"The Whole Is Greater Than the Sum of Its Parts": A MULTI-TEACHER COT DISTILLATION FRAMEWORK

Anonymous authors
Paper under double-blind review

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 supervisions 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.

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 (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. 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 in 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; each model offers unique strengths and specialized knowledge*** (Jiang et al., 2023; Wang et al., 2024). This necessitates distilling multi-teacher capabilities to form 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 distribu-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

tions 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 & 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 interference. 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 parameter conflict and 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 CoT reasoning has empowered LLMs to tackle tasks that previously seemed to be 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 those 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. 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) also adopts a mistake-driven mechanism. However, existing paradigms mainly rely on a single teacher. This reliance confines the student's potential to the boundaries of a specific teacher's knowledge and reasoning biases, necessitating a multi-teacher approach that leverages complementary expertise.

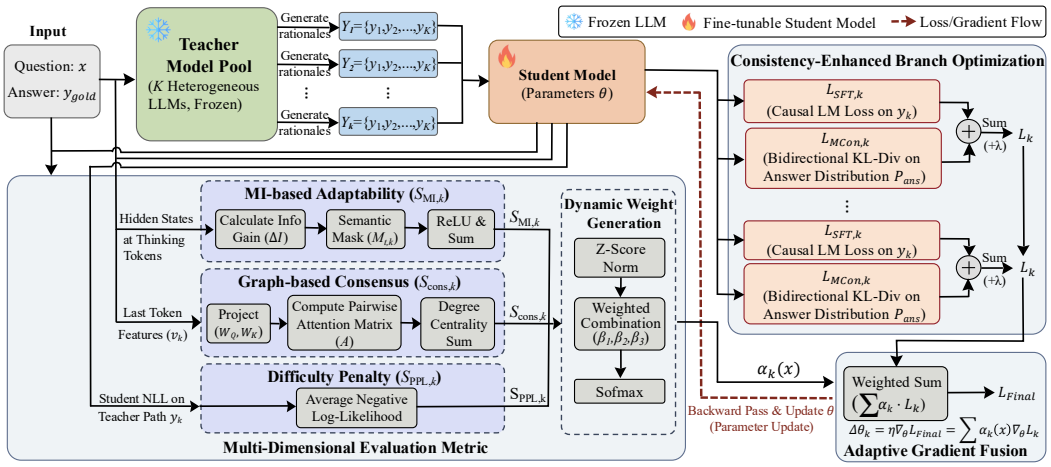


Figure 1: Overview of *COMPACT* framework, which consists of three main components: (1) A frozen ***Teacher Model Pool*** generating diverse rationales; (2) A ***Multi-Dimensional Evaluation Metric*** that dynamically computes Adaptability (S_{MI}), Consensus (S_{cons}), and Difficulty (S_{PPL}) scores;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

2.2 Model Combining and Merging in LLMs.

To break the capability bounds of individual models, recent approaches 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. 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 & Raffel, 2022) utilizes diagonal Fisher information to approximate the posterior distribution. Task Arithmetic (Ilharco et al., 2023) demonstrates 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 (Yu et al., 2024). However, these merging frameworks usually assume that tasks are orthogonal in parameter space. This assumption fails in multi-teacher distillation, where different teachers could provide conflicting reasoning paths that lead 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 coefficient conditioned on 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 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 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 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. A high $S_{MI,k}$ indicates that the reasoning path provided by teacher k triggers "epiphany moments" in the student, 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 noise. 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 the 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	CSQA	StrategyQA	GPQA-D
Qwen2.5-7B						
Base (Qwen2.5-7B-Instruct)	77.20	92.36	90.33	83.45	68.68	30.30
SBS (Hsieh et al., 2023)	77.40$_{\uparrow 0.20}$	94.77$_{\uparrow 2.41}$	93.00$_{\uparrow 2.67}$	83.20$_{\downarrow 0.25}$	67.25$_{\downarrow 1.43}$	27.46$_{\downarrow 2.84}$
MCC (Chen et al., 2023)	82.20$_{\uparrow 5.00}$	90.52$_{\downarrow 1.84}$	91.00$_{\uparrow 0.67}$	81.72$_{\downarrow 1.73}$	67.03$_{\downarrow 1.65}$	26.77$_{\downarrow 3.53}$
Committee (Li et al., 2025)	77.25$_{\uparrow 0.05}$	88.34$_{\downarrow 4.02}$	91.51$_{\uparrow 1.18}$	80.52$_{\downarrow 2.93}$	64.19$_{\downarrow 4.49}$	36.60$_{\uparrow 6.30}$
MoT (Shen et al., 2025)	81.60$_{\uparrow 4.40}$	94.37$_{\uparrow 2.01}$	93.00$_{\uparrow 2.67}$	83.20$_{\downarrow 0.25}$	66.81$_{\downarrow 1.87}$	37.88$_{\uparrow 7.58}$
Ours w/ fusion	**82.70**$_{\uparrow 5.50}$	**95.30**$_{\uparrow 2.94}$	**94.31**$_{\uparrow 3.98}$	**83.59**$_{\uparrow 0.14}$	**72.93**$_{\uparrow 4.25}$	**46.39**$_{\uparrow 16.09}$
Llama3.1-8B						
Base (Llama3.1-8B-Instruct)	46.00	83.89	87.00	74.77	70.74	21.71
SBS (Hsieh et al., 2023)	52.60$_{\uparrow 6.60}$	86.96$_{\uparrow 3.07}$	87.67$_{\uparrow 0.67}$	75.02$_{\uparrow 0.25}$	71.05$_{\uparrow 0.31}$	19.70$_{\downarrow 2.01}$
MCC (Chen et al., 2023)	53.00$_{\uparrow 7.00}$	85.01$_{\uparrow 1.12}$	83.33$_{\downarrow 3.67}$	73.33$_{\downarrow 1.44}$	67.99$_{\downarrow 2.75}$	18.53$_{\downarrow 3.18}$
Committee (Li et al., 2025)	51.93$_{\uparrow 5.93}$	81.83$_{\downarrow 2.06}$	87.84$_{\uparrow 0.84}$	74.95$_{\uparrow 0.18}$	70.87$_{\uparrow 0.13}$	18.47$_{\downarrow 3.24}$
MoT (Shen et al., 2025)	50.80$_{\uparrow 4.80}$	86.73$_{\uparrow 2.83}$	85.00$_{\downarrow 2.00}$	74.94$_{\uparrow 0.17}$	67.25$_{\downarrow 3.49}$	23.74$_{\uparrow 2.03}$
Ours w/ fusion	**53.56**$_{\uparrow 7.56}$	**88.25**$_{\uparrow 4.36}$	**87.93**$_{\uparrow 0.93}$	**76.26**$_{\uparrow 1.49}$	**71.72**$_{\uparrow 0.98}$	**27.27**$_{\uparrow 5.56}$
Qwen2.5-1.5B						
Base (Qwen2.5-1.5B-Instruct)	50.40	72.81	79.33	74.20	58.07	24.75
SBS (Hsieh et al., 2023)	49.80$_{\downarrow 0.60}$	73.91$_{\uparrow 1.10}$	78.52$_{\downarrow 0.81}$	73.46$_{\downarrow 0.74}$	56.55$_{\downarrow 1.52}$	22.99$_{\downarrow 1.76}$
MCC (Chen et al., 2023)	53.60$_{\uparrow 3.20}$	70.54$_{\downarrow 2.27}$	79.80$_{\uparrow 0.47}$	71.32$_{\downarrow 2.88}$	56.99$_{\downarrow 1.08}$	20.71$_{\downarrow 4.04}$
Committee (Li et al., 2025)	58.10$_{\uparrow 7.70}$	73.69$_{\uparrow 0.88}$	77.98$_{\downarrow 1.35}$	71.21$_{\downarrow 2.99}$	59.43$_{\uparrow 1.36}$	27.07$_{\uparrow 2.32}$
MoT (Shen et al., 2025)	54.00$_{\uparrow 3.60}$	74.45$_{\uparrow 1.64}$	75.25$_{\downarrow 4.08}$	71.79$_{\downarrow 2.41}$	58.08$_{\uparrow 0.01}$	28.71$_{\uparrow 3.96}$
Ours w/ fusion	**56.20**$_{\uparrow 5.80}$	**75.08**$_{\uparrow 2.27}$	**81.00**$_{\uparrow 1.67}$	**75.37**$_{\uparrow 1.17}$	**58.73**$_{\uparrow 0.66}$	**35.33**$_{\uparrow 10.58}$

Table 1: Performance comparison of COMPACT and other methods. ↑ and ↓ indicate the relative improvement/decrease compared with the corresponding base model. Baselines are trained under their official 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 (Lightman et al., 2023), GSM8K (Cobbe et al., 2021), SVAMP (Patel et al., 2021). (2) Out-Of-Distribution commonsense benchmarks: CSQA (Talmor et al., 2019), StrategyQA (Geva et al., 2021), GPQA-Diamond (Rein et al., 2023).

Models and Implementation Details. We employ open-source DeepSeek-R1-Llama-70B, Qwen2.5-70B, Llama-3.1-70B and QWQ-32B as teacher models. For students, we select widely-used Qwen2.5-1.5B, Qwen2.5-7B, and Llama3.1-8B to evaluate scalability and architectural universality. We train the models for 8 epochs(1.5B) and 6 epochs(7B/8B) with a batch size of 4 per device and 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 distillation approach enforcing consistency; (4) MoT (Shen et al., 2025), baseline method that averages parameters over all students fine-tuned on various teachers; (5) Committee (Li et al., 2024), a mistakes-aware method that simulates a peer-review among multiple teachers.

4.2 Main Results

Table **??** 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

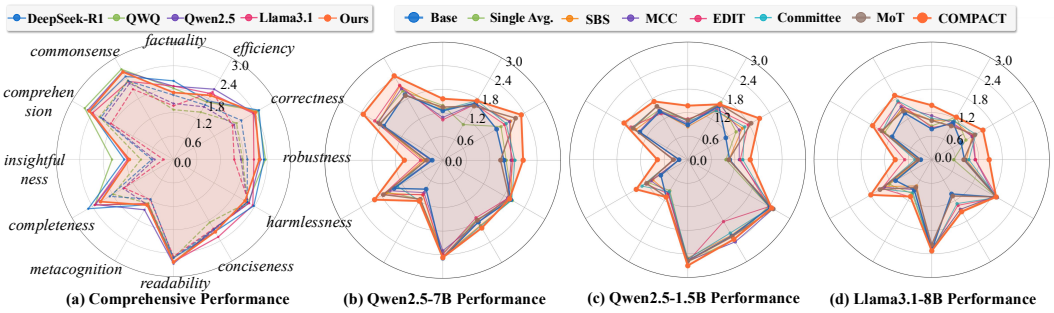


Figure 2: Fine-grained capability evaluation on FLASK framework. (a) compares teachers (DeepSeek-R1-Distill-Llama-70B, QWQ-32B, Qwen2.5-72B, Llama3.1-70B)(solid lines) and their corresponding distilled students (dashed lines), alongside *COMPACT*. (b)(c)(d) show the performance of different distillation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

distillation. Figure 3 shows the dynamic weighting scores for different teachers during training on MATH dataset. Detailed analysis see Appendix C.

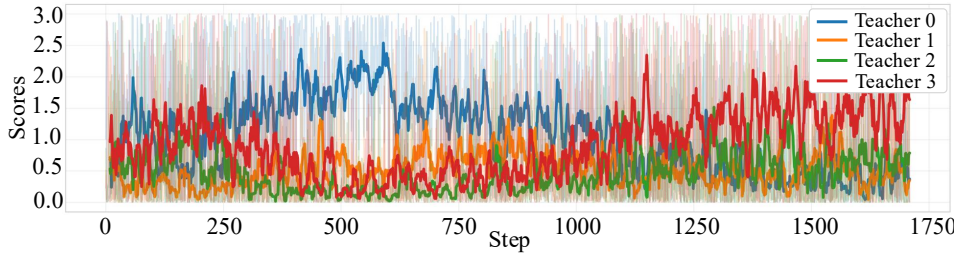


Figure 3: Evolution of dynamic weighting scores for different teachers during training. Teacher 0-3 corresponds to Qwen2.5-72B, Llama3.1-70B, DeepSeek-R1-Llama-70B, and QWQ-32B.

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 (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, particularly in the first and secondary principal components, 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 2(a) visualizes the representational shift on the ID reasoning dataset (MATH500). *Naive Parameter Averaging* exhibits the largest and most chaotic dispersion (high shift). This confirms our hypothesis that static parameter 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 2(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.

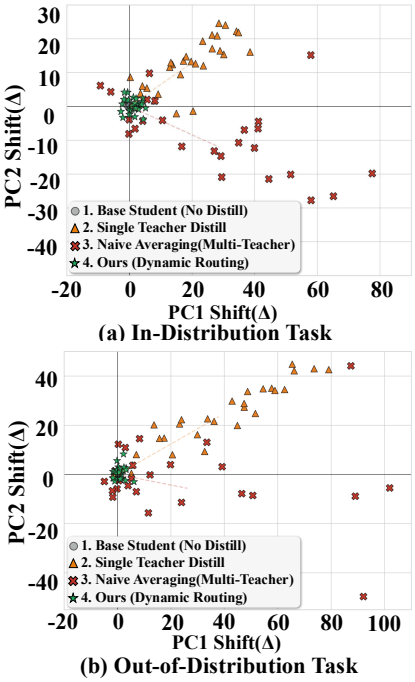


Table 2: Layer-wise PCA shift of latent representation in Qwen2.5-1.5B distilled by various methods. *COMPACT* exhibits minimal shift, confirming its stability in preserving the general capabilities.

Breaking Capability Boundaries through Adaptive Fusion. Figure 2(a) confirms that different teachers excel in specific domains and the "style signature" phenomenon, where students distilled from individual teachers inherit distinct capability biases. While the naive parameter averaging attempts to combine these abilities and shows gains in *Logical Correctness* and *Completeness*, proving that multi-teacher distillation indeed introduces diverse knowledge, 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, the *COMPACT*-distilled student effectively breaks these capability boundaries, even surpassing individual teachers on several metrics.

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 individual 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 within our multi-dimensional evaluation metric, we conducted ablation studies on the Qwen2.5-1.5B student model. The results are summarized in Figure 4. We calculated the average relative degradation rate (%) and the shift in stability (Standard Error of the Mean, SEM) across all tasks.

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."* exhibits significantly higher variance. This suggests that without compatibility awareness, the student is subject to destructive gradient interference, whereas *COMPACT*'s dynamic weighting stabilizes the learning trajectory by selecting compatible paths.

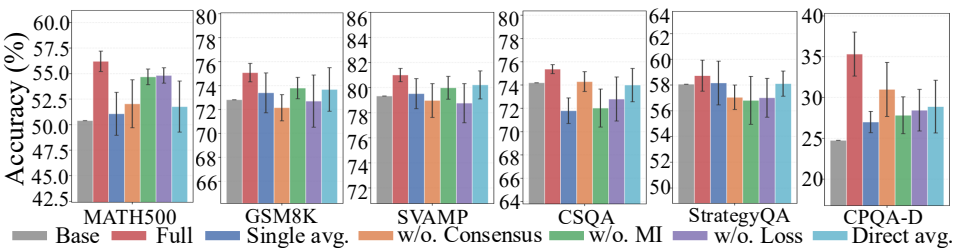


Figure 4: Comprehensive ablation study on model performance.

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 5(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.

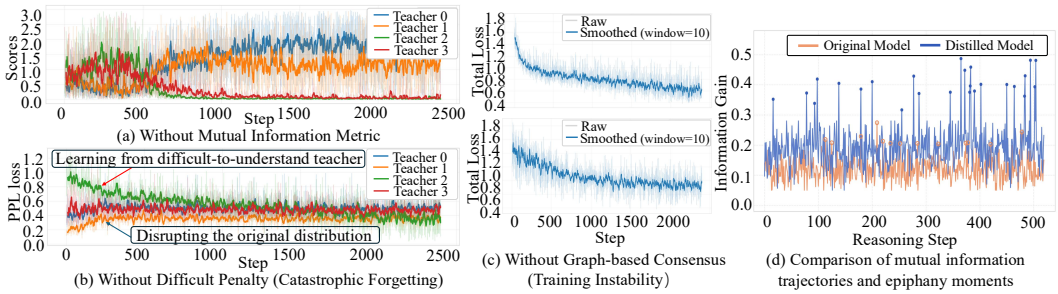


Figure 5: Training dynamics under component ablation.

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 5(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). This confirms that S_{MI} is the decisive factor in prioritizing reasoning depth over mere imitation and the key to the student's insightfulness and generalization capabilities. 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. Similar to consensus, its absence triggers higher volatility: ID variance increases by $1.95\times$ and OOD variance by $1.39\times$. This exclusion also exacerbates catastrophic forgetting (Figure 5(b)); the student aggressively 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 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 teacher strengths, 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.

6 AI USE STATEMENT

We used LLMs for two purposes in the preparation of this work. First, we used LLMs to assist with and polish the writing of the manuscript, including improving clarity, grammar, and readability. Second, we used LLMs for retrieval and discovery, such as identifying potentially relevant prior and related work. All AI-assisted outputs were reviewed and verified. We take full responsibility for the final content of this work, including all text, claims, citations, and other materials produced with the assistance of generative AI.

7 REPRODUCIBILITY STATEMENT

To support reproducibility, we submitted the code for our proposed method as part of the supplementary material. We describe the algorithm, experimental setup, datasets, evaluation protocols, and key implementation details in the main paper and appendix.

REFERENCES

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.

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. URL https://api.semanticscholar.org/CorpusID:268253497.

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*, abs/2502.18001, 2025. URL https://api.semanticscholar.org/CorpusID:276580366.

Karl Cobbe, Vineet Kosaraju, Mohammad Bavarian, Mark Chen, Heewoo Jun, Lukasz Kaiser, Matthias Plappert, Jerry Tworek, Jacob Hilton, Reiichiro Nakano, Christopher Hesse, and John Schulman. Training verifiers to solve math word problems, 2021. URL https://arxiv.org/abs/2110.14168.

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. URL https://api.semanticscholar.org/CorpusID:280018114.

Mor Geva, Daniel Khashabi, Elad Segal, and et al. Did aristotle use a laptop? a question answering benchmark with implicit reasoning strategies. In *Proceedings of EMNLP*, 2021.

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*, abs/2305.15717, 2023. URL https://api.semanticscholar.org/CorpusID:258887629.

Dan Hendrycks, Collin Wang, Karina Maziarz, Dawn Song, and ... Measuring mathematical problem solving with the math dataset. In *Advances in Neural Information Processing Systems*, 2021.

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.

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*, pp. 8003–8017, 2023.

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.

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.

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. URL https://api.semanticscholar.org/CorpusID:259075564.

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. volume 114, pp. 3521–3526. National Academy of Sciences, 2017.

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, pp. 22199–22213, 2022.

Zhuochun Li, Yuelyu Ji, Rui Meng, and Daqing He. Learning from committee: Reasoning distillation from a mixture of teachers with peer-review. *ArXiv*, abs/2410.03663, 2024. URL https://api.semanticscholar.org/CorpusID:273162520.

Zhuochun Li, Yuelyu Ji, Rui Meng, and Daqing He. Learning from committee: Reasoning distillation from a mixture of teachers with peer-review, 2025. URL https://arxiv.org/abs/2410.03663.

Hunter Lightman, Vineet Kosaraju, Yuri Burda, Harrison Edwards, Bowen Baker, Teddy Lee, Jan Leike, John Schulman, Ilya Sutskever, and Karl Cobbe. Let's verify step by step. In *The Twelfth International Conference on Learning Representations*, 2023.

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. URL https://api.semanticscholar.org/CorpusID:270067747.

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)*, pp. 1773–1781, 2023.

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

Arkil Patel, Satwik Bhattamishra, and Navin Goyal. Are nlp models really able to solve simple math word problems?, 2021. URL https://arxiv.org/abs/2103.07191.

Sohan Patnaik, Milan Aggarwal, Sumita Bhatia, and Balaji Krishnamurthy. Learning together to perform better: Teaching small-scale llms to collaborate via preferential rationale tuning. *ArXiv*, abs/2506.02519, 2025. URL https://api.semanticscholar.org/CorpusID:279119870.

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*, abs/2506.02867, 2025. URL https://api.semanticscholar.org/CorpusID:279119212.

Peter Rein, Fabian Balsiger, and et al. Gpqa: A graduate-level google-proof q&a benchmark. *arXiv preprint arXiv:2311.12022*, 2023.

Zhanming Shen, Zeyu Qin, Zenan Huang, Haoxing Chen, Jiaqi Hu, Yihong Zhuang, Guoshan Lu, Gang Chen, and Junbo Zhao. Merge-of-thought distillation. *ArXiv*, abs/2509.08814, 2025. URL https://api.semanticscholar.org/CorpusID:281243827.

Alon Talmor, Jonathan Herzig, Nicholas Lourie, and Jonathan Berant. Commonsenseqa: A question answering challenge targeting commonsense knowledge. In *Proceedings of NAACL-HLT*, 2019.

Junlin Wang, Jue Wang, Ben Athiwaratkun, Ce Zhang, and James Zou. Mixture-of-agents enhances large language model capabilities. *ArXiv*, abs/2406.04692, 2024. URL https://api.semanticscholar.org/CorpusID:270357878.

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.

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.

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*, abs/2508.09883, 2025. URL https://api.semanticscholar.org/CorpusID:280642027.

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*, abs/2505.16831, 2025. URL https://api.semanticscholar.org/CorpusID:278788524.

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. URL https://api.semanticscholar.org/CorpusID:271865581.

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*, abs/2307.10928, 2023. URL https://api.semanticscholar.org/CorpusID:259991144.

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*, abs/2408.03092, 2024. URL https://api.semanticscholar.org/CorpusID:271720178.

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. URL https://api.semanticscholar.org/CorpusID:278602083.

A Appendix

A.1 Details of Tasks and Datasets

We select MATH500, GSM8K, SVAMP, CommonsenseQA, StrategyQA, and GPQA-Diamond to systematically evaluate model performance across two dimensions: mathematical reasoning and commonsense/knowledge reasoning. The basic statistics of these benchmarks are presented in Tables 3–6.

MATH500. MATH500 comprises 500 problems randomly sampled from the MATH dataset (Hendrycks et al., 2021), covering seven subjects across five difficulty levels. Sample distribution in Table 3 and Table 4.

GSM8K. GSM8K (Cobbe et al., 2021) consists of approximately 8.5K high-quality, linguistically diverse grade school math word problems.

SVAMP. SVAMP (Patel et al., 2021) includes 1,000 one-unknown arithmetic word problems (up to grade 4), constructed by applying structural variations to problem statements. Sample distribution in Table 5.

Subject Area	Size	Proportion
Algebra	124	24.8%
Counting & Probability	38	7.6%
Geometry	41	8.2%
Intermediate Algebra	97	19.4%
Number Theory	62	12.4%
Prealgebra	82	16.4%
Precalculus	56	11.2%
Total	**500**	**100.0%**

Table 3: Subject-area distribution of MATH500.

Difficulty Level	Size	Proportion
Level 1	43	8.6%
Level 2	90	18.0%
Level 3	105	21.0%
Level 4	128	25.6%
Level 5	134	26.8%
Total	**500**	**100.0%**

Table 4: Difficulty-level distribution of MATH500.

CommonsenseQA. CommonsenseQA (Talmor et al., 2019) is a multiple-choice question-answering benchmark containing 12,247 examples to test commonsense knowledge.

StrategyQA. StrategyQA (Geva et al., 2021) comprises 2,780 questions requiring multi-step strategy inference to answer questions with implicit reasoning steps, making it a challenging benchmark for compositional and knowledge-aware reasoning.

GPQA-Diamond. GPQA-Diamond (Rein et al., 2023) contains 198 expert-written questions in biology, physics, and chemistry, selected for high discrimination between experts and non-experts. Sample distribution in Table 6.

Operation-type	Size	Proportion
Subtraction	160	53.33%
Addition	59	19.67%
Common-Division	48	16.00%
Multiplication	33	11.00%
Total	**300**	**100.00%**

Table 5: Operation-type distribution of the SVAMP test set.

Domain	Size	Proportion
Chemistry	93	46.97%
Physics	86	43.43%
Biology	19	9.60%
Total	**198**	**100.00%**

Table 6: Domain distribution of the GPQA-Diamond dataset.

A.2 Multi-perspective Dataset

We provide a detailed statistical overview of the multi-teacher dataset here. We prompted DeepSeek-R1-Llama-70B, Qwen2.5-70B, Llama-3.1-70B and QWQ-32B to explicitly generate rationales and curated 497 samples from the MATH dataset, each of which contains four valid teacher-generated rationales and a golden answer, strictly following the main text's filtering and stratification criteria. Table 7 and Table 8 summarize the dataset distribution by subjects and difficulty.

Subject Area	Size	Proportion
algebra	28	14.00%
counting and probability	5	2.50%
geometry	6	3.00%
intermediate algebra	12	6.00%
number theory	33	16.50%
prealgebra	102	51.00%
precalculus	14	7.00%
Total	**200**	**100.0%**

Table 7: Subject distribution of the constructed dataset.

Difficulty Level	Size	Proportion
Level 1	29	14.50%
Level 2	53	26.50%
Level 3	45	22.50%
Level 4	45	22.50%
Level 5	28	14.00%
Total	**200**	**100.0%**

Table 8: Difficulty-level distribution of the constructed dataset.

A.3 Experimental Environment

For brevity, the primary hyperparameter settings and distillation configurations employed in our experiments are listed in Table 9.

Parameter	Value
General Settings	
Optimizer	AdamW
LoRA Target Modules	All linear layers (Attn. & FFN)
Batch Size (per Device)	4
Gradient Accumulation	4
Hardware	2 $\times$ NVIDIA A100 (80GB)
SLMs Optimization	
Learning Rate (1.5B)	5×10^{-5}
Learning Rate (7B/8B)	1×10^{-4}
Training Epochs (1.5B)	8
Training Epochs (7B/8B)	6

Table 9: Settings and hardware implementation.

B Further Analysis

Task-Dependent Preference in Mathematical Reasoning. The training dynamics across mathematical benchmarks elucidate a clear trade-off between reasoning depth and accessibility, governed by task complexity. On the challenging MATH dataset, the student manifests a distinct *curriculum learning* trajectory: it initially prioritizes the less capable generalist teacher (Qwen2.5) to establish foundational logic, before progressively shifting alignment toward the specialized reasoning expert (QWQ) to master intricate reasoning patterns. In contrast, on the elementary GSM8K task, the student eschews the superfluous complexity of stronger reasoners (QWQ/DeepSeek-R1), consistently favoring the simpler and robust solutions of generalist models (Qwen2.5/Llama3.1). This preference confirms that the framework effectively suppresses "over-reasoning" in simpler contexts while selectively activating deep reasoning capabilities solely where necessitated by problem difficulty.

Logic Emergence in Commonsense Reasoning. On the knowledge-intensive CSQA task, the student demonstrates a non-monotonic preference adaptation that eventually absorbs complex reason-

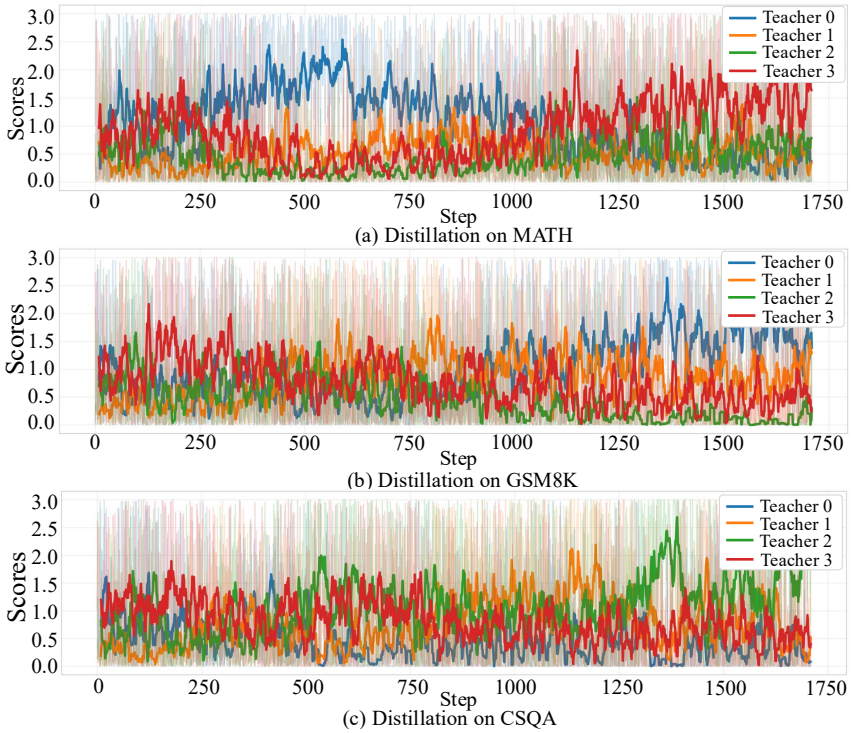


Figure 6: Evolution of dynamic weighting scores for different teachers during training on different datasets. Teacher 0-3 corresponds to Qwen2.5-72B, Llama3.1-70B, DeepSeek-R1-Llama-70B, and QWQ-32B.

ing. Unlike the stable dominance observed in math tasks, the student initially oscillates between less capable generalists (Qwen2.5) before converging on the strong reasoning expert (DeepSeek-R1) in the later training stages. This late-stage shift indicates that while generalist models can handle basic knowledge grounding, the superior structural coherence and robust reasoning ability of the strong teacher are more critical for resolving complex implicit reasoning tasks, which the student effectively internalizes only after sufficient training.

C Information Gain Visualization

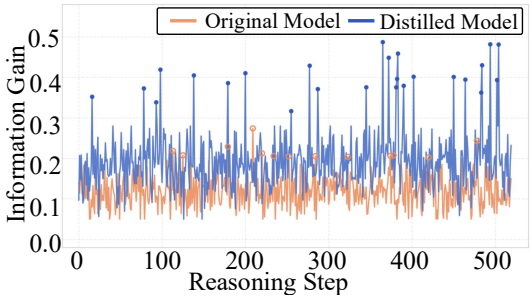


Figure 7: Comparison of mutual information trajectories between the original student model and the *COMPACT*-distilled student model during the reasoning process.

To verify whether our COMPACT framework successfully transfers the ”epiphany” capabilities of strong teachers to the student, we visualize the step-wise Information Gain (MI) on a sample reasoning task. As shown in Figure 7, the original student model exhibits a relatively flat and low-entropy trajectory, suggesting a lack of deep reasoning or logical breakthroughs. In contrast, the distilled student model demonstrates sharp, sparse peaks at specific intervals. These peaks align with the theoretical “Thinking Tokens” , confirming that the student has evolved from rote generation to active, structured reasoning with high-information logical transitions.

D DISTILLATION EFFIECIENCY

A common concern with multi-teacher distillation is the inference overhead. However, COMPACT effectively offsets this cost through extreme sample efficiency, converging to the SOTA performance using only 200 curated samples per teacher (totaling $< 1k$ samples). This drastic reduction in dataset size (compared to vanilla methods that typically mandate thousands of examples), completely neutralizes the cost of querying multiple teachers and computing metric scores, resulting in a shorter total training time.

Furthermore, the efficiency advantage is substantial when compared to advanced iterative or reinforcement learning paradigms. Methods like EDIT (Dai et al., 2025) and Committee (Li et al., 2024) require expensive multi-round generation and multi-stage student retraining, while RL approaches like (Patnaik et al., 2025) necessitate massive rollouts ($> 10k$ trajectories). In contrast, COMPACT employs a single-stage dynamic weighting mechanism, eliminating repetitive inference-train loops. This demonstrates that high-quality, compatibility-aware supervision is more computationally efficient than data scaling or complex iterative optimization.