MIND: From Passive Mimicry to Active Reasoning through Capability-Aware Multi-Perspective CoT Distillation

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

[1]State Key Laboratory of Human-Machine Hybrid Augmented Intelligence, and Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University
[2]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

While Large Language Models (LLMs) have emerged with remarkable capabilities in complex tasks through Chain-of-Thought reasoning, practical resource constraints have sparked interest in transferring these abilities to smaller models. However, achieving both domain performance and cross-domain generalization remains challenging. Existing approaches typically restrict students to following a single golden rationale and treat different reasoning paths independently. Due to distinct inductive biases and intrinsic preferences, alongside the student's evolving capacity and reasoning preferences during training, a teacher's "optimal" rationale could act as out-of-distribution noise. This misalignment leads to a degeneration of the student's latent reasoning distribution, causing suboptimal performance. To bridge this gap, we propose MIND, a capability-adaptive framework that transitions distillation from passive mimicry to active cognitive construction. We synthesize diverse teacher perspectives through a "Teaching Assistant" network. By employing a *Feedback-Driven Inertia Calibration mechanism*, this network utilizes inertia-filtered training loss to align supervision with the student's current adaptability, effectively enhancing performance while mitigating catastrophic forgetting. Extensive experiments demonstrate that MIND achieves state-of-the-art performance on both in-distribution and out-of-distribution benchmarks, and our sophisticated latent space analysis further confirms the mechanism of reasoning ability internalization.

1 Introduction

Large Language Models (LLMs) have exhibited remarkable emergent capabilities in solving complex reasoning tasks (Wei et al., 2022a; Bubeck et al., 2023). As an emergent ability, Chain-of-Thought (CoT) reasoning empowers models with exceptional performance by decomposing intricate problems into intermediate steps (Kojima et al., 2022). However, these capabilities typically mandate massive parameter scales (Wei et al., 2022b), rendering deployment in resource-constrained environments prohibitively expensive. To bridge this gap, CoT Distillation has emerged as a promising paradigm to transfer the reasoning prowess of LLMs into compact Student Models (SLMs) using teacher rationales as supervision (Ho et al., 2022; Magister et al., 2023; Hsieh et al., 2023). Despite progress, we identify several limitations in current approaches that hinder the cultivation of robust reasoning:

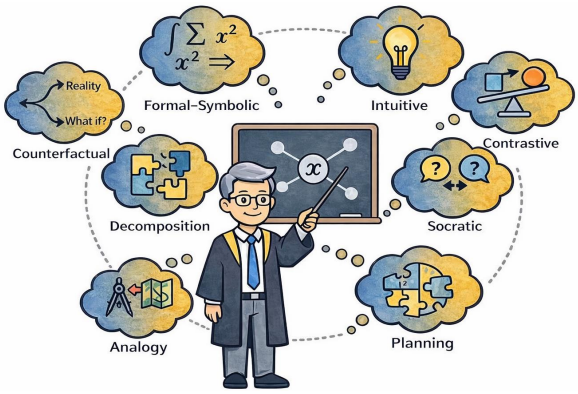

Figure 1: **Reasoning Perspectives inherent in LLMs.** We only demonstrate the distinctive ones used to clearly demonstrate our distillation method. Selectively using a subset for efficiency does not sacrifice performance.

1) *Distribution Collapse via Single-Path Rigidity.* Although teacher LLMs provide diverse reasoning trajectories (Figure 1), SLMs with limited capacity often fail to capture such complex multi-modal reasoning distributions. Rigid single-path supervision forces SLMs to average over diverse modes, causing the loss of strategic diversity to adaptively switch strategies for different questions (Ho et al., 2022; Magister et al., 2023; Chen et al., 2023) and brittle generalization.

2) *Neglect of Structural Synergy among Reasoning Paths.* While recent works leverage multiple

[*]Equal contribution.
[†]Corresponding author.

Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 43622–43636

reasoning paths to improve reliability (Chen et al., 2023; Li et al., 2024b), they typically aggregate results through voting or ranking, implicitly assuming path independence, neglecting the structural synergy where strategies interact to resolve ambiguities and complement partial information (Ainsworth, 2006). The absence of this mutual reinforcement leads to suboptimal supervision.

3) *Misalignment from Static Supervision.* Most critically, traditional "one-size-fits-all" supervision neglects the intrinsic teacher-student cognitive gap. Due to distinct inductive biases and preferences (Chen et al., 2025; Jiang et al., 2025), a teacher's "optimal" reasoning may act as out-of-distribution noise to the student that creates high variance gradients. Furthermore, we discovered that the student's learning capacity and reasoning preference evolve dynamically during training (e.g., preferring explicit step-by-step guidance over abstract leaps in early training steps) (Lin et al., 2025). This misalignment forces the student to learn patterns incompatible with their current capability, resulting in inefficient training and reasoning hallucinations.

To transition SLMs from passive mimicry to active reasoning, we propose Capability-Adaptive **M**ulti-Perspect**I**ve Chai**N**-of-Thought **D**istillation (**MIND**), a dynamic framework that harmonizes diverse reasoning patterns with student-centric adaptive supervision. Inspired by the distinct stylistic signatures of teacher LLMs, MIND constructs a *multi-perspective corpus* to capture varied cognitive reasoning strategies. Through a *dynamic fusion mechanism*, our framework enables SLMs to explicitly internalize these patterns, empowering them to flexibly synthesize appropriate strategies during inference, marking a significant leap from the monotonic reasoning of traditional methods. To bridge the capability gap, we introduce Meta-Gating Network (MetaNet), a "Teaching Assistant" that facilitates cognitive alignment through a *Feedback-Driven Inertia Calibration mechanism*. By leveraging inertia-filtered training loss as a proxy for student adaptability, MetaNet dynamically recalibrates fusion weights to direct supervision toward the most compatible paths. This capability-aligned process mitigates hallucinations and maximizes training efficiency even with minimal samples.

We pioneer framing distillation as a capability-aware cognitive construction process, departing from conventional monotonic supervision. Extensive evaluations on In-Distribution(ID) and Out-of-Distribution(OOD) benchmarks demonstrate that MIND achieves SOTA performance and mitigates catastrophic forgetting. These results suggest that our approach evolves SLMs from rote "task-takers" into genuine "thinkers" capable of universal reasoning. Code at `https://github.com/CAG-Research/MIND-CoT-Distillation.git`. Our main contributions are summarized as follows:

1. We propose MIND, a capability-aware multi-perspective framework that transitions distillation from mimicry to active cognitive construction, effectively enhancing performance.
2. We introduce a "Teaching Assistant" to align teacher supervision with the student's evolving capability, effectively mitigating hallucination and catastrophic forgetting.
3. Extensive experiments show MIND achieves SOTA performance across both ID (+3.27%) and OOD (+7.53%) benchmarks compared to previous SOTA.

2 Related Work

2.1 Chain-of-Thought Capability in LLMs

Large Language Models (LLMs) have demonstrated remarkable emergent capabilities (Wei et al., 2022a; Brown et al., 2020; Nye et al., 2021)., 2021], primarily realized by the Chain-of-Thought (CoT) reasoning paradigm (Wei et al., 2022b; Kojima et al., 2022). By decomposing complex problems into sequential intermediate steps, this strategy enables models to tackle intricate tasks that were previously intractable for standard answer-oriented inference (Wei et al., 2022a; Chowdhery et al., 2023; Imani et al., 2023). This approach has been shown to substantially enhance performance in both zero-shot and few-shot settings, eliciting deductive reasoning abilities that were previously latent (Kojima et al., 2022; Wang et al., 2023b). However, these emergent reasoning capabilities are strongly correlated with model scale; smaller language models (SLMs) often fail to spontaneously generate coherent reasoning chains or exhibit diminished performance compared to their larger counterparts (Magister et al., 2023; Fu et al., 2023), which evidences the importance of CoT distillation to explicitly endow SLMs with structured reasoning capabilities.

2.2 Distill Reasoning Capabilities into SLMs

While CoT Distillation effectively transfers reasoning capabilities from massive LLMs to compact Student Models (SLMs) (Ho et al., 2022; Magister et al., 2023; Hsieh et al., 2023), recent studies

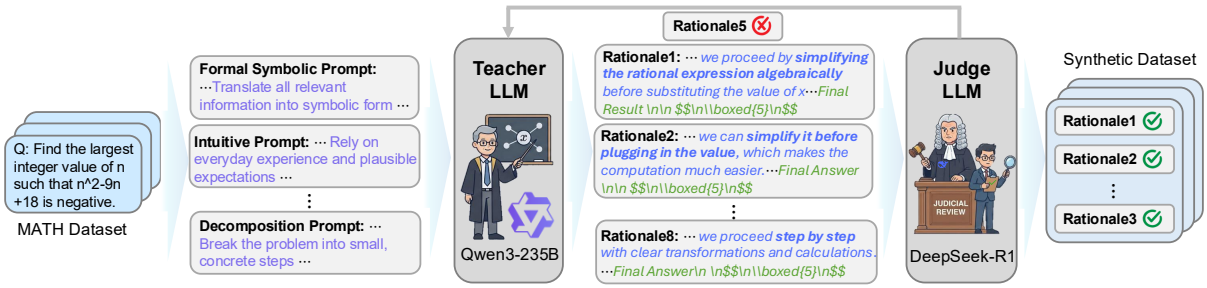

Figure 2: **Demonstration of the dataset construction pipeline.** We adopt a Judge LLM to identify reasoning with poor quality or leading to wrong answers, and maintain the consistency between rationales and the correct answer.

emphasize that the diversity of reasoning paths and the granularity of rationales often impact distillation quality more than the teacher's raw accuracy. Meanwhile, some studies highlight significant limitations in this "imitation-only" approach as it often leads to spurious correlations between questions and answers, restricting generalization to out-of-distribution (OOD) tasks (Dai et al., 2024; Feng et al., 2024; Wang et al., 2023a; Chen et al., 2023).

Crucially, existing methods struggle to harmonize diverse reasoning patterns with dynamic student adaptation. Multi-expert frameworks like (Li et al., 2024b) rely on fixed, task-level weighting, which enforces a static reasoning mode and precludes the intra-step cognitive shifting required for complex problems. Similarly, recent adaptive approaches often rigidify the learning process: (Jiang et al., 2025) strictly segregates intuition from expression, while (Li et al., 2024a) treats distinct strategies as competing options to be linearly weighted rather than complementary perspectives to be synthesized. Furthermore, these methods typically depend on external heuristics or predefined schedules, ignoring the student's internal cognitive state. Although contrastive sampling (Wang et al., 2025) attempts to mitigate this, it incurs prohibitive computational overhead. In contrast, our work is the first to explicitly synthesize a diverse reasoning repertoire synchronized with the student's real-time adaptability, achieving robust performance without auxiliary overhead.

3 Method

3.1 Dataset Construction

Cognitive Perspectives and Prompting. Empirically, we observe that teacher rationales exhibit clear semantic separability (Figure 4(c)), reflecting distinct cognitive modes rather than a chaotic distribution. Leveraging this, we synthesize eight orthogonal Cognitive Perspectives (Figure 1) to maximize representational distinctiveness and coverage. To instantiate these abstract patterns, we utilize the MATH dataset (Hendrycks et al., 2021) as our foundational corpus, chosen for its heterogeneous domains (e.g., algebra, geometry, number theory) that naturally necessitate distinct reasoning depths. Formally, for each sample $(x, y) \in \mathcal{D}_{\text{MATH}}$, we design perspective-specific prompts $\mathcal{P} = \{p_k\}_{k=1}^{8}$ to explicitly instruct the teacher (Qwen3-235B) to employ the k-th strategy as detailed in Figure 2. This yields a multi-perspective dataset where each sample consists of candidate reasoning paths and predictions: $(x, \{r_k\}_{k=1}^{8}, \{\hat{y}_k\}_{k=1}^{8})$.

Quality Filtering and Difficulty Stratification. To ensure corpus quality, we implement a rigorous post-processing pipeline. First, we discard traces where the prediction $\hat{y}_k$ deviates from the ground truth y, retaining only valid samples with correct results. Second, to avoid overfitting to trivial patterns and focus on complex reasoning, we downsample simple problems (MATH Levels 1-2) while preserving all challenging instances (Levels 3-5). The resulting dataset $\mathcal{D}_{MultiPers}$ features a balanced distribution across difficulties and perspectives.

3.2 Latent Space Visualization

To rigorously quantify whether the acquired reasoning patterns represent fundamentally different internal representations or merely superficial lexical variations, we visualize the student's latent reasoning manifold (Figure 4(a)(b)).

We train an encoder to project the last-layer hidden states from eight "specialist" students (each distilled exclusively on a specific perspective) into a low-dimensional reasoning manifold z. To rigorously organize and interpret these representations, we incorporated a Dirichlet Process Mixture Model (DPMM) to govern the structure of the latent space. Unlike parametric models with fixed cluster constraints, the non-parametric DPMM spontaneously infers the underlying structure, allowing the ac-

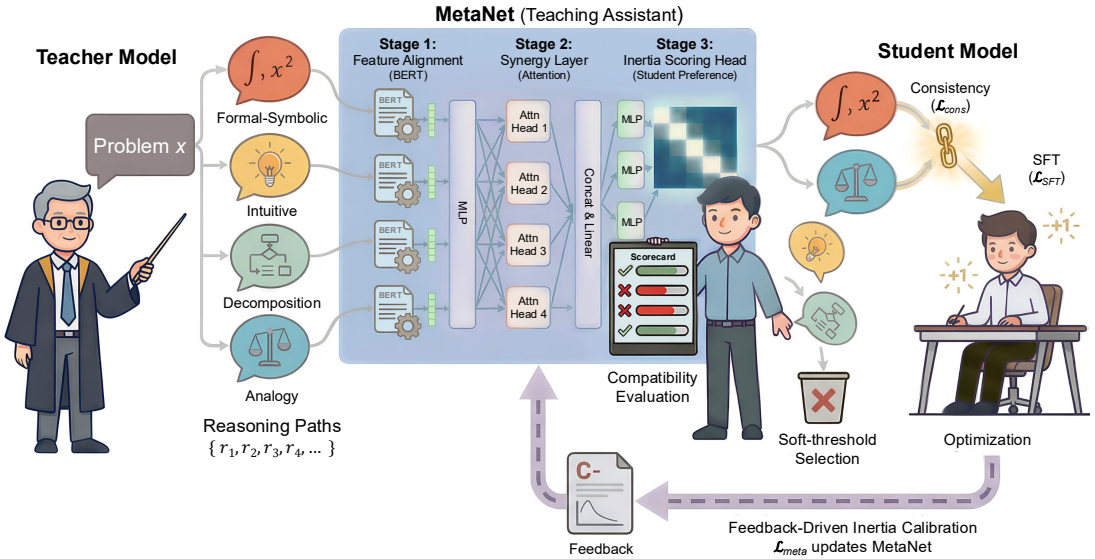

Figure 3: **Overview of our method MIND.** (1) We first prompt the teacher model to generate multi-perspective rationales. (2) Then, we warm up the MetaNet for a few steps on the acquired dataset and recalibrate it using the student's performance feedback. (3) We train the student model with SFT supervision and consistency regularization.

tive components to be determined solely by the observed data distribution. We then visualize the resulting topology of this latent space by t-SNE.

As shown in Figure 4(d), the visualization reveals a highly structured manifold where the eight specialists form distinct, compact clusters with clear decision boundaries. This topological separation confirms that the distillation process has successfully imprinted stable, distinguishable cognitive signatures onto the student's parameter space. It demonstrates that distinct reasoning perspectives correspond to specific activation patterns, indicating the internalization of underlying mechanisms rather than surface-level template memorization.

3.3 Capability-Aware Perspective Fusion

Having established that distinct reasoning perspectives form separable topological clusters within the student's latent space (Section 3.2), the subsequent challenge is to dynamically synthesize these isolated capabilities.

3.3.1 Meta-Gating Network Architecture

We introduce the Meta-Gating Network (MetaNet) acting as a dynamic "Teaching Assistant" to evaluate the compatibility of each teacher-generated reasoning path with the student's current learning state. Formally, given an input question x and a set of K candidate reasoning paths $\mathcal{R} = \{r_k\}_{k=1}^{K}$, MetaNet predicts compatibility scores $\{s_k\}_{k=1}^{K}$ via three key components (Figure 3):

1. Feature Alignment: A frozen Sentence-BERT first encodes the question and reasoning paths into

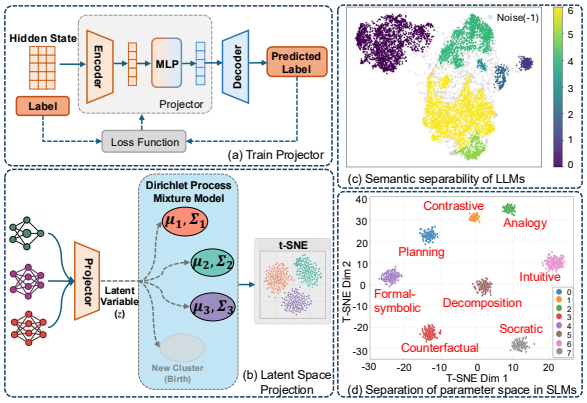

Figure 4: **Methodology and results of the Latent Space Visualization Analysis.** (a) and (b)demonstrate the mechanism, (c) illustrates the semantic level separation of LLM's reasoning paths, (d) provides a rigorous visualization of the differentiation of specifically trained students' latent space.

semantic embeddings h_x and h_{r_k}. A learnable projection layer then fuses these features into a unified latent space: $e_k = \text{MLP}_{align}([h_x; h_{r_k}])$.

2. Perspective Synergy: To capture the structural dependencies among distinct strategies, we employ a Multi-Head Self-Attention layer. The layer aggregates information across perspectives, enhancing the representation of mutually reinforcing strategies: $Z = \text{SelfAttn}(E)$, where $E = [e_1, \ldots, e_K]$.

3. Adaptive Scoring: We utilize K independent MLP heads, parameterized by ϕ_k, to function as a parametric "Inertia Matrix." By filtering short-term training variance, these heads effectively encode the student's accumulating preferences for specific reasoning styles directly into their parameters. The

output, $s_k = \text{MLP}_{score}^{(\phi_k)}(z_k)$, provides a stabilized estimation of the student's dynamic adaptability.

3.3.2 Feedback-Driven Inertia Calibration

Ideally, supervision weights should prioritize patterns aligning with the student's evolving capability. A naive approach might directly utilize the student's real-time training loss to weight different paths. However, instantaneous loss is noisy and fails to reflect long-term cognitive evolution. To address this, we propose a *Feedback-Driven Inertia Calibration Mechanism*.

MetaNet Calibration. Instead of constantly reacting to transient loss fluctuations, we explicitly align MetaNet's predictions with the student's evolving capacity using the student's real-time training loss, $\mathcal{L}_{real} = [-\log P(r_k|x;\theta)]_{k=1}^{K}$, as a ground-truth proxy for "learnability." After warming-up for a few steps, MetaNet is updated simultaneously with the student via a ListNet-based ranking loss to minimize the divergence between predicted scores s and actual performance:

$$\mathcal{L}_{\text{meta}} = D_{\text{KL}}\big(\pi_\tau(s) \,\|\, \pi_\tau(-\mathcal{L}_{\text{real}})\big) \quad (1)$$

where $\pi_\tau = Softmax(\frac{s}{\tau})$ denotes the temperature-scaled score. To ensure stability, we implement a differential update schedule where MetaNet employs a lower learning rate and higher gradient accumulation steps than the student. This induces a necessary optimization hysteresis, allowing the MetaNet to function as a stable anchor that filters stochastic noise while preserving the student's parameter plasticity during early exploration and preventing coupled oscillation.

Consistency-Regularized Supervision. Based on the compatibility scores s predicted by MetaNet for all K perspectives. To focus the student's limited capacity on high-value paths and filter outlier noise, we adaptively select a subset of high-confidence perspectives $\mathcal{I}_{dyn}$ that fall within a tolerance parameter β of the optimal score:

$$\mathcal{I}_{dyn} = \{k \mid s_k \geq \max(\mathbf{s}) * \beta\}$$

Fusion weights are then normalized exclusively within this valid subset:

$$\alpha_k = \frac{\exp(s_k/\tau_{student})}{\sum_{j\in\mathcal{I}_{top}} \exp(s_j/\tau_{student})}, \quad \forall k \in \mathcal{I}_{top}$$

The student optimization objective $\mathcal{L}_{student}$ comprises two terms: a preference-weighted SFT loss and a pairwise consistency regularization loss.

(1) Preference-Weighted SFT ($\mathcal{L}_{SFT}$): We maximize the likelihood of the selected reasoning paths within $\mathcal{I}_{dyn}$, weighted by their compatibility α_k. This ensures the student primarily learns from the perspectives that align best with its current cognitive state:

$$\mathcal{L}_{SFT} = \sum_{k\in\mathcal{I}_{dyn}} \alpha_k \cdot \mathcal{L}_{CE}(r_k|x;\theta) \quad (2)$$

(2) Pairwise Consistency Regularization ($\mathcal{L}_{cons}$): To prevent reasoning fragmentation, we enforce logical consistency among the selected perspectives with a consensus that valid reasoning paths, despite their diverse trajectories, should converge to consistent answer distributions. Unlike prior methods that regularize all paths indiscriminately, we selectively minimize the Jensen-Shannon Divergence (JSD) between the answer probability distributions $P(y|x,r_i)$ and $P(y|x,r_j)$ for every pair of selected perspectives:

$$\mathcal{L}_{cons} = \sum_{\substack{i,j\in\mathcal{I}_{\text{dyn}} \\ i<j}} (\alpha_i \cdot \alpha_j) \cdot \text{JSD}(P_i||P_j) \quad (3)$$

3.4 Student Training

The final student objective is a weighted sum:

$$\mathcal{L}_{total} = \mathcal{L}_{SFT} + \lambda\mathcal{L}_{cons} \quad (4)$$

By filtering out incompatible paths and reinforcing consistency among the compatible ones, this mechanism ensures the student internalizes a coherent and diverse repertoire of reasoning strategies.

4 Experiments

4.1 Experimental Setup

Datasets. We evaluate MIND on two categories of benchmarks to verify task performance and generalization: (1) In-Distribution mathematical problem-solving benchmarks: MATH500 (Lightman et al., 2023), GSM8K (Cobbe et al., 2021), SVAMP (Patel et al., 2021). (2) Out-Of-Distribution datasets commonsense reasoning 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 Qwen3-235B as our teacher model, selected for its state-of-the-art performance. For students, we select widely-used Qwen2.5-1.5B, Qwen2.5-7B, and Llama3.1-8B to rigorously evaluate scalability and architectural universality. Implementation settings are detailed in Table 8.

Baselines. We compare MIND against a diverse set of baselines: (1) Zero-shot CoT on 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 standard vanilla CoT distillation; (3) MCC-KD (Chen et al., 2023), a multi-path distillation approach enforcing consistency; (4) MoDE-CoTD (Li et al., 2024b), a distillation method with mixture of decoupled experts. (5) EDIT (Dai et al., 2025), which emphasizes mistake-driven key reasoning steps. Additionally, we evaluate Single-Perspective Variants (students trained on individual perspectives separately) to validate the effectiveness of our dynamic fusion mechanism.

4.2 Main Results

Table1 summarizes the performance of MIND across diverse student models and benchmarks. MIND consistently outperforms the base model and surpasses all baselines by significant margins. Experiments with Llama3.1-8B and Qwen2.5-1.5B further demonstrate MIND's robustness across different model architectures and scales. We analyze these along two key dimensions: in-distribution performance and out-of-distribution generalization.

MIND Enhancesd In-Distribution Performance. On in-distribution tasks, MIND achieves substantial gains by enabling students to adaptively internalize diverse reasoning strategies. Notably, while traditional distillation often causes parameter-constrained models (e.g., Qwen2.5-1.5B) to overfit—showing regression on tasks with slight distributional shifts—MIND maintains robust improvements comparable to the larger Qwen2.5-7B (avg. gain +3.63 / 5.10%). This suggests that MIND facilitates the acquisition of generalized reasoning capabilities rather than the rote memorization of teacher traces.

MIND Mitigates Catastrophic Forgetting and Boosts Generalization. Unlike baseline methods that suffer from overfitting and catastrophic forgetting on knowledge-intensive OOD benchmarks (e.g., CSQA, StrategyQA), MIND achieves an average OOD accuracy improvement of 3.15 (6.41%). Even compared to EDIT, which explicitly optimized for generalization, MIND demonstrates superior robustness in complex cross-task transfer and adaptability to small models. MIND addresses this by allowing the student to selectively assimilate only those reasoning patterns compatible with its current state, thereby minimizing interference with its pre-existing knowledge space. Remarkably,

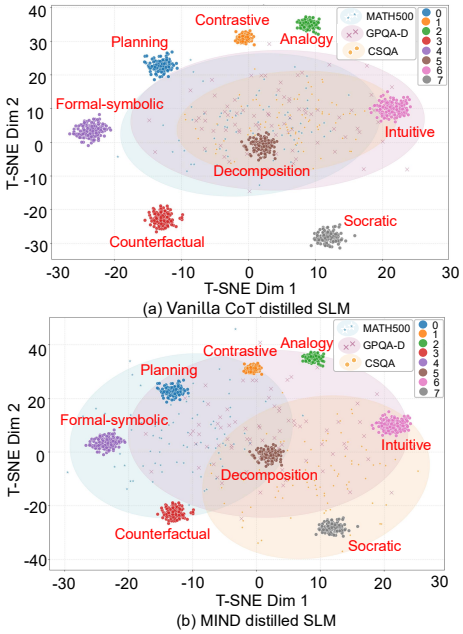

Figure 5: **Topological comparison of the student's latent reasoning manifold.** (a) The Vanilla CoT distilled student exhibits severe mode collapse. (b) The MIND-distilled student demonstrates comprehensive coverage.

on the challenging GPQA-Diamond dataset, MIND delivers a gain of 6.8 (26.57%). We attribute this to our perspective fusion mechanism that enables students to learn diverse cognitive styles rather than mimicking a monolithic reasoning process.

Necessity of Multi-Perspective Fusion. Single-perspective variants ("Ours w/o fusion") significantly underperform MIND and even lag behind baselines on OOD tasks. This degradation suggests that enforcing a rigid, monolithic reasoning mode disrupts cognitive stability, particularly in capacity-constrained models. Thus, dynamic fusion is critical for establishing a robust reasoning manifold that generalizes across domains.

Data Efficiency. MIND achieves these superior results using only 497 training samples, a magnitude fewer than the thousands required by standard distillation baselines. This extreme data efficiency effectively offsets the computational overhead of dynamic preference calculation, rendering the overall training process highly efficient.

4.3 Latent Space Distribution Analysis

To verify the internalization of diverse reasoning paradigms, we visualize the student's intrinsic reasoning manifold using the DPMM-structured encoder derived in Section 4.3. Cognitive inclination is quantified via the Euclidean distance between projected representations and pre-identified cluster centroids in Figure 4(d).

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	↑3.99	83.45	68.68	30.30	↑4.46
SbS (Hsieh et al., 2023)	77.40↑0.20	94.77↑2.41	93.00↑2.67	↑2.23	83.20↓0.25	67.25↓1.43	27.46↓2.84	↑5.97
MCC (Chen et al., 2023)	82.20↑5.00	90.52↓1.84	91.00↑0.67	↑2.71	81.72↓1.73	67.03↓1.65	26.77↓3.53	↑6.76
MoDE (Li et al., 2024b)	77.67↑0.47	94.16↑1.80	93.33↑3.00	↑2.23	83.70↑0.25	67.03↓1.65	24.75↓5.55	↑6.78
EDIT (Dai et al., 2025)	79.50↑2.30	94.28↑2.49	93.50↑3.17	↑1.53	83.80↑0.35	67.50↓1.18	29.10↓1.20	↑5.13
Ours w/o fusion Avg.	80.60±0.40	92.00±1.33	92.33±0.67	↑2.31	81.05±0.96	66.67±0.42	20.20±1.01	↑9.30
Ours w/ fusion	**82.63**↑5.43	**94.92**↑2.56	**94.31**↑3.98	–	**83.98**↑0.52	**70.74**↑2.06	**41.10**↑10.80	–
Llama3.1-8B								
Base (Llama3.1-8B-Instruct)	46.00	83.89	87.00	↑4.11	74.77	70.74	21.71	↑2.25
SbS (Hsieh et al., 2023)	52.60↑6.60	86.96↑3.07	87.67↑0.67	↑0.67	75.02↑0.25	71.05↑0.31	19.70↓2.01	↑4.23
MCC (Chen et al., 2023)	53.00↑7.00	85.01↑1.12	83.33↓3.67	↑2.63	73.33↓1.44	67.99↓2.75	18.53↓3.18	↑3.26
MoDE (Li et al., 2024b)	50.41↑4.41	84.00↑0.11	82.00↓5.00	↑4.28	73.55↓1.22	68.50↓2.24	20.71↓1.00	↑3.24
EDIT (Dai et al., 2025)	52.80↑6.80	86.91↑3.02	87.67↑0.67	↑0.62	74.82↑0.05	71.15↑0.41	20.20↓1.51	↑2.12
Ours w/o fusion Avg.	51.60±0.80	83.96±1.25	82.33±0.67	↑3.78	72.02±1.31	68.21±0.59	17.55±1.52	↑4.92
Ours w/ fusion	**53.73**↑7.73	**87.51**↑3.62	**88.00**↑1.00	–	**75.08**↑0.31	**72.71**↑1.97	**24.75**↑3.04	–
Qwen2.5-1.5B								
Base (Qwen2.5-1.5B-Instruct)	50.40	72.81	79.33	↑3.63	74.20	58.07	24.75	↑2.75
SbS (Hsieh et al., 2023)	49.80↓0.60	73.91↑1.10	78.52↓0.81	↑3.73	73.46↓0.74	56.55↓1.52	22.99↓1.76	↑4.10
MCC (Chen et al., 2023)	53.60↑3.20	70.54↓2.27	79.80↑0.47	↑3.16	71.32↓2.88	56.99↓1.08	20.71↓4.04	↑5.43
MoDE (Li et al., 2024b)	52.60↑2.20	70.31↓2.50	79.33↑0.00	↑3.72	75.14↑0.94	55.46↓2.61	21.71↓3.04	↑4.33
EDIT (Dai et al., 2025)	51.20↑0.80	74.50↑1.69	79.60↑0.27	↑2.71	74.47↑0.27	57.22↓0.85	23.74↓1.01	↑3.27
Ours w/o fusion Avg.	52.00±0.80	70.52±0.36	77.66±1.07	↑4.41	71.57±0.62	53.49±2.11	16.67±3.03	↑7.83
Ours w/ fusion	**54.09**↑3.69	**76.50**↑3.69	**82.83**↑3.50	–	**74.77**↑0.57	**59.17**↑1.10	**31.31**↑6.56	–

Table 1: Performance comparison of MIND and other methods. All baselines are trained on their original settings.

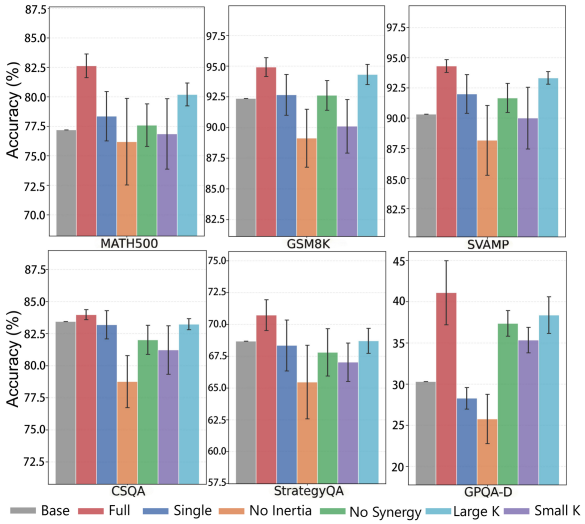

Figure 6: Comprehensive ablation study on model performance. We use top-K to replace the soft-threshold to better control the number of perspectives for clarity.

Figure 5 reveals a stark contrast between distillation paradigms. The vanilla baseline (Figure 5(a)) exhibits a decision space concentrated on narrow reasoning modes, reflecting an overfit to rigid templates and leading to brittle OOD generalization. Conversely, MIND-distilled student (Figure 5(b)) covers the entire reasoning manifold with disentangled primitives. This broad coverage allows the model to navigate its cognitive topology, activate the most appropriate reasoning mode, or a synergistic combination tailored to specific problems.

Furthermore, the specific task distributions in Figure 5 validate the efficacy of the Synergy Layer. We observe that the model does not randomly sample perspectives but exhibits task-adaptive activation: logic-intensive tasks (e.g., MATH500) gravitate towards symbolic reasoning clusters, while semantic-intensive tasks (e.g., CSQA) shift towards intuitive ones. Notably, for the highly complex GPQA-Diamond dataset, which demands heterogeneous capabilities, the student's representations are distributed across multiple synergistic clusters, demonstrating the model's capacity to compose sophisticated strategies from basic primitives. These results empirically confirm that MIND equips students with a versatile cognitive landscape, directly driving superior performance in all scenarios.

4.4 Ablation Study

We analyze the impact of the *Inertia Calibration Mechanism* and the *Perspective Synergy Layer* on the student model's training stability. Comprehensive results are illustrated in Figure 6.

Impact of Inertia Calibration Mechanism. Removing the parametric inertia matrix exposes the system to instantaneous noise($\alpha_k \propto \exp(-\mathcal{L}_{real}^{(k)})$), degenerates the system into a reactive weighting scheme susceptible to high-variance aleatoric uncertainty, leading to four failure modes (Figure 7(a)): *1) Greedy Rebound:* The model initially exploits easy samples for rapid loss reduction (greedy optimization) but suffers a sharp rebound

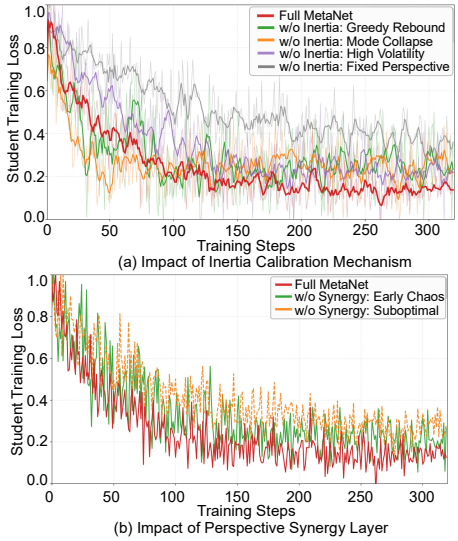

Figure 7: Effect of the Inertia Calibration Mechanism and Synergy Layer. Moving average was performed on the training loss in (a) to enhance the overall trend.

when these shortcuts fail on complex instances, leading to unstable convergence. *2) Mode Collapse:* Premature plateauing at a sub-optimal level as the model overfits to a single "shortcut" perspective that offers low initial loss, discarding valuable reasoning paths. *3) High Volatility:* Violent weight fluctuations between mini-batches, resulting in a jagged loss trajectory that hinders optimization. *4) Fixed Perspective:* Converges to a single perspective with a smooth but slow descent on loss curve, failing to leverage curriculum-based efficiency.

Impact of Perspective Synergy Layer. The Synergy Layer aggregates cross-perspective information. Ablating this layer forces the scoring heads to evaluate paths in isolation, ignoring structural dependencies and leading to two significant degradations shown in Figure 7(b): *1) Early Chaos:* Without synergy, the model struggles to distinguish between conflicting paths, leading to high-variance noise in the early stages of training. The lack of a "soft voting" mechanism delays the formation of a stable consensus on optimal reasoning modes. *2) Suboptimal Convergence:* The inability to leverage mutual reinforcement hinders the minimization of the overall training loss, resulting in a slower convergence rate and a higher final loss plateau.

5 Analysis

To determine whether the student's reasoning differentiation arises from intrinsic cognitive evolution or merely as an artifact of MetaNet's enforcement, we leverage MetaNet as a dynamic "cognitive probe" to monitor the training dynamics.

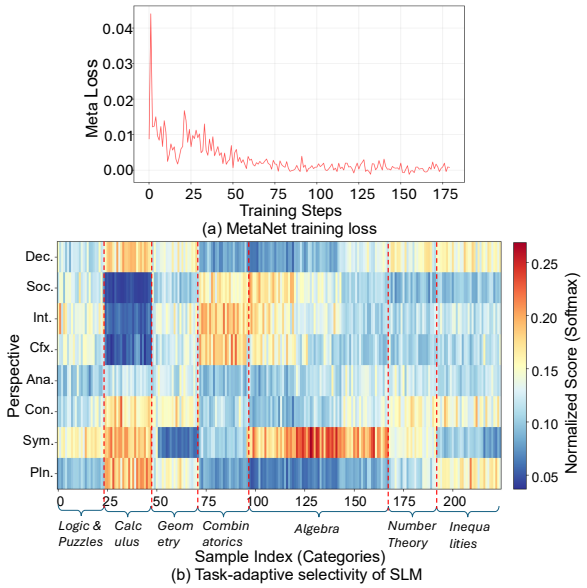

Figure 8: Evaluation to answer the question: Does the student exhibit an intrinsic, evolving preference for specific cognitive styles, or is this differentiation merely an artifact of the MetaNet's enforcement?

Synchronized Training Dynamics. Figure 8(a) illustrates the co-evolution of the student and MetaNet. The initial volatility in MetaNet's loss reflects the student's rapid cognitive restructuring during early distillation, where MetaNet acts as a lagging indicator recalibrating to shifting representations. The subsequent stabilization confirms that the student has settled into a stable reasoning manifold, achieving high-fidelity alignment with MetaNet's weighting.

Task-Specific Cognitive Preferences. Figure 7(b) validates that the student's preferences are semantically grounded. The model demonstrates task-adaptive selectivity (e.g., prioritizing spatial heuristics for geometry versus symbolic derivation for algebra). This confirms that MIND empowers the student to autonomously navigate its cognitive repertoire rather than mimicking static templates, providing a mechanistic explanation for the robust generalization observed in our experiments.

6 Conclusion

In this work, we proposed MIND, which transforms distillation from passive mimicry into active cognitive construction by dynamically synthesizing diverse reasoning perspectives aligned with the student's evolving capacity. MIND achieves SOTA results on ID and OOD benchmarks. Our comprehensive experiments confirm that SLMs can transcend imitation to become compact models equipped with robust, universal reasoning capabilities.

7 Acknowledgement

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.

8 Limitations

There are two potential limitations of our work: (1) While our eight synthesized cognitive perspectives prove highly effective for complex logical and mathematical reasoning tasks, they may not fully encompass the reasoning needs of highly subjective or creative domains (e.g., literary analysis or open-ended storytelling). Extending the MIND framework to such non-deterministic tasks would likely necessitate defining a new set of domain-specific cognitive primitives.

(2) Our current evaluation is primarily conducted on widely-used open-source student models. Future work should extend to a broader spectrum of student architectures and sizes to fully establish the universality of MIND. Additionally, validating the framework with a wider array of teacher models, including proprietary closed-source LLMs, remains an important direction to further explore the boundaries of capability transfer.

References

Shaaron Ainsworth. 2006. Deft: A conceptual framework for considering learning with multiple representations. *Learning and instruction*, 16(3):183–198.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, and 1 others. 2020. Language models are few-shot learners. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, pages 1877–1901.

Sébastien Bubeck, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yuanzhi Li, Scott Lundberg, and 1 others. 2023. Sparks of artificial general intelligence: Early experiments with gpt-4. *arXiv preprint arXiv:2303.12712*.

Hongzhan Chen, Siyue Wu, Xiaojun Quan, Rui Wang, Ming Yan, and Ji Zhang. 2023. Mcc-kd: Multi-cot consistent knowledge distillation. *arXiv preprint arXiv:2310.14747*.

Xinghao Chen, Zhijing Sun, Guo Wenjin, Miaoran Zhang, Yanjun Chen, Yirong Sun, Hui Su, Yijie Pan, Dietrich Klakow, Wenjie Li, and 1 others. 2025. Unveiling the key factors for distilling chain-of-thought reasoning. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 15094–15119.

Aakanksha Chowdhery, Sharan Narang, Jacob Devlin, Maarten Bosma, Gaurav Mishra, Adam Roberts, Paul Barham, Hyung Won Chung, Charles Sutton, Sebastian Gehrmann, and 1 others. 2023. Palm: Scaling language modeling with pathways. *Journal of Machine Learning Research*, 24(240):1–113.

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. 2021. Training verifiers to solve math word problems. *Preprint*, arXiv:2110.14168.

Chengwei Dai, Kun Li, Wei Zhou, and Songlin Hu. 2024. Improve student's reasoning generalizability through cascading decomposed cots distillation. *arXiv preprint arXiv:2405.19842*.

Chengwei Dai, Kun Li, Wei Zhou, and Songlin Hu. 2025. Capture the key in reasoning to enhance CoT distillation generalization. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 441–465, Vienna, Austria. Association for Computational Linguistics.

Tao Feng, Yicheng Li, Li Chenglin, Hao Chen, Fei Yu, and Yin Zhang. 2024. Teaching small language models reasoning through counterfactual distillation. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 5831–5842.

Yao Fu, Hao Peng, Litu Ou, Ashish Sabharwal, and Tushar Khot. 2023. Specializing smaller language models towards multi-step reasoning. In *International Conference on Machine Learning*, pages 10421–10430. PMLR.

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

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

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

Yifan Hou, Jiaoda Li, Yu Fei, Alessandro Stolfo, Wangchunshu Zhou, Guangtao Zeng, Antoine Bosselut, and Mrinmaya Sachan. 2023. Towards a mechanistic interpretation of multi-step reasoning capabilities of language models. *ArXiv*, abs/2310.14491.

Cheng-Yu Hsieh, Chun-Liang Li, Chih-Kuan Yeh, Hootan Nakhost, Yasuhisa Fujii, Alex Ratner, Ranjay Krishna, Chen-Yu Lee, and Tomas Pfister. 2023. 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.

Shima Imani, Liang Du, and Harsh Shrivastava. 2023. Math-prompter: Mathematical reasoning using large language models. *arXiv preprint arXiv:2303.05398*.

Wangyi Jiang, Yaojie Lu, Hongyu Lin, Xianpei Han, and Le Sun. 2025. Teach small models to reason by curriculum distillation. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 7423–7433.

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

Chenglin Li, Qianglong Chen, Liangyue Li, Caiyu Wang, Feng Tao, Yicheng Li, Zulong Chen, and Yin Zhang. 2024a. Mixed distillation helps smaller language models reason better. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 1673–1690.

Xiang Li, Shizhu He, Jiayu Wu, Zhao Yang, Yao Xu, Yang jun Jun, Haifeng Liu, Kang Liu, and Jun Zhao. 2024b. Mode-cotd: Chain-of-thought distillation for complex reasoning tasks with mixture of decoupled lora-experts. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 11475–11485.

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

Jhe-Hao Lin, Yi Yao, Chan-Feng Hsu, Hong-Xia Xie, Hong-Han Shuai, and Wen-Huang Cheng. 2025. Perspective-aware teaching: Adapting knowledge for heterogeneous distillation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4178–4187.

Lucie Charlotte Magister, Jonathan Mallinson, Jakub Adamek, Eric Malmi, and Aliaksei Severyn. 2023. 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.

Maxwell Nye, Anders Johan Andreassen, Guy Gur-Ari, Henryk Michalewski, Jacob Austin, David Bieber, David Dohan, Aitor Lewkowycz, Maarten Bosma, David Luan, and 1 others. 2021. Show your work: Scratchpads for intermediate computation with language models. In *Deep Learning for Code Workshop at NeurIPS*.

Arkil Patel, Satwik Bhattamishra, and Navin Goyal. 2021. Are nlp models really able to solve simple math word problems? *Preprint*, arXiv:2103.07191.

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

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

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

Wei Wang, Zhaowei Li, Qi Xu, Yiqing Cai, Hang Song, Qi Qi, Ran Zhou, Zhida Huang, Tao Wang, and Li Xiao. 2025. QCRD: Quality-guided contrastive rationale distillation for large language models. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 14345–14356.

Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc Le, Ed Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. 2023b. Self-consistency improves chain of thought reasoning in language models. In *International Conference on Learning Representations (ICLR)*.

Jason Wei, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama, Maarten Bosma, Denny Zhou, Donald Metzler, and 1 others. 2022a. Emergent abilities of large language models. *arXiv preprint arXiv:2206.07682*.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, and Denny Zhou. 2022b. Chain-of-thought prompting elicits reasoning in large language models. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 35, pages 24824–24837.

XiaoYu Xu, Xiang Yue, Yang Liu, Qingqing Ye, Haibo Hu, and Minxin Du. 2025. Unlearning isn't deletion: Investigating reversibility of machine unlearning in llms. *ArXiv*, abs/2505.16831.

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 2–5.

MATH500.MATH500 comprises 500 problems randomly sampled from the MATH dataset (Hendrycks et al., 2021), covering seven subjects across five difficulty levels.

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.

CommonsenseQA. CommonsenseQA (Talmor et al., 2019) contains 12,247 examples testing commonsense knowledge.

StrategyQA. StrategyQA (Geva et al., 2021) comprises 2,780 questions requiring multi-step strategy inference to answer questions with implicit reasoning steps.

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.

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 2: 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 3: Difficulty-level distribution of MATH500.

A.2 Multi-perspective Dataset

We provide a detailed statistical overview of the multi-perspective dataset here. Using Qwen3-235B and eight perspective prompts (Appendix A.4), we curated 497 samples from the MATH dataset, strictly following the main text's filtering and stratification criteria. Each sample contains eight valid CoT rationales yielding correct answers. Tables 6 and 7 summarize the dataset distribution by subject area and difficulty.

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 4: 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 5: Domain distribution of the GPQA-Diamond dataset.

Subject Area	Size	Proportion
algebra	187	37.6%
counting_and_probability	31	6.2%
geometry	25	5.0%
intermediate_algebra	43	8.7%
number_theory	59	11.9%
prealgebra	122	24.5%
precalculus	30	6.0%
Total	**497**	**100.0%**

Table 6: Subject-area distribution of the constructed MATH-derived dataset.

Difficulty Level	Size	Proportion
Level 1	74	14.9%
Level 2	137	27.6%
Level 3	124	24.9%
Level 4	112	22.5%
Level 5	50	10.1%
Total	**497**	**100.0%**

Table 7: Difficulty-level distribution of the constructed MATH-derived dataset.

A.3 Experimental Environment

For brevity, the primary hyperparameter settings and fine-tuning configurations employed in our experiments are listed in Table 8.

A.4 Multi-perspective Prompts

To encapsulate the broad reasoning manifold and mitigate the bias of any single reasoning mode, we elicited diverse CoT rationales by simulating

Parameter	Value
General Settings	
Optimizer	AdamW
LoRA Target Modules	All linear layers (Attn. & FFN)
Batch Size (per Device)	4
Gradient Accumulation	4
Hardware	2 × 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
MetaNet Configuration	
Warmup Stage	1 Epoch @ 1×10^{-4}
Calibration Stage LR	5×10^{-5}

Table 8: Hyperparameter settings and hardware configuration for MIND distillation.

eight teacher "stylistic signatures". By employing the prompts listed in Table 9, we tasked Qwen3-235B with constructing a multi-perspective dataset. This diversity ensures that the student model is exposed to a rich spectrum of cognitive dimensions, fostering superior generalization across complex tasks.

B Quantitative analysis of reasoning paths

While the t-SNE clusters demonstrate separation, this may come from superficial lexical differences rather than semantic distinctions in reasoning logic. To rigorously prove that the distinct perspectives represent cognitive differentiation rather than lexical pattern, we conducted two additional layer-wise analyses on the student model (Qwen2.5-7B, 28 layers):

1. Following the settings in (Xu et al., 2025), we conducted a Layer-wise PCA Trajectory Analysis. We projected the internal hidden states of the eight specialist student models, distilled with specific reasoning perspectives, through PCA, and tracked their trajectory evolution from the first to the final layer.

2. We further introduce *layer-wise probing* similar to the settings in (Hou et al., 2023) to verify that linearly separable semantic representations emerge within the model's middle layers. By extracting hidden states from the eight specialist student models at the final prompt token (capturing *reasoning initiation* prior to token generation) and training simple linear classification probes at different layers,

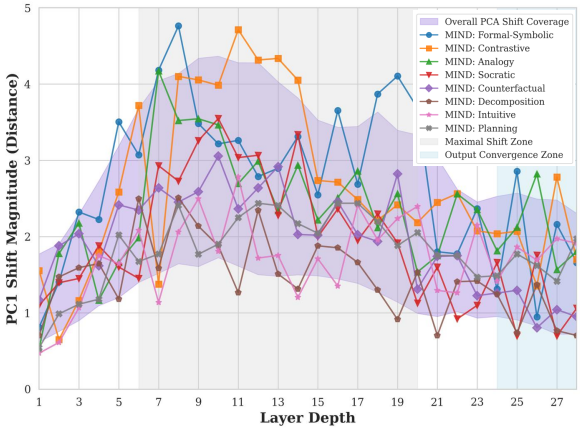

Figure 9: Layer-wise absolute PC1 shift magnitude relative to base model

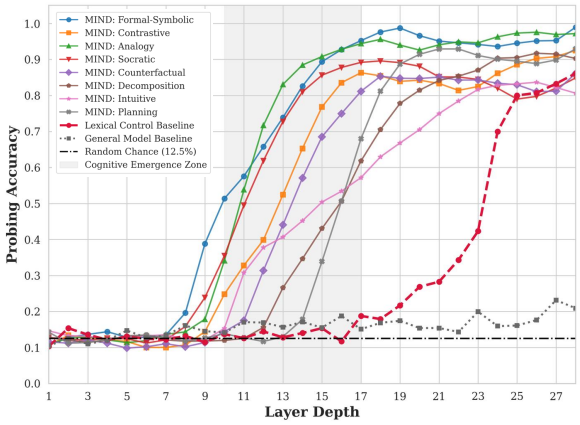

Figure 10: Layer-wise probing accuracy

we identify where the distilled student model begins to accurately predict the active reasoning perspective. We set three baselines: 1) *Random Chance* (fixed at 12.5% accuracy); 2) *Lexical Control Baseline*: finetune the student model using samples with only lexical variations but no logical differences (lexical variation as labels); and 3) *General Model Baseline*: derived from the hidden states of the Qwen2.5-7B without finetuning (perspectives as labels).

B.1 Result of Layer-wise PCA Trajectory Analysis

In LLMs, early-to-middle layers handle semantic abstraction and logical reasoning, while late layers translate internal representation to specific vocabulary. If the DPMM cluster separation observed in the last-layer hidden states were merely at the lexical level, the eight trajectories would remain highly entangled throughout the early and middle layers and only diverge at the final layers.

Our results in Figure 9 refute this hypothesis.

Perspective	Description
Formal Symbolic (Sym.)	Solve the problem using strict formal reasoning. Translate all relevant information into symbolic or logical representations and follow a sequence of explicit, verifiable deductions. Intuitive leaps are avoided in favor of equations, logic operators, or formal inference rules. The final answer is presented after the last formal step.
Intuitive (Int.)	Solve the problem using intuitive and heuristic reasoning. Concepts are explained in simple, human-like terms, relying on everyday experience and plausible expectations. Heavy formalism is avoided in favor of natural understanding. The final answer is stated after the intuitive reasoning process.
Decomposition (Dec.)	Break the problem into small, concrete sub-goals. Each step explicitly addresses a component of the problem and incrementally leads toward the solution. Reasoning steps are numbered (e.g., Step 1, Step 2, ...) without skipping intermediate logic. The final answer is provided at the end.
Planning (Pln.)	Begin with a high-level plan outlining the strategy for solving the problem. Each component of the plan is then expanded into detailed reasoning. The structure is made explicit in the order of *Plan*, *Execution*, and *Answer*. The final answer follows the completed breakdown.
Analogy (Ana.)	Solve the problem by constructing an analogy to a simpler or more familiar situation. The analogy is explained clearly, and each of its elements is mapped back to the original problem. The inferred solution from the analogy leads to the final answer.
Socratic (Soc.)	Employ a self-questioning approach to guide reasoning. At each step, guiding questions (e.g., "What is known?", "What follows?", "Why?") are posed and explicitly answered. This iterative questioning drives the reasoning process, culminating in the final answer.
Contrastive (Con.)	Solve the problem by comparing multiple candidate explanations or answers. Each option is analyzed in terms of its plausibility, strengths, and weaknesses. Through contrast and elimination, the most appropriate answer is identified and stated.
Counterfactual (Cfx.)	Apply counterfactual reasoning by considering how the outcome would change if certain conditions were altered. Alternative scenarios are analyzed to reveal critical dependencies. These insights are then used to determine the correct answer under the actual conditions. The final answer is stated accordingly.

Table 9: Prompt templates of eight reasoning perspectives used in the multi-perspective dataset construction.

The trajectories diverge in the middle layers (at 1/3 to 2/3 depth) along distinct subspace paths, before ultimately converging toward their final representations. Notably, the differences within the output space are smaller than those in the middle layers, proving that these reasoning perspectives invoke distinct feature combinations in the model's deep semantic space well before the vocabulary generation stage.

Together, these layer-wise mechanistic analyses quantitatively demonstrate that the eight cognitive perspectives in our method diverge significantly within the model's deep semantic space well before output lexical formatting.

B.2 Result of Layer-wise Probing Analysis

As shown in Figure 10, the linear separability of our eight perspectives emerges significantly at Layer 14 (approx. 40% depth), reaching 90% accuracy by Layer 19. In contrast, the *Lexical Control* baseline becomes separable after Layer 26 (approx. 90% depth). While the general model baseline confirms these differentiation are not originally appear in the undistilled model.

C Computational Overhead

We decompose the computational overhead into *Forward* and *Backward* passes for detailed analysis:
1.*Forward Pass*: As shown in Table 10, the forward computation overhead is merely 2.51×10^8 FLOPs, accounting for $< 0.001\%$ of the student model's training cost. While the initial text processing by Sentence-BERT (process sequences of length $K \times L$ for feature extraction) does introduce approximately 6.28% in computational overhead, this specific computation is entirely decoupled from the actual training loop and is more accurately categorized as a one-time, offline post-processing cost during the dataset generation process.
2.*Backward Pass*: As shown in Table 11, the backward pass exclusively operates on a lightweight network comprising only 15.74M parameters. MetaNet's backward pass requires only 5.03×10^8 FLOPs, 0.0008% of the massive $5.73 \times$

Component	Complexity / $\mathcal{O}(\cdot)$	State	Peak RAM (Weights)	FLOPs
Feature Alignment†	$\mathcal{O}(K \cdot L \cdot d_{\text{bert}}^2)$	Frozen	$\sim$220 MB (FP16)	$\approx 3.60 \times 10^{12}$
Synergy Attention†	$\mathcal{O}(K^2 \cdot d + K \cdot d^2)$	Trainable	$\sim$29 MB (FP32)	$\approx 1.17 \times 10^{8}$
Scoring MLPs	$\mathcal{O}(K \cdot d^2)$	Trainable	$\sim$34 MB (FP32)	$\approx 1.34 \times 10^{8}$
Total MetaNet	—	—	$\sim$283 MB	$\approx 3.60 \times 10^{12}$
Student LLM	$\mathcal{O}(L^2 \cdot d_{\text{llm}})$	Trainable	$\sim$28 GB – 42 GB	$\approx 5.73 \times 10^{13}$
Ratio*	—	—	$\approx$0.81%	$\approx$6.28%

Notation: † Sentence-BERT, ‡ with projection, * MetaNet / Student. L-Sequence Length, K-Number of total perspectives, d-Tensor Dimension

Table 10: MetaNet Forward Computation Breakdown

Component	Complexity / $\mathcal{O}(\cdot)$	State	Peak RAM (Grads+AdamW)	FLOPs
Feature Alignment†	$\mathcal{O}(K \cdot L \cdot d_{\text{bert}}^2)$	Frozen	0 MB (*No gradients*)	0
Synergy Attention†	$\mathcal{O}(K^2 \cdot d + K \cdot d^2)$	Trainable	$\sim$87 MB (FP32)	$\approx 2.34 \times 10^{8}$
Scoring MLPs	$\mathcal{O}(K \cdot d^2)$	Trainable	$\sim$102 MB (FP32)	$\approx 2.68 \times 10^{8}$
Total MetaNet	—	—	$\sim$189 MB	$\approx 5.02 \times 10^{8}$
Student LLM	$\mathcal{O}(L^2 \cdot d_{\text{llm}})$	Trainable	$\sim$28 GB – 42 GB	$\approx 5.73 \times 10^{13}$
Ratio*	—	—	$< 0.67\%$	$\approx 0.00087\%$

Notation: † Sentence-BERT, ‡ with projection, * MetaNet / Student. L-Sequence Length, K-Number of total perspectives, d-Tensor Dimension.

Table 11: MetaNet Backward Computation Breakdown

10^{13} FLOPs required for the 7B student model's backward pass. The total memory footprint of MetaNet's weights, gradients, and AdamW first/second-order momentum states is approximately 252MB. Even when including the weights of the frozen BERT, the peak memory footprint stays well below 500MB. Relative to the tens of gigabytes necessary for training the student LLM, this introduces no memory bottleneck.

D Dataset Scaling

To strictly verify whether the gains come from MetaNet or simply the *Naive Diversity* of 8 perspectives, we conduct two additional experiments.

D.1 Mixture of 8 perspectives

We mix the 497 samples (each with 8 perspectives) and directly train the student models to detect the path diversity bias. We compared our full MIND framework against a *Naive Diversity* baseline (Mixed), where the student model is finetuned on the combined multi-perspective dataset without MetaNet's guidance.

As shown in Table 12, simply mixing diverse reasoning paths leads to performance degradation across almost all tasks. This is particularly severe in small size student models like Qwen2.5-1.5B, which suffers massive drops. These results validate our hypothesis that different perspectives inherently contain distinct inductive biases. Without the mechanism to filter and align these rationales with the student's real-time capability, "naive diversity" effectively acts as out-of-distribution noise, causing perspective conflicts and confusing the student model.

D.2 Scale the dataset to verify the benefit of diverse reasoning paths

We further investigated whether simply scaling up the volume of the "Naive Diversity" dataset could improve the performance. We increase the base samples from 500 to 2000 (effectively scaling the reasoning paths from 4,000 to 16,000) to finetune Qwen2.5-7B.

As shown in Table 13, scaling the diverse dataset without MetaNet selection fails to yield sustained improvements. While there is a marginal improvement on in-distribution tasks' performance (MATH500, GSM8K, SVAMP) before 1000 samples, which could be attributed to the benefit of increasing training set size, scaling beyond this point triggers severe performance degeneration across all benchmarks. Out-of-distribution and knowledge-intensive tasks suffer the most. This result pro-

Model	MATH500	Mixed	GSM8K	Mixed	SVAMP	Mixed	CSQA	Mixed	StrategyQA	Mixed	GPQA-D	Mixed
Qwen2.5-7b	**82.63**	78.20	**94.92**	92.83	**94.31**	92.33	**83.98**	80.09	**70.74**	67.28	**41.10**	27.27
Llama3.1-8b	**53.73**	49.76	87.51	**87.74**	**88.00**	87.67	75.08	**75.15**	**72.71**	67.69	**24.75**	18.18
Qwen2.5-1.5b	**54.09**	43.40	**76.50**	69.29	**82.83**	78.00	**74.77**	67.01	**59.17**	46.51	**31.31**	16.67

Table 12: Direct Mixing to Detect Path Diversity Bias

Base Samples (Effective Paths)	MATH500	GSM8K	SVAMP	CSQA	StrategyQA	GPQA-D
500 (4,000)	78.20	92.83	92.33	**80.09**	**67.28**	**27.27**
800 (6,400)	**78.45**	92.90	**92.55**	79.20	64.76	25.75
1000 (8,000)	77.83	**92.95**	90.35	75.05	62.40	23.74
1400 (11,200)	74.16	90.67	89.01	72.13	60.12	15.15
1600 (12,800)	73.63	90.13	88.23	71.80	59.55	15.15
1800 (14,400)	70.92	89.42	87.00	70.21	57.73	12.79
2000 (16,000)	68.50	88.60	85.59	66.07	53.50	10.61

Table 13: Scaling on Dataset with Diverse Paths

Effective Paths	MATH500	GSM8K	SVAMP	CSQA	StrategyQA	GPQA-D
12,800	75.56	91.26	89.12	74.77	61.15	25.26
16,000	71.13	90.07	87.33	70.99	57.05	19.16

Table 14: Scaling on Standard CoT Dataset

vides clear evidence of catastrophic forgetting and mode collapse. As the volume of conflicting reasoning paths increases, the unguided student model struggles to harmonize the conflicting supervision signals.

To verify the performance drop is caused by perspective conflict rather than purely the catastrophic forgetting caused by a large domain-specific dataset, we generate standard CoT rationales by prompting the teacher model directly on MATH questions to acquire a larger dataset of 16,000 samples. We train Qwen2.5-7B with dataset scale 12,800 and 16,000 to show how the student model itself will scale to the same large amount of samples.

As shown in Table 14, Qwen2.5-7B finetuned on standard CoT with large dataset shows relatively less performance degradation, confirming our hypothesis.

These experiments jointly prove that the success of MIND stems from the adaptive supervision adjustment enabled by MetaNet, rather than the mere volume or naive diversity of the teacher's reasoning paths.