

# Two Experts Are All You Need for Steering Thinking: Reinforcing Cognitive Effort in MoE Reasoning Models Without Additional Training

Mengru Wang<sup>\*,1,2</sup>, Xingyu Chen<sup>\*,1</sup>, Yue Wang<sup>\*,1</sup>, Zhiwei He<sup>\*,1</sup>, Jiahao Xu<sup>1</sup>, Tian Liang<sup>1</sup>,  
Qiuzhi Liu<sup>1</sup>, Yunzhi Yao<sup>2</sup>, Wenxuan Wang<sup>1</sup>, Ruotian Ma<sup>1</sup>, Haitao Mi<sup>1</sup>,  
Ningyu Zhang<sup>†,2</sup>, Zhaopeng Tu<sup>†,1</sup>, Xiaolong Li<sup>1</sup>, and Dong Yu<sup>1</sup>

<sup>1</sup>Tencent

<sup>2</sup>Zhejiang University

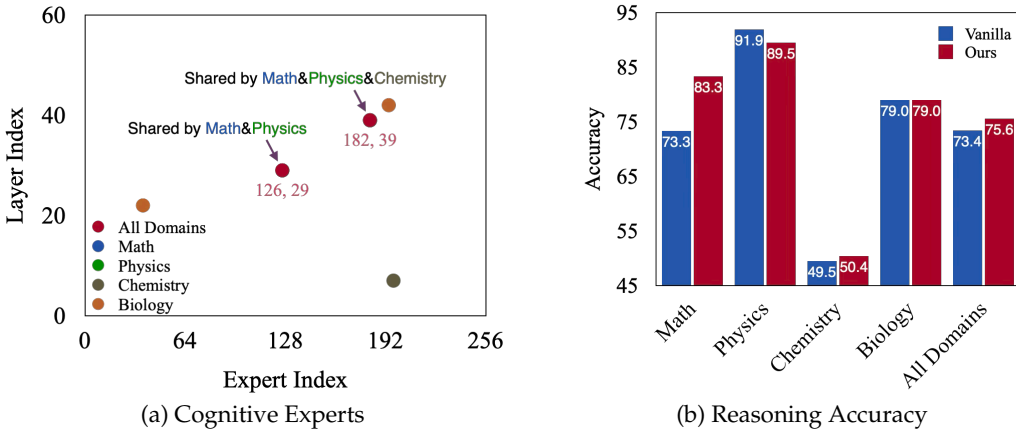


Figure 1: (a) Illustration of cognitive experts identified across domains. (b) Reinforcing only the top two experts (in red color) can improve reasoning accuracy without additional training.

## Abstract

Mixture-of-Experts (MoE) architectures within Large Reasoning Models (LRMs) have achieved impressive reasoning capabilities by selectively activating experts to facilitate structured cognitive processes (Guo et al., 2025; Team, 2025). Despite notable advances, existing reasoning models often suffer from cognitive inefficiencies like overthinking (Chen et al., 2024) and underthinking (Wang et al., 2025a). To address these limitations, we introduce a novel inference-time steering methodology called **Reinforcing Cognitive Experts (RICE)**, designed to improve reasoning performance without additional training or complex heuristics. Leveraging normalized Pointwise Mutual Information (nPMI), we systematically identify specialized experts, termed **cognitive experts** that orchestrate meta-level reasoning operations characterized by tokens like “<think>”. Empirical evaluations with leading MoE-based LRMs (DeepSeek-R1 and Qwen3-235B) on rigorous quantitative and scientific reasoning benchmarks demonstrate noticeable and consistent improvements in reasoning accuracy, cognitive efficiency, and cross-domain generalization. Crucially, our lightweight approach substantially outperforms prevalent reasoning-steering techniques, such as prompt design and decoding constraints, while preserving the model’s general instruction-following skills. These results highlight reinforcing cognitive experts as a promising, practical, and interpretable direction to enhance cognitive efficiency within advanced reasoning models.

<sup>\*</sup>Equal Contribution. Work was done when Mengru, Xingyu, Yue, and Zhiwei were interning at Tencent.

<sup>†</sup>Correspondence to: Zhaopeng Tu <zptu@tencent.com> and Ningyu Zhang <zhangningyu@zju.edu.cn>.

## 1 Introduction

Models capable of extended reasoning, often referred to as Large Reasoning Models (LRMs) like OpenAI’s o1 (Jaech et al., 2024) and DeepSeek-R1 (Guo et al., 2025), have significantly advanced machine intelligence, largely by scaling test-time computation (Ji et al., 2025; Zhang et al., 2025a). Despite their impressive capabilities, these LRMs remain susceptible to inefficiencies (Sui et al., 2025a; Feng et al., 2025; Pan et al., 2025; Qu et al., 2025; Chen et al., 2025a; Wang et al., 2025b; Wu et al., 2025; Lu et al., 2025). Prior work has sought to address these issues through approaches such as preference optimization (Chen et al., 2024), decoding penalties (Wang et al., 2025a), and various other techniques. In this work, we tackle these problems from a novel perspective: potential expert specialization in Mixture-of-Experts (MoE) architecture.

Due to the computational resource efficiency brought about by its sparsity, the MoE architecture (Dai et al., 2024a; DeepSeek-AI et al., 2024; Xue et al., 2024) has been increasingly adopted by state-of-the-art (SOTA) models, such as DeepSeek-R1 (Guo et al., 2025) and Qwen3 (Team, 2025). This sparse, specialized activation paradigm bears a conceptual resemblance to functional specialization in the human brain, where targeted interventions can modulate cognitive functions and behaviors (Reinhart and Nguyen, 2019; Wischnewski et al., 2023; Oathes et al., 2023; Grover et al., 2021). Inspired by this principle, we systematically investigate whether undesirable reasoning behaviors in MoE-based LRMs correlate with the activation patterns of specific experts, and critically, if strategic manipulation of these experts can ameliorate such issues.

We introduce an approach to identify and modulate key experts integral to the reasoning process. By analyzing the co-occurrence of explicit linguistic markers of thought (e.g., ‘<think>’ and ‘</think>’) with individual expert activations, we pinpoint a subset of experts highly correlated with the model’s cognitive deliberations. We designate these critical experts as **cognitive experts**. Through extensive experimentation with SOTA MoE-reasoning models DeepSeek-R1 (Guo et al., 2025) and Qwen3-235B (Team, 2025) on challenging mathematic and scientific reasoning benchmarks, we demonstrate that selectively amplifying **as few as two cognitive experts** can enhance both reasoning depth and efficiency. Notably, our approach achieves marked accuracy improvements while reducing token usage in critical reasoning tasks, outperforming existing steering methods such as prompting and decoding constraints (Wang et al., 2025a).

**On the Cognitive Expert.** The “cognitive expert” proposed in this work is a hypothetical construct. Given the complexity of LRMs, we provide no theoretical justification for its existence; our conclusions are purely empirical.

Moreover, we showcase impressive generalization and robustness of cognitive expert modulation, observing consistent improvements in unseen and more complex reasoning scenarios while maintaining or even enhancing general instruction-following capabilities. Our findings provide strong evidence that modulating selective experts responsible for meta-level reasoning is effective, efficient, and broadly applicable across domains, paving the way for lightweight and interpretable model steering in increasingly sophisticated MoE-based reasoning models.

Our main contributions are:

1. We propose a normalized Pointwise Mutual Information (nPMI) method for identifying cognitive experts within LRMs that are highly correlated with reasoning behavior, requiring only a single forward propagation and no additional training.
2. We introduce a lightweight inference-time steering strategy, named “reinforcing cognitive experts”, that effectively enhances reasoning depth and accuracy without requiring any additional training or supervision signals.
3. Through comprehensive experiments on two prevalent MoE reasoning models and rigorous benchmarks, we empirically validate the efficacy, generalizability, and robustness of cognitive expert modulation, demonstrating significant improvements in cognitive efficiency and problem-solving accuracy.

## 2 Identifying Cognitive Experts

In this section, we leverage normalized Pointwise Mutual Information (nPMI) (Bouma, 2009) to quantify the correlation between model thinking and each expert in a Mixture of Experts (MoE) reasoning model. We hypothesize that there are some “cognitive experts” selected by nPMI metric, which orchestrate meta-level reasoning for complex tasks.

### 2.1 Expert Specialization in MoE Models

In large reasoning models, deep thinking is manifested through key tokens, such as “<think>” to initiate reasoning, “</think>” to terminate it, and tokens like “recheck” to guide introspection. In the MoE framework, these tokens are generated after forward propagation through various model components, including the MoE routing mechanism that assigns them to specialized experts, with weights determining each expert’s contribution.

Formally, let us consider an MoE framework (DeepSeek-AI et al., 2024) with  $N$  experts, denoted  $\{E_1, \dots, E_i, \dots, E_N\}$ , at each layer. For each input token  $x$ , a gating function selects a subset  $S \subset \{E_1, \dots, E_N\}$  of  $O$  experts ( $O \leq N$ ), where  $|S| = O$ , and assigns weights  $w_i$  (with  $\sum_{i \in S} w_i = 1$ ) to each selected expert  $E_i \in S$ . The output  $h_x$  for the token  $x$  is computed as:

$$h_x = \sum_{i \in S} w_i \cdot E_i(x), \quad \text{where } |S| = O, \quad (1)$$

where  $E_i(x)$  represents the output of expert  $E_i$ , and  $w_i$  is the weight of the  $i$ -th selected expert. Prior work on MoE models shows that expert routing is often token-dependent (Xue et al., 2024), but recent study (Olson et al., 2025) indicates that DeepSeek-R1’s advanced reasoning enables its expert routing to focus on semantic specialization, surpassing token-dependent methods. We hypothesize that experts with consistently high co-occurrence scores with thinking tokens serve as key “cognitive experts”<sup>1</sup> responsible for meta-level reasoning.

**Measuring Correlation of Specialized Experts and Thinking Tokens** To examine whether a given expert consistently governs the model’s reasoning process, we measure the co-occurrence between its activation and specific reasoning-related marker tokens, such as “<think>”, “</think>”, and others. Formally, let  $x$  represent a token and  $y$  denote an expert  $E_i$ . We measure their association using pointwise mutual information (PMI). The PMI of  $x$  and  $y$  is defined as

$$\text{PMI}(x, y) = \log_2 \frac{p(x, y)}{p(x)p(y)} = \log_2 \frac{p(y|x)}{p(y)}, \quad (2)$$

where  $p(x, y)$  is the joint probability that  $x$  and  $y$  both occur, while  $p(x)$  and  $p(y)$  are their individual (marginal) probabilities, and  $p(y|x)$  is the conditional probability that  $y$  occurs given  $x$ .

For interpretability, we normalize PMI to the range  $[-1, +1]$ , yielding

$$\text{nPMI}(x, y) = \frac{\text{PMI}(x, y)}{-\log_2 p(x, y)}. \quad (3)$$

Thus,  $\text{nPMI}(x, y) \approx -1$  indicates that events  $x$  and  $y$  never co-occur,  $\text{nPMI}(x, y) = 0$  implies independence, and  $\text{nPMI}(x, y) \approx +1$  indicates they appear almost exclusively together (complete co-occurrence).

Let  $M$  denote the number of instances in a dataset, and let  $T$  represent the total number of tokens across all instances in the dataset. We denote by  $k_n$  the number of times the expert  $E_i$  is activated specifically when the thinking token (e.g. “<think>”) appears, and by  $K_n$  the total number of times  $E_i$  is activated across all tokens (including both thinking and non-thinking tokens). Since the

<sup>1</sup>Due to the complexity of LRMs, the “cognitive expert” proposed in this work is a hypothetical concept, and our findings are supported by empirical evidence rather than theoretical validation.

reasoning model generally generates one thinking start and end token for each instance, then we can achieve the following functions when  $x$  denotes “<think>” or “</think>”:

$$p(y = E_i|x) = \frac{k_n}{M}, \quad p(y = E_i) = \frac{K_n}{T}, \quad p(x, y = E_i|x) = \frac{k_n}{T}. \quad (4)$$

$$\text{nPMI}(x, y = E_i) = \frac{\log_2(\frac{k_n}{M}) + \log_2(\frac{T}{K_n})}{\log_2(\frac{T}{k_n})}. \quad (5)$$

Intuitively, if an expert  $E_i$  is activated almost exclusively during “<think>” and rarely (or never) at other tokens,  $k_n \approx K_n \approx M$ ,  $\text{nPMI}(x = \text{<think>}, y = E_i) \approx \frac{\log_2 1 + \log_2(\frac{T}{M})}{\log_2(\frac{T}{M})} \approx +1$ , indicating that this expert is effectively tied to the thinking marker. In other words, the expert’s entire usage focuses on activating the thinking token. Such specialists are prime candidates for “cognitive experts”, given their consistently high co-occurrence with the thinking marker tokens.

## 2.2 Identify Cognitive Experts

We observe that some experts exhibit high nPMI scores with both <think> and </think>, indicating a *bimodal association*. This suggests their broad involvement in the reasoning process rather than specialization in its initiation. To prioritize experts specialized in initiating (rather than terminating) reasoning, we adopt the following selection strategy:

We define a set of thinking tokens  $\Pi = \{\text{<think>}, \text{</think>}, \text{Alternatively}\}$ . The final nPMI score of expert  $E_i$  for thinking is formulated as:

$$\text{nPMI}_{E_i} = \sum_{t \in \Pi} c_x \cdot \text{nPMI}(x, y = E_i), \quad (6)$$

where  $t$  is a thinking token in set  $\Pi$ ,  $c_x$  denotes the coefficient associated with the token  $t$ , assigned as  $c_{\text{<think>}} = 1$ ,  $c_{\text{</think>}} = -1$ , and  $c_{\text{Alternatively}} = -1$ .

We select the top- $l$  experts based on their final nPMI scores for reasoning to form the *cognitive expert set*  $P$ . The weight of expert  $E_i$  is steered according to the following condition:

$$w_i = \begin{cases} w_i \cdot \beta & \text{if } E_i \in S \text{ and } E_i \in P, \\ w_i & \text{otherwise,} \end{cases} \quad (7)$$

where  $P = \{E_j \mid \text{nPMI}_{E_j} \text{ is among the top } l \text{ scores}\}$  denotes the set of *cognitive experts*,  $S$  is the subset selected by the gating function in Eq. 1, and  $\beta$  is the reinforcement multiplier. In other words, once these cognitive experts are identified, we can reinforce reasoning in the MoE model by controlling their steering multiplier  $\beta$ .

## 3 Experiments

**Research Questions** In this study, we investigate the following research questions:

- RQ1: Can the identified cognitive experts effectively enhance cognitive effort within MoE models?
- RQ2: Do cognitive experts differ across various domains (e.g., Math, physics, chemistry, and biology)?
- RQ3: Does reinforcing specific cognitive experts negatively impact the general problem-solving capabilities of MoE models?

### 3.1 Experimental Setup

**MoE-based Reasoning Models.** Currently available open-source MoE architectures tailored for large reasoning models tasks include DeepSeek-R1 (Guo et al., 2025) and Qwen3-235B (Team, 2025).

Table 1: Domain-specific cognitive experts (e.g., (layer ID, expert ID)) identified in DeepSeek-R1. “All” aggregates data across all four domains.

Domain	Identified Experts Ranked by nPMI Score				
	1st	2nd	3rd	4th	5th
Math	(39, 182)	(29, 126)	(14, 114)	(27, 45)	(16, 129)
Physics	(29, 126)	(39, 182)	(36, 53)	(39, 46)	(24, 159)
Chemistry	(7, 197)	(39, 182)	(22, 37)	(29, 106)	(29, 126)
Biology	(42, 194)	(22, 37)	(37, 241)	(43, 61)	(39, 188)
All	(39, 182)	(29, 126)	(29, 106)	(4, 214)	(50, 120)

DeepSeek-R1 selects 8 experts from a total of 256 at each layer, whereas Qwen3-235B selects 8 experts from a total of 128. We primarily use the DeepSeek-R1 (671B) model for our experiments, supplemented by additional evaluations on the Qwen3-235B model to examine the generalizability of cognitive experts. Note that we provide more experimental details in §A.

**Benchmarks.** We evaluate our approach on two challenging benchmarks designed specifically to test the reasoning abilities necessary for solving scientific problems across diverse domains:

- **AIME (MAA Committees)**: a dataset from the American Invitational Math Examination, which assesses advanced mathematical problem-solving skills. We use two recent test sets, AIME2024 and AIME2025, each comprising 30 problems.
- **GPQA Diamond (Rein et al., 2024)**: a comprehensive dataset of 198 expert-crafted multiple-choice questions in biology, chemistry, and physics, designed to test advanced scientific reasoning skills.

### 3.2 Effectiveness of Cognitive Experts

**Identification of Cognitive Experts.** To address RQ1, we first identify cognitive experts within two MoE reasoning models – DeepSeek-R1 (Guo et al., 2025) and Qwen3-235B (Team, 2025) – across four scientific domains. Taking Math as an illustrative example, we first use DeepSeek-R1 to generate answers on the AIME2024 dataset, simultaneously recording the expert selections at each token position during forward propagation. Next, we employ the nPMI metric defined in Eq. 6 to identify the top five experts that exhibit the strongest statistical association with reasoning-related marker tokens (e.g., “<think>”). These experts are thus identified as the key cognitive experts specialized for mathematical reasoning. Analogously, we apply this procedure to the biology, chemistry, and physics instances in the GPQA Diamond dataset to identify cognitive experts in these respective domains. As for Qwen3-235B, we follow a similar procedure but generate domain-specific responses with the Qwen3-235B model itself. This ensures consistent identification signals that correspond directly to the model under examination.

Cognitive experts identified within DeepSeek-R1 are summarized in Table 1. An analogous summary for Qwen3-235B is provided and discussed in Appendix B.1. From Table 1, we observe that the top two cognitive experts in the math, physics, and the aggregated “All” domains are remarkably consistent: (39, 182) and (29, 126). This strongly suggests these experts play critical and reliable roles in reasoning tasks requiring increased cognitive effort, particularly in quantitative and logic-intensive domains. The significant overlap observed between math and physics further implies a shared underlying cognitive strategy—likely focusing on symbolic manipulation and structured logical inference—which the model employs consistently across these domains. Additionally, the repeated appearance of certain experts in multiple domains supports our hypothesis: a subset of experts encodes generalized reasoning capabilities applicable across diverse scientific fields. Therefore, these cross-domain patterns indicate that DeepSeek-R1 may encode robust domain-general cognitive mechanisms, with some experts serving as reusable computational building blocks suitable for abstract reasoning and logical problem-solving tasks.

Table 2: Effect of Deepseek-R1 on AIME24 with reinforced cognitive experts, evaluated across different multipliers and varying numbers of Math-domain cognitive experts. “Random” denotes two randomly chosen experts. The row with Multiplier 1 denotes the performance of vanilla DeepSeek-R1 on AIME24.

Multiplier	Top1	Top2	Top3	Top4	Top5	Random
1				73.3		
2	70.0	70.0	76.7	73.3	73.3	70.0
4	76.7	<b>83.3</b>	73.3	66.7	76.7	73.3
8	76.7	73.3	<b>83.3</b>	73.3	73.3	70.0
16	80.0	80.0	1.7	76.7	73.3	73.3
32	70.0	<b>83.0</b>	73.3	73.3	73.3	76.7
64	80.0	<b>83.3</b>	60.0	53.3	50.0	66.7
128	70.0	<b>83.3</b>	43.3	26.7	13.3	63.3
256	73.3	60.0	10.0	6.7	0.0	73.3
512	63.3	46.7	6.7	3.3	0.0	63.3

Table 3: Performance of our approach on the AIME24 and generalization on the unseen AIME25.

Benchmark	Method	Accuracy	Thoughts	#Tokens
AIME24	DeepSeek-R1	73.3	12.0	9,219
	+RICE {(39,182), (29,126)}	<b>83.3</b>	10.2	8,317
AIME25	DeepSeek-R1	63.3	17.0	11,310
	+RICE {(39,182), (29,126)}	<b>73.3</b>	15.2	12,072
AIME24	Qwen3-235B	86.7	20.1	10,956
	+RICE {(70,47), (23,115)}	86.7	16.2	10,722
AIME25	Qwen3-235B	66.7	19.7	15,013
	+RICE {(70,47), (23,115)}	<b>73.3</b>	16.8	13,935

**Reinforcing Cognitive Experts.** Once identified, we reinforce the cognitive experts identified from the Math domain (AIME24) and evaluate their performance under different reinforcement configurations on the same benchmark AIME24 (Table 2). The optimal hyperparameters – the number of cognitive experts  $l$  and the steering multiplier  $\beta$ —are selected based on this evaluation and used in all subsequent experiments. We then assess the generalization ability of these reinforced experts on the unseen, more challenging tasks from AIME25 (Table 3).

From Table 2, we observe that **reinforcing two top-ranked cognitive experts significantly enhances the model’s reasoning ability**. Notably, using two experts with the reinforcement multiplier of 4, 32, 64, or 128 achieves the highest accuracy of 83.3%. In contrast, applying an excessively large multiplier (e.g., 512) causes a dramatic drop in accuracy, often to near zero. This failure mode is characterized by the model repetitively generating meaningless tokens, suggesting that overly aggressive reinforcement disrupts the model’s generation dynamics. Overall, moderate reinforcement of well-identified cognitive experts leads to consistent improvements, whereas over-reinforcement or random expert selection results in performance degradation. However, reinforcing two randomly selected experts across a wide multiplier range (2 to 512) yields minimal performance variation. Therefore, we use *two experts with a reinforcement multiplier 64 for all subsequent experiments*.

We directly apply the cognitive experts identified from AIME24 to solve unseen and more challenging reasoning problems in AIME25. As shown in Table 3, these cognitive experts generalize well to the AIME25 test set. For DeepSeek-R1, the accuracy improves from 63.3% to 73.3% when guided by the identified cognitive experts. Similarly, for Qwen3-235B, accuracy increases from 66.7% to 73.3%. The above phenomenon demonstrates **the transferability and robustness of the expert selection across tasks with higher cognitive demands**.



Table 4: Effect of cognitive experts of Deepseek-R1 across different domains.

Domain	Math	Physics	Chemistry	Biology	Average
R1	73.3	91.9	49.5	79.0	73.4
Math	<b>83.3</b>	89.5	50.4	79.0	<b>75.6</b>
Physics	<b>83.3</b>	89.5	50.4	79.0	<b>75.6</b>
Chemistry	80.0	<b>95.4</b>	<b>52.7</b>	68.4	74.1
Biology	73.3	93.0	47.3	73.9	71.9
All	<b>83.3</b>	89.5	50.4	79.0	<b>75.6</b>

Crucially, the observed accuracy improvements do not necessarily entail increased computational cost in terms of token usage, supporting our hypothesis that our method encourages deeper thinking rather than just longer outputs. Our cognitive expert strategy, despite improving average accuracy of Deepseek-R1 on AIME24, uses more efficient reasoning thought<sup>2</sup> (10.2 vs 12.0) and tokens (8,317 vs 9,219) on average compared to the baseline. This efficiency phenomenon is also observed in Qwen3-2-35B, where the substantial accuracy gain (+6.6%) is accompanied by a notable reduction in thought (16.8 vs 19.7) and token count (13,935 vs 15,013). This suggests that **reinforcing cognitive experts helps the model to reason more effectively**, focusing computational effort more productively within the reasoning process without generating excessive verbosity. The reasoning effectiveness can be clearly observed in Table 7, where our RICE demonstrates deeper and more consistent reasoning, leading directly to the correct answer. In contrast, vanilla DeepSeek-R1 exhibits more frequent shifts in reasoning and fails to commit to its initially correct deductions. Additional pass@k performance using the model’s officially recommended top- $p$  sampling strategy (provided in §B.2) further supports this observation.

### 3.3 Performance of Cognitive Experts across Domains

To address RQ2, we evaluate the transferability of domain-specific cognitive experts by applying expert sets identified from one domain to others. As the top-2 experts selected from *Math*, *Physics*, and the *All* domains are identical, their results are the same across domains. As shown in Table 4, we have several observations:

**Cognitive Experts Generalize Well across Domains.** Our evaluation, summarized in Fig 1 and Table 4, clearly illustrates the efficacy of the identified cognitive experts in enhancing the DeepSeek-R1 model’s reasoning capability across multiple domains. Leveraging cognitive experts identified from aggregated data (“All” domains) shows marked overall improvement, raising the average accuracy from 73.4% to 75.6%. Notably, substantial improvement is observed in the Math tasks (from 73.3% to 83.3%). Moderate accuracy gains are also seen in Chemistry (from 49.5% to 50.4%) and minor degradation observed in Physics (from 91.9% to 89.5%), indicating broad applicability and effectiveness of these general reasoning modulators across diverse problem sets. Biology tasks show stable performance, unaffected by general expert modulation.

**Domain-specific Expert Sets Provide Targeted Gains.** Further analysis demonstrates the nuanced implications of domain-specific cognitive experts. Chemistry-identified experts outperform general experts significantly within their native Chemistry domain (49.5% to 52.7%) and notably enhance Physics performance (91.9% to 95.4%), highlighting potential cross-domain synergies between physics and chemistry reasoning processes. However, this specialization lowers the accuracy in Math (from 83.3% with general experts to 80.0%) and more substantially limits the Biology domain performance (from 79.0% to 68.4%). Similarly, Biology-derived experts enhance task-specific results (from 91.9% to 93.0% in Physics) but degrade average performance across other domains, indicating

<sup>2</sup>The “Thoughts” metric refers to the underthinking score introduced in prior work (Wang et al., 2025a), which quantifies reasoning efficiency, with lower values indicating higher efficiency.

further that specialized expert selections may negatively impact general cognitive reasoning by reinforcing overly specialized activations.

**No Evidence of Harmful Side-effects on Other Domains.** Our experimental findings clearly confirm that cognitive experts, either chosen from aggregated cross-domain data or specific domains, constitute effective cognitive modulators that enhance model reasoning accuracy and efficiency. General-purpose expert adjustments deliver robust cross-domain improvements, demonstrating their fundamental importance to reasoning processes regardless of subject matter. Meanwhile, domain-specialized expert modulation illustrates substantial potential for targeted cognitive improvements, particularly within closely related scientific domains. Together, these insights validate our proposed approach as versatile, effective, and immediately deployable for enhancing efficiency, accuracy, and overall reasoning proficiency of existing MoE-based large reasoning models.

### 3.4 Impact of Reinforced Cognitive Experts on General Capabilities

To address RQ3, we investigate whether reinforcing cognitive experts negatively impacts the model’s general capabilities, such as instruction-following. To this end, we evaluate reinforced models on the ArenaHard benchmark (Li et al., 2024) to assess potential adverse impacts on general capabilities. The ArenaHard benchmark, designed to evaluate instruction-following capabilities, comprises 500 challenging queries spanning diverse scenarios. We randomly select 50 queries as the test data and employ GPT-4-Turbo to judge pairwise comparisons of outputs against the GPT-4-0613 baseline.

#### Reinforcing Cognitive Experts Maintains or Slightly Improves General Instruction-following Capabilities.

Our experimental evaluation on the ArenaHard benchmark demonstrates that reinforcing the identified cognitive experts does not adversely impact the model’s capability to handle general, challenging instruction-following tasks. As shown in Table 5, models steered by cognitive experts derived from each domain consistently maintain or marginally improve upon the baseline DeepSeek-R1 accuracy of 91.0%. Specifically, the domain-specific cognitive experts from Chemistry and Biology show notable accuracy enhancements (from 91.0% to 94.0% in Chemistry; from 91.0% to 93.0% in Biology), underscoring the potential for positive transfer of reasoning-rich expert reinforcement to general-purpose capabilities. Moreover, the general experts (“All” domain) also marginally improve performance (to 92.0%), confirming that cognitive expert-control has a neutral-to-beneficial impact on general instruction-following capabilities.

Table 5: Effect of reinforced cognitive experts of Deepseek-R1 on ArenaHard.

Method	Accuracy	#Token
Vanilla	91.0	2,919
<i>Reinforce Experts from different domains</i>		
Math	92.0	2,933
Physics	92.0	2,933
Chemistry	94.0	3,332
Biology	93.0	3,072
All	92.0	2,933

**Steering of Cognitive Experts Results in Moderately Increased Verbosity.** The analysis of token counts further reveals that cognitive expert steering moderately increases model verbosity in response generation. For example, Chemistry and Biology models increase average token counts notably (from 2,919 to 3,332 tokens and from 2,919 to 3,072 tokens, respectively), highlighting that the activation of certain domain-specific cognitive experts may favor more detailed deliberations. Nevertheless, the overall increase in verbosity is moderate, indicating a desirable balance between detail-oriented reasoning and response conciseness.

#### Overall, Reinforcing Cognitive Experts Does not Hinder but rather Supports General Capabilities.

These findings collectively confirm our approach as effective and safe for targeted, lightweight interventions. Reinforcing cognitive experts significantly enhances model performance within their original domains and has either neutral or positive effects on general-purpose instruction-following benchmarks. The moderate increase in verbosity indicates richer, more thoughtful reasoning, aligning with the intended goal of encouraging deeper cognitive processing without sacrificing prac-



tality. This highlights the practicality and versatility of our approach in improving existing MoE model reasoning efficacy and general cognitive capabilities through strategic expert modulation.

### 3.5 Comparison with Other Steering Methods

We compare our RICE against two prevalent inference-time methods for reasoning tasks: prompt engineering and decoding constraints. Specifically, we analyze two prompt configurations: placing the prompt before the `<think>` token (Prompt<sub>before</sub>) and after the `<think>` token (Prompt<sub>after</sub>), with details provided in Appendix A.1. For decoding constraints, we adopt a strategy similar to TIP from Wang et al. (2025a), which curtails the generation of alternative solutions to promote coherent and focused reasoning. In our work, we constrain the thinking mark tokens (e.g., `<think>` and “Alternatively”) rather than inefficient tokens (e.g., “messy”), and refer to our variant as TIP<sub>t</sub>.

Table 6 compares our cognitive expert modulation method against prompting (both before and after the `<think>` token) and decoding constraints (TIP) on the challenging AIME benchmarks. Our approach achieves the highest average improvement, boosting DeepSeek-R1’s performance from a baseline of 68.3% to 78.7%, surpassing the best-performing baseline TIP<sub>t</sub> by 2.0% in average accuracy. Notably, our method achieves particularly pronounced gains on AIME24 (83.3%) compared to Prompt<sub>before</sub>

Table 6: Comparison with other steering methods on AIME24 and AIME25.

Method	AIME24	AIME25	Avg.
DeepSeek-R1	73.3	63.3	68.3
Prompt <sub>before</sub>	76.7	73.3	75.0
Prompt <sub>after</sub>	76.7	66.7	71.7
TIP <sub>t</sub>	80.0	73.3	76.7
RICE	83.3	73.3	78.7

(76.7%) and TIP<sub>t</sub> (80.0%), indicating a superior ability to enhance cognitive effort and deep reasoning through targeted modulation of expert routing weights. These results validate our hypothesis that strategically steering cognitive experts significantly reinforces appropriate cognitive behavior and enhances reasoning accuracy more effectively than generic prompting and decoding constraints.

Table 7: An example of our RICE corrects **wrong reasoning (in red)** generated by Deepseek-R1 into **correct reasoning (in green)**.

Problem	Let ABCDEF be a convex equilateral hexagon in... and EF has side lengths 200, 240, and 300. Find the side length of the hexagon.
DeepSeek-R1	Okay, so I need to find the side length...the triangle formed by extending AB, CD, and EF has sides 200, 240, and 300. Let me sketch this mentally... <b>Alternatively</b> , maybe we can model the hexagon as being formed by three pairs of parallel sides... <b>Alternatively</b> , think of the hexagon as a combination of three parallelograms... <b>Alternatively</b> , using the concept of similar triangles... <b>Alternatively</b> , maybe each pair of opposite sides of the hexagon is parallel to a different side of the triangle... <b>I will posit that the answer is 80...Therefore, the answer is 75/2...Let me verify this result...Thus, <math>s = 1200 / 23</math>. Wait, this <b>contradicts</b> our previous result...I think the answer is 80...based on the thorough step-by-step analysis, <math>\boxed{1200/23}</math> is the conclusion.</b>
+RICE	Okay, so I need to find the side length...Let me draw a rough sketch in my mind... <b>So</b> , the triangle is such that each side is aligned with one of the extended sides of the hexagon...But maybe using <b>coordinate geometry could help here</b> ....Not sure, maybe this is <b>getting too complicated</b> ... <b>Alternatively</b> , maybe using ratios... <b>Alternatively</b> , <b>after struggling through all these approaches, perhaps the answer is related to the harmonic mean of the triangle’s sides...Therefore, I think the answer is 80. But need to verify...Therefore, the side length of the hexagon is <math>\boxed{80}</math>.</b>

## 4 Related Work

**Large Reasoning Models** Large Reasoning Models (LRMs) significantly enhances the reasoning capabilities of large language models (LLMs) (Jaech et al., 2024; Xia et al., 2025). Prominent implementations include OpenAI’s o1 (Jaech et al., 2024), QwQ (Qwen, 2024), Qwen3 (Team, 2025), DeepSeek-R1 (Guo et al., 2025), Claude 3.7 (Anthropic, 2025) and Kimi-1.5 (Team et al., 2025) achieve human-like reasoning by leveraging scaled test-time computation. In particular, the open-source DeepSeek-R1 utilizes a Mixture-of-Experts (MoE) architecture (Dai et al., 2024b) with sparsely activated parameters, selectively activating only 8 out of 256 experts per layer (DeepSeek-AI et al., 2024). This MoE architecture has been widely adopted in recent LLMs (Llama, 2025; Muennighoff et al., 2024; Shazeer et al., 2017), achieving an optimal balance between computational efficiency and competitive performance in complex reasoning tasks.

**MoE Models** Previous research on Mixture of Experts (MoE) models indicates that expert routing is primarily token-dependent (Xue et al., 2024). However, Olson et al. (2025) demonstrate that DeepSeek-R1’s advanced reasoning capabilities enable its routing mechanism to achieve greater semantic specialization and structured cognitive processing, representing a substantial advancement over prior MoE models. Subsequently, Hazra et al. (2025) train sparse autoencoders (SAEs) on DeepSeek-R1, identifying interpretable features such as backtracking, division, and rapid response patterns within the SAEs space. However, training SAEs is computationally intensive, posing significant resource demands. We employ the normalized Pointwise Mutual Information (nPMI) metric to evaluate expert specialization, requiring only a single forward propagation.

**Efficient Thinking** Despite significant advancements, large reasoning models continue to encounter substantial cognitive challenges, such as the overthinking (Chen et al., 2024; Sui et al., 2025b; Cuadron et al., 2025; Zaremba et al., 2025) and underthinking phenomenon (Wang et al., 2025a; Qu et al., 2025; Ballon et al., 2025). Subsequent efforts address these issues through rule-based stop, decoding constraints (Tran et al., 2025; Wang et al., 2025a; Ding et al., 2025; Xiang et al., 2025; Ma et al., 2025; Muennighoff et al., 2025; Han et al., 2024; Aytas et al., 2025; Zhang et al., 2025b), steering vectors (Chen et al., 2025b; Cyberek and Evans, 2025), and parameters tuning (Sun et al., 2025; Chen et al., 2024; Ye et al., 2025; Aggarwal and Welleck, 2025; Hao et al., 2024). There are also some works specifically designed to improve reasoning capabilities in MoE architectures by re-mixing experts through gradient-based optimization (Li et al., 2025) or by expert pruning via sparse dictionary learning (Tang et al., 2025). However, the resource-intensive nature of expert re-mixing algorithms makes them impractical to scale to large models such as 671B-parameter systems, whereas our method is lightweight and directly applicable to such large-scale settings. Generally, in contrast to the above strategies that primarily rely on crafted rules, extensive labeled data, or computationally expensive parameter training, our *reinforcing cognitive experts* approach achieves more efficient and deeper reasoning with only a single forward propagation, without requiring any supervision signals or additional training.

## 5 Conclusion and Future Work

In this work, we explore “cognitive experts” within MoE-based LRMs, proposing a computationally efficient method based on nPMI to identify experts central to cognitive deliberation. We empirically show that steering these experts allows steering of reasoning processes with minimal additional computational overhead or training burden. Critically, the identified experts demonstrated strong transferability across multiple scientific domains, suggesting a fundamental, domain-general cognitive function within the reasoning model. Our findings highlight the promising role of expert specialization as a mechanistic lever for enhancing cognitive control, interpretability, and efficiency in large-scale reasoning architectures.

Future directions include deeper investigations into the structural properties and broader applicability of cognitive experts, as well as integration with other cognitive control strategies to further

enhance reasoning robustness. By uncovering this hidden layer of functional specialization within MoE models, we may open new avenues for fine-grained control over neural reasoning processes, more closely mirroring the modularity observed in biological cognitive systems.

## Limitations and Broader Impacts

The internal coordination mechanisms of long-range reasoning models are inherently complex, and our nPMI-based approach may not fully capture all relevant interactions. Future work should explore more sophisticated metrics for expert identification. Besides, our validation was constrained by the current availability of open-source MoE architectures designed for long-range reasoning, limited to DeepSeek-R1 (Guo et al., 2025) and Qwen3-235B (Team, 2025). Additional testing across more diverse architectures is warranted. The ability to precisely control reasoning processes in large reasoning models has significant implications for both AI safety and efficiency. Our method’s minimal computational overhead makes it particularly promising for real-world applications where resource constraints are critical. The observed cross-domain transferability of cognitive experts suggests exciting possibilities for developing more general and adaptable AI systems.

## References

- Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, et al. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*, 2025.
- Qwen Team. Qwen3: Think deeper, act faster. 2025. URL <https://qwenlm.github.io/zh/blog/qwen3/>.
- Xingyu Chen, Jiahao Xu, Tian Liang, Zhiwei He, Jianhui Pang, Dian Yu, Linfeng Song, Qiuzhi Liu, Mengfei Zhou, Zhuosheng Zhang, Rui Wang, Zhaopeng Tu, Haitao Mi, and Dong Yu. Do NOT think that much for  $2+3=?$  on the overthinking of o1-like llms. *CoRR*, abs/2412.21187, 2024. doi: 10.48550/ARXIV.2412.21187. URL <https://doi.org/10.48550/arXiv.2412.21187>.
- Yue Wang, Qiuzhi Liu, Jiahao Xu, Tian Liang, Xingyu Chen, Zhiwei He, Linfeng Song, Dian Yu, Juntao Li, Zhuosheng Zhang, Rui Wang, Zhaopeng Tu, Haitao Mi, and Dong Yu. Thoughts are all over the place: On the underthinking of o1-like llms. *CoRR*, abs/2501.18585, 2025a. doi: 10.48550/ARXIV.2501.18585. URL <https://doi.org/10.48550/arXiv.2501.18585>.
- Aaron Jaech, Adam Kalai, Adam Lerer, Adam Richardson, Ahmed El-Kishky, Aiden Low, Alec Helyar, Aleksander Madry, Alex Beutel, Alex Carney, et al. Openai o1 system card. *arXiv preprint arXiv:2412.16720*, 2024.
- Yixin Ji, Juntao Li, Hai Ye, Kaixin Wu, Jia Xu, Linjian Mo, and Min Zhang. Test-time computing: from system-1 thinking to system-2 thinking. *arXiv preprint arXiv:2501.02497*, 2025.
- Qiyuan Zhang, Fuyuan Lyu, Zexu Sun, Lei Wang, Weixu Zhang, Zhihan Guo, Yufei Wang, Irwin King, Xue Liu, and Chen Ma. What, how, where, and how well? a survey on test-time scaling in large language models. *arXiv preprint arXiv:2503.24235*, 2025a.
- Yang Sui, Yu-Neng Chuang, Guanchu Wang, Jiamu Zhang, Tianyi Zhang, Jiayi Yuan, Hongyi Liu, Andrew Wen, Shaochen Zhong, Hanjie Chen, et al. Stop overthinking: A survey on efficient reasoning for large language models. *arXiv preprint arXiv:2503.16419*, 2025a.
- Sicheng Feng, Gongfan Fang, Xinyin Ma, and Xinchao Wang. Efficient reasoning models: A survey. *arXiv preprint arXiv:2504.10903*, 2025.
- Qianjun Pan, Wenkai Ji, Yuyang Ding, Junsong Li, Shilian Chen, Junyi Wang, Jie Zhou, Qin Chen, Min Zhang, Yulan Wu, et al. A survey of slow thinking-based reasoning llms using reinforced learning and inference-time scaling law. *arXiv preprint arXiv:2505.02665*, 2025.

- Xiaoye Qu, Yafu Li, Zhaochen Su, Weigao Sun, Jianhao Yan, Dongrui Liu, Ganqu Cui, Daizong Liu, Shuxian Liang, Junxian He, et al. A survey of efficient reasoning for large reasoning models: Language, multimodality, and beyond. *arXiv preprint arXiv:2503.21614*, 2025.
- Qiguang Chen, Libo Qin, Jinhao Liu, Dengyun Peng, Jiannan Guan, Peng Wang, Mengkang Hu, Yuhang Zhou, Te Gao, and Wanxiang Che. Towards reasoning era: A survey of long chain-of-thought for reasoning large language models. *arXiv preprint arXiv:2503.09567*, 2025a.
- Rui Wang, Hongru Wang, Boyang Xue, Jianhui Pang, Shudong Liu, Yi Chen, Jiahao Qiu, Derek Fai Wong, Heng Ji, and Kam-Fai Wong. Harnessing the reasoning economy: A survey of efficient reasoning for large language models. *arXiv preprint arXiv:2503.24377*, 2025b.
- Tong Wu, Chong Xiang, Jiachen T Wang, and Prateek Mittal. Effectively controlling reasoning models through thinking intervention. *arXiv preprint arXiv:2503.24370*, 2025.
- Ximing Lu, Seungju Han, David Acuna, Hyunwoo Kim, Jaehun Jung, Shrimai Prabhumoye, Niklas Muennighoff, Mostofa Patwary, Mohammad Shoeybi, Bryan Catanzaro, et al. Retro-search: Exploring untaken paths for deeper and efficient reasoning. *arXiv preprint arXiv:2504.04383*, 2025.
- Damai Dai, Chengqi Deng, Chenggang Zhao, RX Xu, Huazuo Gao, Deli Chen, Jiashi Li, Wangding Zeng, Xingkai Yu, Yu Wu, et al. Deepseekmoe: Towards ultimate expert specialization in mixture-of-experts language models. *arXiv preprint arXiv:2401.06066*, 2024a.
- DeepSeek-AI, Aixin Liu, Bei Feng, Bing Xue, Bingxuan Wang, Bochao Wu, Chengda Lu, Chenggang Zhao, Chengqi Deng, Chenyu Zhang, Chong Ruan, Damai Dai, Daya Guo, Dejian Yang, Deli Chen, Dongjie Ji, Erhang Li, Fangyun Lin, Fucong Dai, Fuli Luo, Guangbo Hao, Guanting Chen, Guowei Li, H. Zhang, Han Bao, Hanwei Xu, Haocheng Wang, Haowei Zhang, Honghui Ding, Huajian Xin, Huazuo Gao, Hui Li, Hui Qu, J. L. Cai, Jian Liang, Jianzhong Guo, Jiaqi Ni, Jiashi Li, Jiawei Wang, Jin Chen, Jingchang Chen, Jingyang Yuan, Junjie Qiu, Junlong Li, Junxiao Song, Kai Dong, Kai Hu, Kaige Gao, Kang Guan, Kexin Huang, Kuai Yu, Lean Wang, Lecong Zhang, Lei Xu, Leyi Xia, Liang Zhao, Litong Wang, Liyue Zhang, Meng Li, Miaojun Wang, Mingchuan Zhang, Minghua Zhang, Minghui Tang, Mingming Li, Ning Tian, Panpan Huang, Peiyi Wang, Peng Zhang, Qiancheng Wang, Qihao Zhu, Qinyu Chen, Qiushi Du, R. J. Chen, R. L. Jin, Ruiqi Ge, Ruisong Zhang, Ruizhe Pan, Runji Wang, Runxin Xu, Ruoyu Zhang, Ruyi Chen, S. S. Li, Shanghao Lu, Shangyan Zhou, Shanhuang Chen, Shaoqing Wu, Shengfeng Ye, Shengfeng Ye, Shirong Ma, Shiyu Wang, Shuang Zhou, Shuiping Yu, Shunfeng Zhou, Shuting Pan, T. Wang, Tao Yun, Tian Pei, Tianyu Sun, W. L. Xiao, and Wangding Zeng. Deepseek-v3 technical report. *CoRR*, abs/2412.19437, 2024. URL <https://doi.org/10.48550/arXiv.2412.19437>.
- Fuzhao Xue, Zian Zheng, Yao Fu, Jinjie Ni, Zangwei Zheng, Wangchunshu Zhou, and Yang You. Openmoe: An early effort on open mixture-of-experts language models. In *Forty-first International Conference on Machine Learning, ICML 2024, Vienna, Austria, July 21-27, 2024*. OpenReview.net, 2024. URL <https://openreview.net/forum?id=1YDeZU8Lt5>.
- Robert MG Reinhart and John A Nguyen. Working memory revived in older adults by synchronizing rhythmic brain circuits. *Nature neuroscience*, 22(5):820–827, 2019.
- Miles Wischniewski, Ivan Alekseichuk, and Alexander Opitz. Neurocognitive, physiological, and biophysical effects of transcranial alternating current stimulation. *Trends in Cognitive Sciences*, 27(2):189–205, 2023.
- Desmond J Oathes, Romain JP Duprat, Justin Reber, Ximo Liang, Morgan Scully, Hannah Long, Joseph A Deluise, Yvette I Sheline, and Kristin A Linn. Non-invasively targeting, probing and modulating a deep brain circuit for depression alleviation. *Nature Mental Health*, 1(12):1033–1042, 2023.
- Shrey Grover, John A Nguyen, Vighnesh Viswanathan, and Robert MG Reinhart. High-frequency neuromodulation improves obsessive–compulsive behavior. *Nature medicine*, 27(2):232–238, 2021.

- Gerlof Bouma. Normalized (pointwise) mutual information in collocation extraction. *Proceedings of GSCL*, 30:31–40, 2009.
- Matthew Lyle Olson, Neale Ratzlaff, Musashi Hinck, Man Luo, Sungduk Yu, Chendi Xue, and Vasudev Lal. Semantic specialization in moe appears with scale: A study of deepseek r1 expert specialization. *arXiv preprint arXiv:2502.10928*, 2025.
- MAA Committees. Aime problems and solutions. [https://artofproblemsolving.com/wiki/index.php/AIME\\_Problems\\_and\\_Solutions](https://artofproblemsolving.com/wiki/index.php/AIME_Problems_and_Solutions).
- David Rein, Betty Li Hou, Asa Cooper Stickland, Jackson Petty, Richard Yuanzhe Pang, Julien Dirani, Julian Michael, and Samuel R. Bowman. GPQA: A graduate-level google-proof q&a benchmark. In *First Conference on Language Modeling*, 2024. URL <https://openreview.net/forum?id=Ti67584b98>.
- Tianle Li, Wei-Lin Chiang, Evan Frick, Lisa Dunlap, Tianhao Wu, Banghua Zhu, Joseph E. Gonzalez, and Ion Stoica. From crowdsourced data to high-quality benchmarks: Arena-hard and benchmark-builder pipeline. *CoRR*, abs/2406.11939, 2024. URL <https://doi.org/10.48550/arXiv.2406.11939>.
- Shijie Xia, Yiwei Qin, Xuefeng Li, Yan Ma, Run-Ze Fan, Steffi Chern, Haoyang Zou, Fan Zhou, Xiangkun Hu, Jiahe Jin, et al. Generative ai act ii: Test time scaling drives cognition engineering. *arXiv preprint arXiv:2504.13828*, 2025.
- Qwen. Qwq: Reflect deeply on the boundaries of the unknown. 2024. URL <https://qwenlm.github.io/blog/qwq-32b-preview/>.
- Anthropic. Claude 3.7 sonnet. 2025. URL <https://www.anthropic.com/claude/sonnet>.
- Kimi Team, Angang Du, Bofei Gao, BOWEI Xing, Changjiu Jiang, Cheng Chen, Cheng Li, Chenjun Xiao, Chenzhuang Du, Chonghua Liao, et al. Kimi k1. 5: Scaling reinforcement learning with llms. *arXiv preprint arXiv:2501.12599*, 2025.
- Damai Dai, Chengqi Deng, Chenggang Zhao, R. X. Xu, Huazuo Gao, Deli Chen, Jiashi Li, Wangding Zeng, Xingkai Yu, Y. Wu, Zhenda Xie, Y. K. Li, Panpan Huang, Fuli Luo, Chong Ruan, Zhifang Sui, and Wenfeng Liang. Deepseekmoe: Towards ultimate expert specialization in mixture-of-experts language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, ACL 2024, Bangkok, Thailand, August 11-16, 2024, pages 1280–1297. Association for Computational Linguistics, 2024b. URL <https://doi.org/10.18653/v1/2024.acl-long.70>.
- Llama. The llama 4 herd: The beginning of a new era of natively multimodal ai innovation. 2025. URL <https://www.llama.com/models/llama-4/>.
- Niklas Muennighoff, Luca Soldaini, Dirk Groeneveld, Kyle Lo, Jacob Morrison, Sewon Min, Weijia Shi, Pete Walsh, Oyvind Tafjord, Nathan Lambert, Yuling Gu, Shane Arora, Akshita Bhagia, Dustin Schwenk, David Wadden, Alexander Wettig, Binyuan Hui, Tim Dettmers, Douwe Kiela, Ali Farhadi, Noah A. Smith, Pang Wei Koh, Amanpreet Singh, and Hannaneh Hajishirzi. Olmoe: Open mixture-of-experts language models. *CoRR*, abs/2409.02060, 2024. doi: 10.48550/ARXIV.2409.02060. URL <https://doi.org/10.48550/arXiv.2409.02060>.
- Noam Shazeer, Azalia Mirhoseini, Krzysztof Maziarsz, Andy Davis, Quoc V. Le, Geoffrey E. Hinton, and Jeff Dean. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net, 2017. URL <https://openreview.net/forum?id=B1ckMDqlg>.
- Dron Hazra, Max Loeffler, Murat Cubuktepe, Levon Avagyan, Liv Gorton, Mark Bissell, Owen Lewis, Thomas McGrath, and Daniel Balsam. Under the hood of a reasoning model. 2025. URL <https://www.goodfire.ai/blog/under-the-hood-of-a-reasoning-model>.



- Yang Sui, Yu-Neng Chuang, Guanchu Wang, Jiamu Zhang, Tianyi Zhang, Jiayi Yuan, Hongyi Liu, Andrew Wen, Shaochen Zhong, Hanjie Chen, and Xia Ben Hu. Stop overthinking: A survey on efficient reasoning for large language models. *CoRR*, abs/2503.16419, 2025b. doi: 10.48550/ARXIV.2503.16419. URL <https://doi.org/10.48550/arXiv.2503.16419>.
- Alejandro Cuadron, Dacheng Li, Wenjie Ma, Xingyao Wang, Yichuan Wang, Siyuan Zhuang, Shu Liu, Luis Gaspar Schroeder, Tian Xia, Huanzhi Mao, Nicholas Thumiger, Aditya Desai, Ion Stoica, Ana Klimovic, Graham Neubig, and Joseph E. Gonzalez. The danger of overthinking: Examining the reasoning-action dilemma in agentic tasks. *CoRR*, abs/2502.08235, 2025. doi: 10.48550/ARXIV.2502.08235. URL <https://doi.org/10.48550/arXiv.2502.08235>.
- Wojciech Zaremba, Evgenia Nitishinskaya, Boaz Barak, Stephanie Lin, Sam Toyer, Yaodong Yu, Rachel Dias, Eric Wallace, Kai Xiao, Johannes Heidecke, et al. Trading inference-time compute for adversarial robustness. *arXiv preprint arXiv:2501.18841*, 2025.
- Marthe Ballon, Andres Algaba, and Vincent Ginis. The relationship between reasoning and performance in large language models—o3 (mini) thinks harder, not longer. *arXiv preprint arXiv:2502.15631*, 2025.
- Bao Hieu Tran, Nguyen Cong Dat, Nguyen Duc Anh, and Hoang Thanh-Tung. Learning to stop overthinking at test time. *CoRR*, abs/2502.10954, 2025. doi: 10.48550/ARXIV.2502.10954. URL <https://doi.org/10.48550/arXiv.2502.10954>.
- Yifu Ding, Wentao Jiang, Shunyu Liu, Yongcheng Jing, Jinyang Guo, Yingjie Wang, Jing Zhang, Zengmao Wang, Ziwei Liu, Bo Du, Xianglong Liu, and Dacheng Tao. Dynamic parallel tree search for efficient LLM reasoning. *CoRR*, abs/2502.16235, 2025. doi: 10.48550/ARXIV.2502.16235. URL <https://doi.org/10.48550/arXiv.2502.16235>.
- Kun Xiang, Zhili Liu, Zihao Jiang, Yunshuang Nie, Kaixin Cai, Yiyang Yin, Runhui Huang, Haoxiang Fan, Hanhui Li, Weiran Huang, et al. Can atomic step decomposition enhance the self-structured reasoning of multimodal large models? *arXiv preprint arXiv:2503.06252*, 2025.
- Wenjie Ma, Jingxuan He, Charlie Snell, Tyler Griggs, Sewon Min, and Matei Zaharia. Reasoning models can be effective without thinking. *arXiv preprint arXiv:2504.09858*, 2025.
- Niklas Muennighoff, Zitong Yang, Weijia Shi, Xiang Lisa Li, Li Fei-Fei, Hannaneh Hajishirzi, Luke Zettlemoyer, Percy Liang, Emmanuel J. Candès, and Tatsunori Hashimoto. s1: Simple test-time scaling. *CoRR*, abs/2501.19393, 2025. doi: 10.48550/ARXIV.2501.19393. URL <https://doi.org/10.48550/arXiv.2501.19393>.
- Tingxu Han, Zhenting Wang, Chunrong Fang, Shiyu Zhao, Shiqing Ma, and Zhenyu Chen. Token-budget-aware LLM reasoning. *CoRR*, abs/2412.18547, 2024. doi: 10.48550/ARXIV.2412.18547. URL <https://doi.org/10.48550/arXiv.2412.18547>.
- Simon A. Aytes, Jinheon Baek, and Sung Ju Hwang. Sketch-of-thought: Efficient LLM reasoning with adaptive cognitive-inspired sketching. *CoRR*, abs/2503.05179, 2025. doi: 10.48550/ARXIV.2503.05179. URL <https://doi.org/10.48550/arXiv.2503.05179>.
- Jintian Zhang, Yuqi Zhu, Mengshu Sun, Yujie Luo, Shuofei Qiao, Lun Du, Da Zheng, Huajun Chen, and Ningyu Zhang. Lightthinker: Thinking step-by-step compression. *arXiv preprint arXiv:2502.15589*, 2025b.
- Runjin Chen, Zhenyu Zhang, Junyuan Hong, Souvik Kundu, and Zhangyang Wang. Seal: Steerable reasoning calibration of large language models for free. *arXiv preprint arXiv:2504.07986*, 2025b.
- Hannah Cyberek and David Evans. Steering the censorship: Uncovering representation vectors for llm “thought” control. *arXiv preprint arXiv:2504.17130*, 2025.
- Chung-En Sun, Ge Yan, and Tsui-Wei Weng. Thinkedit: Interpretable weight editing to mitigate overly short thinking in reasoning models. *arXiv preprint arXiv:2503.22048*, 2025.

- Yixin Ye, Zhen Huang, Yang Xiao, Ethan Chern, Shijie Xia, and Pengfei Liu. LIMO: less is more for reasoning. *CoRR*, abs/2502.03387, 2025. doi: 10.48550/ARXIV.2502.03387. URL <https://doi.org/10.48550/arXiv.2502.03387>.
- Pranjal Aggarwal and Sean Welleck. L1: controlling how long A reasoning model thinks with reinforcement learning. *CoRR*, abs/2503.04697, 2025. doi: 10.48550/ARXIV.2503.04697. URL <https://doi.org/10.48550/arXiv.2503.04697>.
- Shibo Hao, Sainbayar Sukhbaatar, DiJia Su, Xian Li, Zhiting Hu, Jason Weston, and Yuandong Tian. Training large language models to reason in a continuous latent space. *CoRR*, abs/2412.06769, 2024. doi: 10.48550/ARXIV.2412.06769. URL <https://doi.org/10.48550/arXiv.2412.06769>.
- Zhongyang Li, Ziyue Li, and Tianyi Zhou. C3po: Critical-layer, core-expert, collaborative pathway optimization for test-time expert re-mixing. *arXiv preprint arXiv:2504.07964*, 2025.
- Yuanbo Tang, Yan Tang, Naifan Zhang, Meixuan Chen, and Yang Li. Unveiling hidden collaboration within mixture-of-experts in large language models. *arXiv preprint arXiv:2504.12359*, 2025.

## A Experiment Setup

### A.1 Baselines

We evaluate our cognitive experts in comparison with two widely used inference-time techniques for reasoning tasks: prompt engineering and decoding constraint. In particular, we consider two types of prompt placements in our analysis — one positioned before the `<think>` token (Promptbefore) and the other placed after it (Promptafter), defined as follows:

#### Prompt before `<think>`

```
<|begin_of_sentence|><|User|> <context>
You are an expert math-solving assistant who prioritizes clear, concise solutions. You solve
problems in a single thought process, ensuring accuracy and efficiency. You seek clarification
when needed and respect user preferences even if they are unconventional.
</context>

<solving_rules>
- Try to complete every idea you think of and don't give up halfway
- Don't skip steps
- Display solution process clearly
- Ask for clarification on ambiguity
</solving_rules>

<format_rules>
- Use equations and explanations for clarity
- Keep responses brief but complete
- Provide step-by-step reasoning if needed
</format_rules>

PROBLEM: {problem}

OUTPUT: Please think carefully and follow above rules to get the correct answer for
PROBLEM. Focus on clear, concise solutions while maintaining a helpful, accurate
style.<|Assistant|> <think> \n
```

Prompt after <think>
<pre> &lt; begin_of_sentence &gt;&lt; User &gt; &lt;context&gt; You are an expert math-solving assistant who prioritizes clear, concise solutions. You solve problems in a single thought process, ensuring accuracy and efficiency. You seek clarification when needed and respect user preferences even if they are unconventional. &lt;/context&gt;  PROBLEM: {problem}  &lt;think&gt; \n  Please think carefully and follow these rules to find the correct answer for PROBLEM.  &lt;solving_rules&gt; - Try to complete every idea you think of and don't give up halfway - Don't skip steps - Display solution process clearly - Ask for clarification on ambiguity &lt;/solving_rules&gt;  &lt;format_rules&gt; - Use equations and explanations for clarity - Keep responses brief but complete - Provide step-by-step reasoning if needed &lt;/format_rules&gt;  Focus on clear, concise solutions while maintaining a helpful and accurate style.  OUTPUT: </pre>

## A.2 Experiments Compute Resources

We conduct our DeepSeek-R1 experiments on 16 H100 GPUs using `vllm==0.7.0`. It is worth noting that for experiments on the Qwen3-235B-A22B model, we use `vllm==0.8.5.post` because the recently released Qwen3-235B-A22B models are only compatible with `vllm` versions  $\geq 0.8.5$ .

## B Experiment details and Results

### B.1 Cognitive Experts of Qwen3-235B

We employ Qwen3-235B to generate responses on the AIME2024 dataset, while recording the expert assignments at each token during the forward propagation. Subsequently, we apply the nPMI measure defined in Eq. 6 to identify the top five experts that exhibit the highest statistical dependence on reasoning-related indicators, such as the “<think>” token. These selected experts are thus regarded as the core cognitive components specialized in mathematical reasoning. Table 8 demonstrates the cognitive experts across math, physics, chemistry, and biology domain.

### B.2 Pass@k performance of cognitive experts

Table 9 presents the Pass@k performance of our cognitive expert modulation approach compared to vanilla baselines across two model architectures. On DeepSeek-R1, our method demonstrates consistent improvements in Pass@8 accuracy (+0.9% on AIME24 and +1.6% on AIME25) despite showing marginal variations in Pass@1 performance. For Qwen3-235B-A22B, our approach achieves

Table 8: Identified cognitive experts of Qwen3-235B. Each entry (layer ID, expert ID) denotes the Qwen3-235B model layer ID and expert ID. “All” combines data from all domains.

Domain	Identified Experts				
	Top-1	Top-2	Top-3	Top-4	Top-5
Math	(39, 182)	(29, 126)	(14, 114)	(27, 45)	(16, 129)
Physics	(2, 28)	(74, 65)	(4, 44)	(25, 103)	(7, 36)
Chemistry	(32, 58)	(26, 30)	(68, 35)	(37, 57)	(25, 103)
Biology	(2, 28)	(26, 30)	(67, 15)	(82, 29)	(25, 103)
All	(25, 103)	(26, 30)	(82, 29)	(67, 15)	(37, 57)

Table 9: Pass@k performance of our cognitive experts on Deepseek-R1 and Qwen3-235B-A22B. For each problem, we generated 16 responses with a temperature of 0.6 and a top p value of 0.95.

Benchmark	Method	pass@1	pass@8	#Tokens
AIME24	DeepSeek-R1	74.8	88.3	8,822
	+RICE {(39,182), (29,126)}	76.0	89.2	9,001
AIME25	DeepSeek-R1	68.5	84.7	10,875
	+RICE {(39,182), (29,126)}	67.7	86.3	11,294
AIME24	Qwen3-235B	84.0	93.0	10,946
	+RICE {(70,47), (23,115)}	85.0	91.6	10,706
AIME25	Qwen3-235B	82.7	88.3	12,546
	+RICE {(70,47), (23,115)}	82.1	89.7	12,373

higher Pass@1 accuracy (+1.0% on AIME24) while showing competitive Pass@8 performance ( $\pm 1.4\%$  on AIME25), with consistent reductions in computational cost (2.2% fewer tokens on AIME24 and 1.4% fewer on AIME25).

Under the Pass@1 metric (Guo et al., 2025), the cognitive experts identified in AIME24 exhibit limited generalization to AIME25. This may be attributed to the top-p sampling mechanism, which partially dilutes the effectiveness of our approach, leading to average performance. In contrast, our method demonstrates strong generalization under greedy sampling, improving AIME25 accuracy from 63.3% to 73.3% on DeepSeek-R1 and from 66.7% to 73.3% on Qwen3-235B (in Table 3). Thus, our approach enables correct answers in a single sampling step, eliminating the need for extensive sampling and verification, thereby enhancing sampling efficiency.

### B.3 Renormalization

We investigate the DeepSeek Mixture-of-Experts (MoE) architecture, where each token selects 8 of 256 experts, with weights normalized to sum to 1. We examine steering specific expert weights under two conditions: with and without renormalization. The effects of the steering coefficient (reinforce factor) are presented in Table 10, with generalization performance analyzed in Table 11.

Table 10 evaluates the reinforce factor’s effect on two cognitive experts. Without renormalization, accuracy peaks at 83.3% (factors 4, 32, 64, 128) but drops to 3.3% at 2048, with erratic token counts (e.g., 16,836). With renormalization, accuracy remains stable (73.3%–83.3%) across most factors, with token counts varying moderately (8,383–9,508), though it declines to 66.7% at factor 256. Renormalization thus enhances robustness at higher steering coefficients.

We evaluate the generalization performance of cognitive experts, identified using normalized Pointwise Mutual Information (nPMI) within Mixture-of-Experts (MoE)-based large reasoning models, comparing three strategies: Vanilla R1, Renormalized, and Without Renormalized



Table 10: Reinforce factor effects of two cognitive experts with/without renormalization

Reinforce Factor	wo/Renormalization		Renormalization	
	<i>Acc</i>	<i>Token</i>	<i>Acc</i>	<i>Token</i>
1 (R1)	73.3	9,291	73.3	9,291
2	70.0	9,103	80.0	8,463
4	83.3	8,145	80.0	8,383
8	73.3	9,502	70.0	8,818
16	80.0	8,493	73.3	9,133
32	83.3	8,337	83.3	8,956
64	83.3	8,317	80.0	9,508
128	83.3	9,490	73.3	9,091
256	60.0	7,986	66.7	8,719
512	46.7	6,270	80.0	8,786
1024	23.3	4,378	73.3	8,564

Table 11: Generalization capacity of two cognitive experts selected from AIME24, with or without renormalization.

Strategy	AIME25	Physics	Chemistry	Biology	Average
Vanilla R1	63.3	91.9	49.5	79.0	70.9
Renormalized	63.3	90.7	52.7	68.4	68.8
wo/Renormalized	73.3	89.5	50.4	79.0	<b>73.1</b>

(wo/Renormalized). Table 11 reports performance across AIME25, Physics, Chemistry, Biology, and their average for experts selected from AIME24.

The wo/Renormalized strategy demonstrates superior generalization, achieving an average score of 73.1, compared to 70.9 for Vanilla R1 and 68.8 for Renormalized. This 4.3-point improvement over Renormalized is driven by notable gains in AIME25 (73.3 vs. 63.3) and Biology (79.0 vs. 68.4). In Physics, Vanilla R1 (91.9) outperforms wo/Renormalized (89.5, -2.4), while in Chemistry, Renormalized (52.7) surpasses wo/Renormalized (50.4), indicating domain-specific trade-offs.