 8.60% on OOD tasks. Its absence triggers higher volatility: ID variance increases by $1.95\times$ and OOD variance by $1.39\times$. This also exacerbates catastrophic forgetting (Figure 6(b)); the student overfits to the hard distribution of Teacher 2 at the expense of its proficiency with the generalist teachers, validating the penalty's role as a crucial compatibility adjuster that prevents distributional drift and preserves general capabilities.

5 Conclusion

In this work, we propose *COMPACT*, a novel compatibility-aware multi-teacher distillation framework that dynamically fuses heterogeneous teacher gradients through instance-level weighting. By integrating multi-dimensional metrics for consensus, adaptability, and difficulty, *COMPACT* successfully synthesizes the complementary strengths of diverse teachers, effectively resolving the "cancellation effect" of static merging and minimizing representational drift. Extensive experiments demonstrate that our method achieves state-of-the-art performance with high data efficiency while preserving the student's general capabilities, offering a robust approach for distilling diverse intelligence into compact models.

Contribution Statement

Jin Cui and Jiaqi Guo contributed equally to this work. Boran Zhao is the corresponding author.

Acknowledgements

This work was supported in part by Fundamental and Interdisciplinary Disciplines Breakthrough Plan of the Ministry of Education of China under Grant JYB2025XDXM504, and National Natural Science Foundation of China No.62302381, No.52441602. The Authors are with the National Key Laboratory of Human-Machine Hybrid Augmented Intelligence and Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an, Shaanxi, China.

References

[Chen *et al.*, 2023] Hongzhan Chen, Siyue Wu, Xiaojun Quan, Rui Wang, Ming Yan, and Ji Zhang. MCC-KD: Multi-CoT Consistent Knowledge Distillation. *arXiv preprint arXiv:2310.14747*, 2023.

[Chen *et al.*, 2024] Xin Chen, Hanxian Huang, Yanjun Gao, Yi Wang, Jishen Zhao, and Ke Ding. Learning to maximize mutual information for chain-of-thought distillation. In *Annual Meeting of the Association for Computational Linguistics*, 2024.

[Chen *et al.*, 2025] Xinghao Chen, Zhijing Sun, Wenjin Guo, Miaoran Zhang, Yanjun Chen, Yirong Sun, Hui Su, Yijie Pan, Dietrich Klakow, Wenjie Li, and Xiaoyu Shen. Unveiling the key factors for distilling chain-of-thought reasoning. *arXiv preprint arXiv:2502.18001*, 2025.

[Dai *et al.*, 2025] Chengwei Dai, Kun Li, Wei Zhou, and Songlin Hu. Capture the key in reasoning to enhance cot distillation generalization. In *Annual Meeting of the Association for Computational Linguistics*, 2025.

[Gudibande *et al.*, 2023] Arnav Gudibande, Eric Wallace, Charles Burton Snell, Xinyang Geng, Hao Liu, P. Abbeel, Sergey Levine, and Dawn Song. The false promise of imitating proprietary llms. *arXiv preprint arXiv:2305.15717*, 2023.

[Ho *et al.*, 2022] Namgyu Ho, Laura Schmid, and Se-Young Yun. Large language models are reasoning teachers. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2022.

[Hsieh *et al.*, 2023] Cheng-Yu Hsieh, Chun-Liang Li, Chih-Kuan Yeh, Hootan Nakhost, Yasuhisa Fujii, Alex Ratner, Ranjay Krishna, Chen-Yu Lee, and Tomas Pfister. Distilling step-by-step! outperforming larger language models with less training data and smaller model sizes. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 8003–8017, 2023.

[Ilharco *et al.*, 2022] Gabriel Ilharco, Mitchell Wortsman, Samir Yitzhak Gadre, Shuran Song, Hannaneh Hajishirzi, Simon Kornblith, Ali Farhadi, and Ludwig Schmidt. Patching open-vocabulary models by interpolating weights. *Advances in Neural Information Processing Systems*, 35:29262–29277, 2022.

[Ilharco *et al.*, 2023] Gabriel Ilharco, Marco Tulio Ribeiro, Mitchell Wortsman, Suchin Gururangan, Ludwig Schmidt, Hannaneh Hajishirzi, and Ali Farhadi. Editing models with task arithmetic. In *International Conference on Learning Representations (ICLR)*, 2023.

[Jiang *et al.*, 2023] Dongfu Jiang, Xiang Ren, and Bill Yuchen Lin. LLM-Blender: Ensembling Large Language Models with Pairwise Ranking and Generative Fusion. In *Annual Meeting of the Association for Computational Linguistics*, 2023.

[Kirkpatrick *et al.*, 2017] James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, et al. Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences*, 114(13):3521–3526, 2017.

[Kojima *et al.*, 2022] Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. Large language models are zero-shot reasoners. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 35, pages 22199–22213, 2022.

[Li *et al.*, 2025] Zhuochun Li, Yuelyu Ji, Rui Meng, and Daqing He. Learning from committee: Reasoning distillation from a mixture of teachers with peer-review. *arXiv preprint arXiv:2410.03663*, 2025.

[Liu *et al.*, 2024] Chengyuan Liu, Shihang Wang, Yangyang Kang, Lizhi Qing, Fubang Zhao, Changlong Sun, Kun Kuang, and Fei Wu. More than catastrophic forgetting: Integrating general capabilities for domain-specific llms. In *Conference on Empirical Methods in Natural Language Processing*, 2024.

[Magister *et al.*, 2023] Lucie Charlotte Magister, Jonathan Mallinson, Jakub Adamek, Eric Malmi, and Aliaksei Severyn. Teaching small language models to reason. In *Proceedings of the 61st annual meeting of the association for computational linguistics (volume 2: short papers)*, pages 1773–1781, 2023.

[Matena and Raffel, 2022] Michael Matena and Colin Raffel. Merging models with fisher-weighted averaging. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 35, pages 22458–22469, 2022.

[Mitchell *et al.*, 2022] Eric Mitchell, Charles Lin, Antoine Bosselut, Chelsea Finn, and Christopher D Manning. Fast model editing at scale. In *International Conference on Learning Representations (ICLR)*, 2022.

[Qian *et al.*, 2025] Chen Qian, Dongrui Liu, Hao Wen, Zhen Bai, Yong Liu, and Jing Shao. Demystifying reasoning dynamics with mutual information: Thinking tokens are information peaks in llm reasoning. *arXiv preprint arXiv:2506.02867*, 2025.

[Shen *et al.*, 2025] Zhanming Shen, Zeyu Qin, Zenan Huang, Haoxing Chen, Jiaqi Hu, Yihong Zhuang, Guoshan Lu, Gang Chen, and Junbo Zhao. Merge-of-thought distillation. *arXiv preprint arXiv:2509.08814*, 2025.

[Wang *et al.*, 2023] Peifeng Wang, Zhengyang Wang, Zheng Li, Yifan Gao, Bing Yin, and Xiang Ren. Scott: Self-consistent chain-of-thought distillation. *arXiv preprint arXiv:2305.01879*, 2023.

[Wang *et al.*, 2024] Junlin Wang, Jue Wang, Ben Athiwaratkun, Ce Zhang, and James Zou. Mixture-of-agents enhances large language model capabilities. *arXiv preprint arXiv:2406.04692*, 2024.

[Wei *et al.*, 2022] Jason Wei, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama, Maarten Bosma, Denny Zhou, Donald Metzler, et al. Emergent abilities of large language models. *arXiv preprint arXiv:2206.07682*, 2022.

[Wu *et al.*, 2025] Xiaojun Wu, Xiaoguang Jiang, Huiyang Li, Jucai Zhai, Dengfeng Liu, Qiaobo Hao, Huang Liu, Zhiguo Yang, Ji Xie, Ninglun Gu, Jin Yang, Kailai Zhang, Yelun Bao, and Jun Wang. Beyond scaling law: A data-efficient distillation framework for reasoning. *arXiv preprint arXiv:2508.09883*, 2025.

[Xu *et al.*, 2025] Xiaoyu Xu, Xiang Yue, Yang Liu, Qingqing Ye, Haibo Hu, and Minxin Du. Unlearning isn't deletion: Investigating reversibility of machine unlearning in llms. *arXiv preprint arXiv:2505.16831*, 2025.

[Yang *et al.*, 2024] Enneng Yang, Li Shen, Guibing Guo, Xingwei Wang, Xiaochun Cao, Jie Zhang, and Dacheng Tao. Model merging in llms, mllms, and beyond: Methods, theories, applications, and opportunities. *ACM Computing Surveys*, 2024.

[Ye *et al.*, 2023] Seonghyeon Ye, Doyoung Kim, Sungdong Kim, Hyeonbin Hwang, Seungone Kim, Yongrae Jo, James Thorne, Juho Kim, and Minjoon Seo. FLASK: Fine-grained Language Model Evaluation based on Alignment Skill Sets. *arXiv preprint arXiv:2307.10928*, 2023.

[Yu *et al.*, 2024] Le Yu, Bowen Yu, Haiyang Yu, Fei Huang, and Yongbin Li. Extend model merging from fine-tuned to pre-trained large language models via weight disentanglement. *arXiv preprint arXiv:2408.03092*, 2024.

[Zhuang *et al.*, 2025] Xianwei Zhuang, Zhihong Zhu, Zhichang Wang, Xuxin Cheng, and Yuexian Zou. UniCoTT: A Unified Framework for Structural Chain-of-Thought Distillation. In *International Conference on Learning Representations*, 2025.