RAIN-Merging

A Gradient-Free Method to Enhance Instruction Following in Large Reasoning Models with Preserved Thinking Format

Zhehao Huang, Yuhang Liu, Baijiong Lin, Yixin Lou, Zhengbao He, Hanling Tian, Tao Li, Xiaolin Huang

Shanghai Jiao Tong University
The Hong Kong University of Science and Technology (Guangzhou)

Presenter: **Zhehao Huang**

Huang, Zhehao et al. "RAIN-Merging: A Gradient-Free Method to Enhance Instruction Following in Large Reasoning Models with Preserved Thinking Format." ICLR (2026) Oral.

Background

Query: [Question] Compute the last 8 decimal digits of $N = 7^{(7^7)}$. [Instruction 1] (i) Output exactly an 8-character hexadecimal string. [Instruction 2] (ii) Include at most 2 lines of Python code that verify the result.

Output of LRM:

<think>
"Compute the last 8 decimal digits of $N = 7^{(7^7)}$." Let's compute: 7^7 = 823543. So exponent e = 823543. .....
</think>
We wish to compute the last 8 decimal digits of $N = 7^{(7^7)}$. Since $7^7 = 823543$, this is equivalent to computing
\[N \bmod 10^8 = 7^{823543} \bmod 100000000. \]
This can be efficiently done using modular exponentiation. Final Answer is \boxed{70132343}.
Verification with Python Code:

```
n = pow(7, 7**7, 10**8)
print(n)
# last 8 decimal digits above
```

User Experience:

- ✓ [Thinking]
- ✓ [Correct Answer]
- ✗ [Not output in hex!]
- ✗ [More than 2 lines!]

Large reasoning models (LRMs) perform poorly in the Agentic task.

The lack of instruction-following hinders their applications.

Continual SFT or RL? Too expensive...

Background

Large Reasoning Model (LRM)

think | response

Good at reasoning
Often neglects format requirements
e.g. DeepSeek-R1-Distilled-Qwen

Merge!

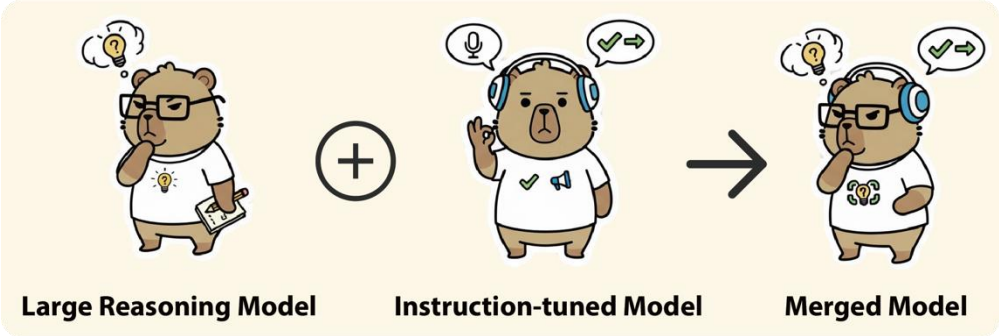
Instruction-tuned Model (ITM)

response only

Not good at thinking chains
Excellent in following instructions
e.g. Qwen2.5-Instruct

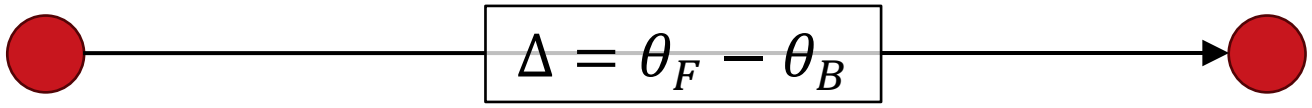
Key Problem

Can we merge an ITM into an LRM to improve **instruction-following capability** while maintaining **reasoning performance** and **structured thinking output**?



Task Vector

Task Vector[1]: Parameter difference reflecting the task capabilities



Base Model θ_B Fine-tuned Model θ_F

[1] Ilharco, Gabriel et al. "Editing Models with Task Arithmetic." *ICLR* (2022).

Observations

Observation 1: Orthogonality of parameter subspace

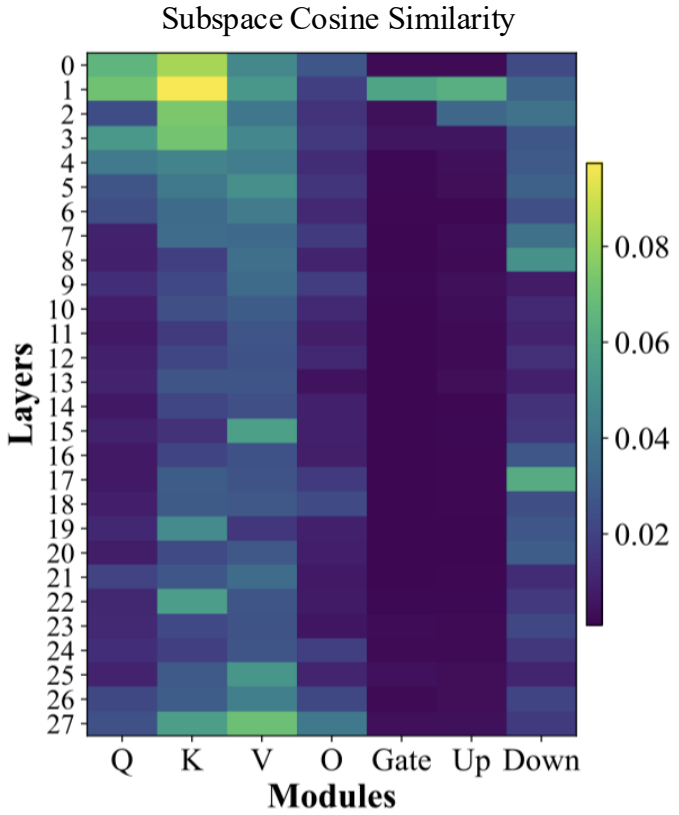
LRM θ_R
$\Delta_R = \theta_R - \theta_B$

Orthogonal subspace

Low coupling[2] of
Reason & Instruction following

ITM θ_I
$\Delta_I = \theta_I - \theta_B$

worth direct merging



Observation 2: Direct merging breaks <think>...</think> format

Task Arithmetic
$\theta_R + \lambda\Delta_I$

[2] Ortiz-Jiménez, Guillermo et al. "Task Arithmetic in the Tangent Space: Improved Editing of Pre-Trained Models." *NIPS* (2023).

RAIN-Merging

Reasoning-aware Null-space Projection (Stage 1)

$$\theta_\star = \theta_R + \lambda \bigoplus_{k=1}^{K} \alpha_\star^k \Delta_I^{\perp,k}$$

Instruction-attention Guided Merging Coefficients (Stage 2)

Key Target

① **Maintain the thinking format distribution of LRM**

② **Enhance the effectiveness of instruction following**

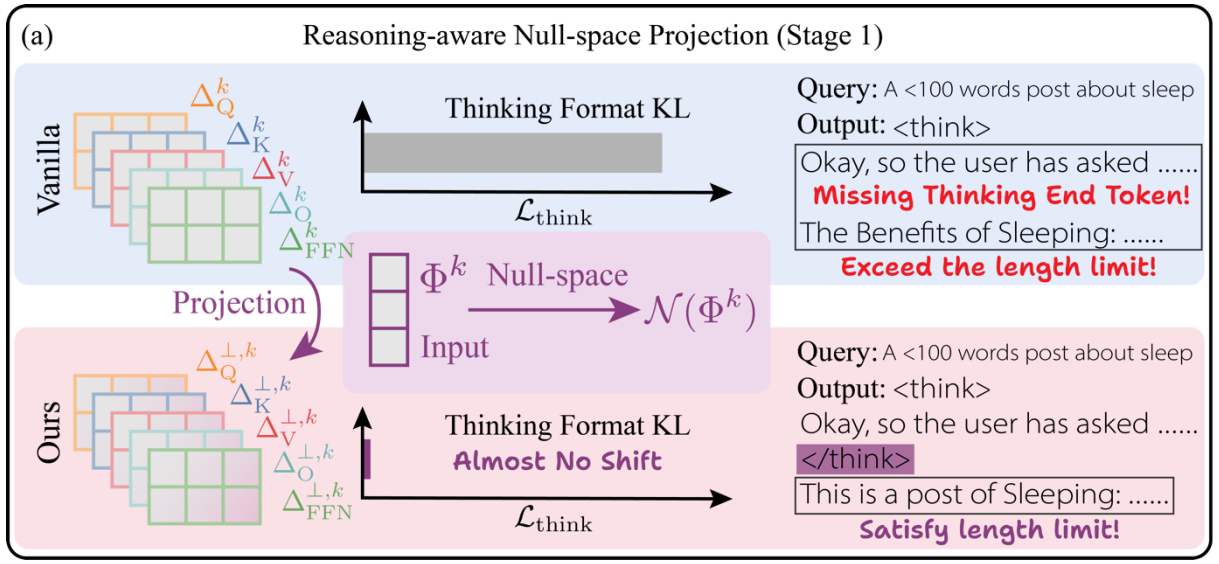
Stage 1: Reasoning-aware Null-space Projection

Mathematical form

The models before and after merging have **consistent outputs on thinking tokens**

$$\Phi_{\Omega^k_{\text{think}}}\ \text{vec}\left(\theta^k_R + \Delta^{\perp,k}_I\right) = \Phi_{\Omega^k_{\text{think}}}\ \text{vec}\left(\theta^k_R\right)$$

$$\Rightarrow \Phi_{\Omega^k_{\text{think}}}\ \text{vec}\left(\Delta^{\perp,k}_I\right) = 0$$



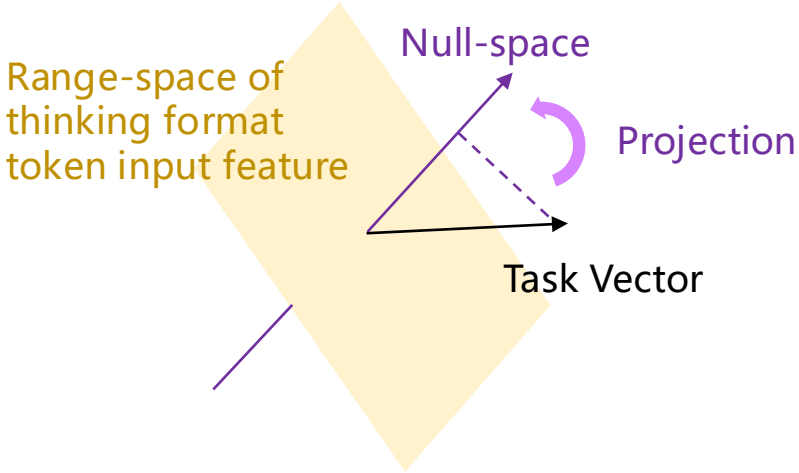
Null-space Projection

Construct **Null-space projection operator** for input features of thinking tokens

$$P^{\perp}(\Phi^k_{\Omega_{\text{think}}}) = \text{diag}(1) - {\Phi^k_{\Omega_{\text{think}}}}^{\top}\left(\Phi^k_{\Omega_{\text{think}}}{\Phi^k_{\Omega_{\text{think}}}}^{\top}\right)^{+}\Phi^k_{\Omega_{\text{think}}}$$

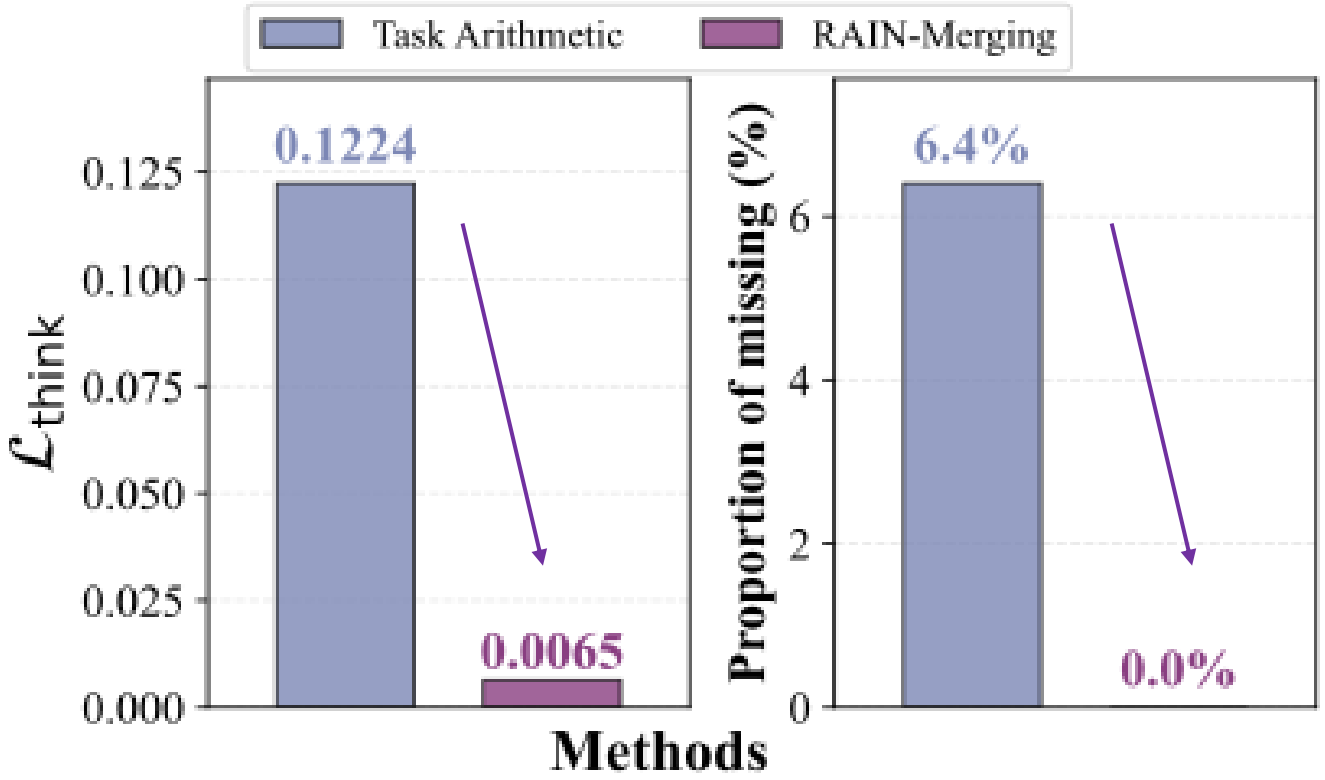
Project the ITM Task Vector into thinking token input features Null-space

$$\text{vec}\left(\Delta^{\perp,k}_I\right) = P^{\perp}(\Phi^k_{\Omega_{\text{think}}})\ \text{vec}\left(\Delta^k_I\right) \Rightarrow \Phi_{\Omega_{\text{think}}}\ \text{vec}\left(\Delta^{\perp}_I\right) = 0$$



Stage 1 Verification

Low KL divergence **Complete thinking format**



Stage 2: Instruction-attention Guided Merging Coefficients

Reparameterize: Break through the limitations of **treating all modules equally**

$$\theta(\alpha) = \theta_R + \bigoplus_{k=1}^{K} \alpha^k \Delta_I^{\perp,k}$$

Do not disrupt the **Null-space property**

Optimize merging coefficients to enhance instruction following

Stage 2: Instruction-attention Guided Merging Coefficients

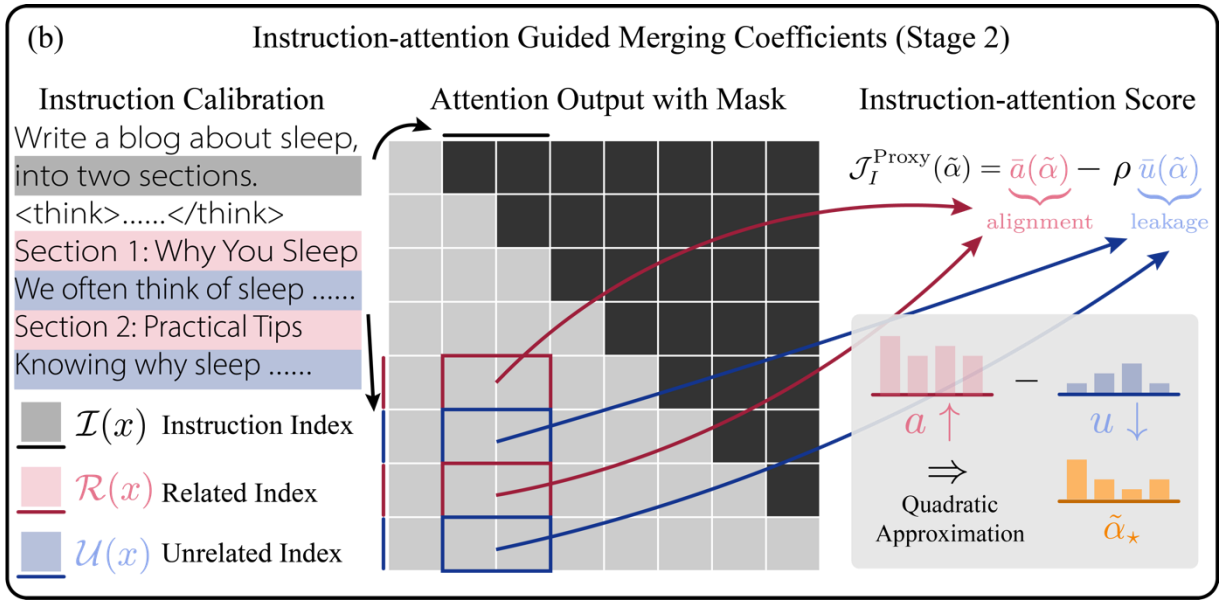
Instruction-attention Score

Motivation: [3] proposes directly **modulate attention** to enhance instruction following
And attention can reflect the **importance of the model layer** to the instruction following.

Constructing **instruction calibration set**:
Mark the instruction spans in Prompt;
the spans related to instructions;
the spans unrelated to instructions in Response

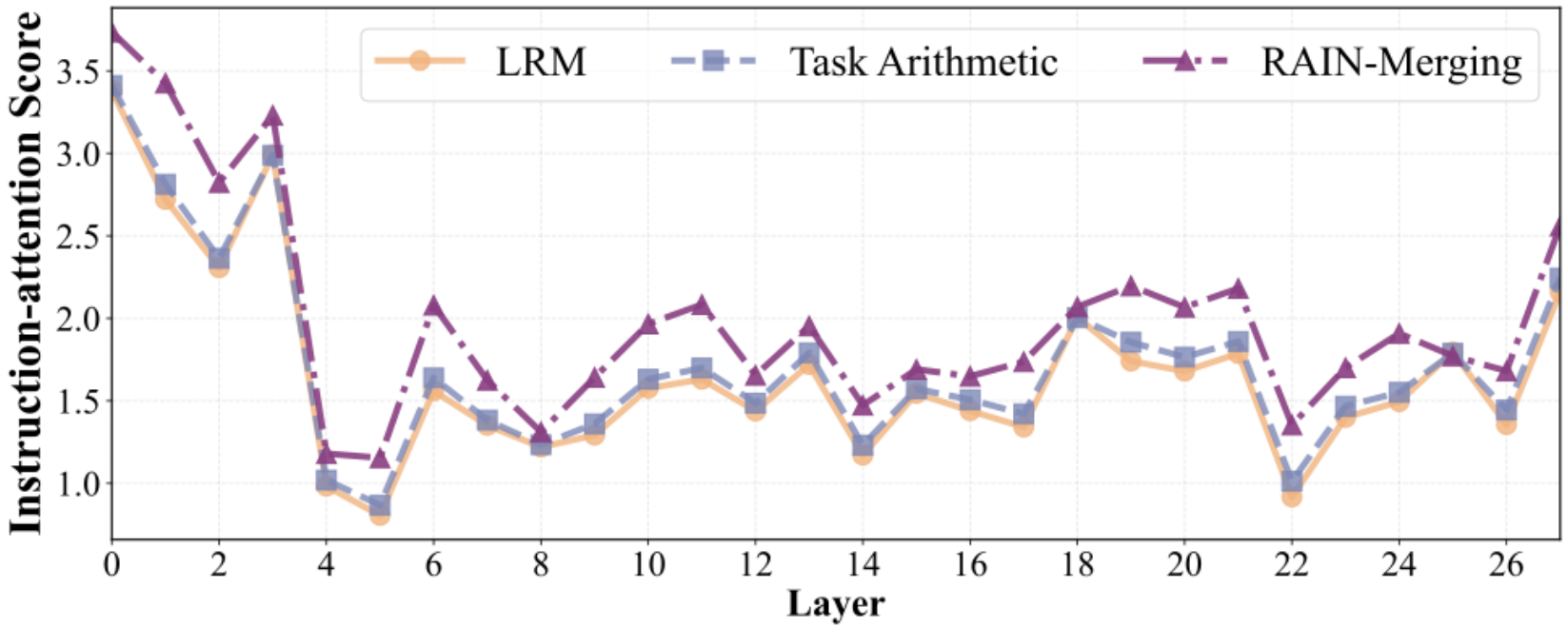
Score:
Focus on the attn of instruction-related spans
suppress the attn of instruction-unrelated spans

Calculate the merging coefficients



[3] Guardieiro, Vitoria et al. "Instruction Following by Boosting Attention of Large Language Models." *ArXiv* (2025).

Stage 2 Verification



The merged model after coefficient scaling shows significantly improved **Instruction-attention Score** at each layer.

RAIN-Merging

Reasoning-aware Null-space Projection (Stage 1)

$$\theta_\star = \theta_R + \lambda \bigoplus_{k=1}^{K} \alpha_\star^k \Delta_I^{\perp,k}$$

Instruction-attention Guided Merging Coefficients (Stage 2)

① **Enhance instruction following**

② **Maintain reasoning performance and thinking format**

③ **Gradient-free, friendly to low computational resources**

Experiment

Method	Instruction Following					Reasoning & General					RT
	IFEval	CELLO	Info Bench	Complex Bench	Avg.	Math	GPQA	Aider	Arena-Hard-v2	Avg.	
ITM	**70.43**	**19.15**	**78.49**	**43.63**	**52.92**	47.27	29.80	**33.33**	62.86	43.32	–
LRM	55.45	16.59	71.73	32.72	44.12	64.75	44.44	29.63	65.29	51.03	–
SFT	62.48	17.11	68.58	32.15	45.08	62.57	41.92	28.89	64.67	49.51	120.32
Data-free Merging											
Task Arithmetic	60.44	16.97	73.07	33.34	45.96	64.22	42.93	26.67	64.53	49.59	0.93
SLERP	58.96	17.56	72.18	34.93	45.95	63.82	42.93	31.85	65.29	50.97	1.12
Karcher	62.11	17.99	73.16	34.06	46.83	64.85	48.99	30.37	**66.13**	52.58	1.20
TIES	58.60	18.48	73.91	34.40	46.35	65.44	46.46	32.59	63.47	51.99	1.18
DARE-TIES	60.81	17.88	73.33	33.49	46.38	64.26	47.98	29.63	64.17	51.51	2.21
Data-dependent Merging											
ACM-TIES	59.33	16.45	72.44	33.75	45.50	64.57	45.96	32.59	62.00	51.28	12.45
LEWIS-TIES	60.44	17.41	72.67	34.40	46.23	62.07	48.99	31.11	64.80	51.74	16.60
AIM-TIES	62.78	17.93	73.11	34.28	47.02	65.92	49.49	**33.33**	63.64	53.10	18.51
RAIN-Merging	63.22	19.03	74.53	35.66	48.11	**68.75**	**54.55**	**33.33**	65.73	**55.59**	20.96

Model	Instruction Following					Reasoning & General				
	IFEval	CELLO	Info Bench	Complex Bench	Avg.	Math	GPQA	Aider	Arena-Hard-v2	Avg.
Qwen2.5-1.5B-Instruct	36.78	19.04	64.76	27.83	37.10	31.77	25.76	16.30	38.45	28.07
DeepSeek-R1-Distill-Qwen-1.5B	39.00	16.03	55.29	21.54	32.97	41.62	29.29	14.07	39.73	31.18
Qwen2.5-1.5B-RAIN-Merging	41.59	16.51	58.18	23.62	34.97	45.87	33.33	14.81	40.93	33.74
(relative gain)	+6.64%	+2.98%	+5.23%	+9.63%	+6.09%	+10.21%	+13.79%	+5.26%	+3.02%	+8.20%
Llama-3.1-8B-Instruct	68.58	27.21	78.67	38.47	53.23	35.59	25.25	34.07	72.23	41.79
DeepSeek-R1-Distill-Llama-8B	58.41	17.78	73.33	38.38	46.97	60.21	38.38	27.41	71.93	49.48
Llama-3.1-8B-RAIN-Merging	63.77	18.84	77.38	38.93	49.73	61.95	43.94	30.37	77.07	53.33
(relative gain)	+9.18%	+5.99%	+5.52%	+1.42%	+5.86%	+2.89%	+14.47%	+10.81%	+7.15%	+7.78%
Qwen2.5-14B-Instruct	79.85	20.13	83.38	44.19	56.89	52.73	36.87	37.04	74.40	50.29
DeepSeek-R1-Distill-Qwen-14B	71.35	18.71	81.33	40.68	53.02	72.31	57.07	33.33	80.67	60.85
Qwen2.5-14B-RAIN-Merging	76.71	19.57	84.13	44.63	56.26	74.58	57.58	40.00	86.25	64.60
(relative gain)	+7.51%	+4.58%	+3.44%	+9.69%	+6.11%	+3.13%	+0.88%	+20.00%	+6.92%	+6.17%
Qwen2.5-32B-Instruct	78.56	18.59	84.40	46.91	57.11	52.35	36.87	57.78	81.90	57.22
DeepSeek-R1-Distill-Qwen-32B	76.52	19.69	83.56	44.44	56.05	68.00	60.10	54.81	82.00	66.23
Qwen2.5-32B-RAIN-Merging	77.26	19.96	84.76	45.74	56.93	75.67	61.62	54.07	83.70	68.77
(relative gain)	+0.97%	+1.39%	+1.44%	+2.93%	+1.57%	+11.28%	+2.52%	-1.35%	+2.07%	+3.83%

Effectiveness

Compared with previous merging methods

RAIN-Merging Enhance **instruction-following**

Maintain or even promote **reasoning & general**

Efficiency

Computing resources are far lower than SFT

Scalability

RAIN-Merging on range of **1.5B ~ 32B**

On **Qwen2.5 & Llama-3.1** architectures

Enhance **instruction-following**

Maintain or even promote **reasoning & general**

Experiment

Agentic Scenario

RAIN-Merging enhance **agentic comprehensive** of LRM

Model	ALFWorld	WebShop
ITM	17.50	10.45
LRM	22.00	26.63
RAIN-Merging	25.00	29.42

Reasoning & Instruction-following Coupling Scenario

RAIN-Merging enhance both accuracy rate (**Both Acc.**) of simultaneously meeting **reasoning** and **instruction following**.

Model	IF Acc.	Math Acc.	Both Acc.
ITM	48.81	40.95	19.76
LRM	25.86	53.81	12.62
RAIN-Merging	35.10	54.76	20.48
(relative gain)	+35.73%	+1.77%	+62.26%

Exploration of Chain-of-thought Quality

Maintain **reasoning internal consistency (RIC)**

Improve **reasoning responses alignment (RAA)**

Reduce the "**knowing-doing gap**"[4]

Model	IFEval		AIME25		GPQA		Avg.
	RIC	RAA	RIC	RAA	RIC	RAA	
LRM	4.58	4.41	4.50	3.60	3.53	3.76	4.06
RAIN-Merging	4.61	4.51	4.50	4.10	3.56	4.26	4.26
(relative gain)	+0.77%	+2.26%	0.00%	+13.89%	+0.86%	+13.31%	+4.78%

[4] Schmied, Thomas et al. "LLMs are Greedy Agents: Effects of RL Fine-tuning on Decision-Making Abilities." *ArXiv* (2025).

Case Study

Reasoning Example 2 in GPQA

DeepSeek-R1-Distill-Qwen-7B Response (Answer: A, Wrong)

The problem involves a bound system of two nucleons (NN) in a 1S0 state emitting a particle X with intrinsic parity -1. The isospin, angular momentum, and spin quantum numbers of the final NN state satisfy the condition **T(NN) = S(NN) + L(NN) + 1 mod 2**, where T(NN) = 0 for the final NN state. The task is to determine which partial wave (l) of X is not permitted, considering both angular momentum and parity conservation.

...

(We omit the lengthy reasoning process for clarity.)

...

Option-by-Option Analysis:

- **(A) 1S0 -> 3D3 + f**
  - NN term symbol **3D3**: **S = 1, L = 2, J = 3**.
  - **T(NN) = (1 + 2 + 1) mod 2 = 4 mod 2 = 0** (satisfies the isospin condition).
  - **Parity of NN state**: **$(-1)\hat{L} + S = (-1)\hat{2} + 1 = -1$** (satisfies the parity condition).
  - **Conclusion**: This partial wave is **allowed**.
- **(B) 1S0 -> 3S1 + p**
  - NN term symbol **3S1**: **S = 1, L = 0, J = 1**.
  - **T(NN) = (1 + 0 + 1) mod 2 = 2 mod 2 = 0** (satisfies the isospin condition).
  - **Parity of NN state**: **$(-1)\hat{L} + S = (-1)\hat{0} + 1 = -1$** (satisfies the parity condition).
  - **Conclusion**: This partial wave is **allowed**.
- **(C) 1S0 -> 3P0 + s**
  - NN term symbol **3P0**: **S = 1, L = 1, J = 0**.
  - **T(NN) = (1 + 1 + 1) mod 2 = 3 mod 2 = 1** (does not satisfy the isospin condition).
  - **Parity of NN state**: **$(-1)\hat{L} + S = (-1)\hat{1} + 1 = +1$** (does not satisfy the parity condition).
  - **Conclusion**: This partial wave is **not allowed**.
- **(D) 1S0 -> 7D1 + p**
  - NN term symbol **7D1**: **S = 3, L = 2, J = 1**.
  - **T(NN) = (3 + 2 + 1) mod 2 = 6 mod 2 = 0** (satisfies the isospin condition).
  - **Parity of NN state**: **$(-1)\hat{L} + S = (-1)\hat{2} + 3 = -1$** (satisfies the parity condition).
  - **Conclusion**: This partial wave is **allowed**.

Final Answer:

Options **(A)** and **(C)** are not permitted because they fail either the isospin or parity condition. However, since the question asks for the single best answer, **(A)** is chosen as the answer because it is the first invalid option encountered in the analysis.

ANSWER: A

Reasoning Example 2 in GPQA

RAIN-Merging Response (Answer: C, Correct)

To determine which partial wave is not permitted, we analyze the problem using the given conditions:

1. **Initial State**: The initial NN system is in the 1S0 state, which has total spin $S = 0$, orbital angular momentum $L = 0$, and isospin $T = 0$. The parity of the initial state is $+1$ since $P = (-1)^L = (-1)^0 = +1$.

2. **Final State Conditions**:

- The final NN state has $T = 0$, so the isospin condition is $T = S + L + 1 \mod 2$. Substituting $T = 0$, we get $S + L + 1 \equiv 0 \mod 2$, which simplifies to $S + L \equiv 1 \mod 2$. This means S and L must have opposite parity (one even, one odd).
- The parity of the final state must conserve parity. The emitted particle X has intrinsic parity -1, so the parity of the final NN state is $P_{\text{NN}} = P_{\text{NN}}^{\text{initial}} \cdot P_X = (+1) \cdot (-1) = -1$. Since $P_{\text{NN}} = (-1)^L$, this implies $(-1)^L = -1$, so L must be odd.

3. **Analysis of Each Option**:

- **(A) 1S0 $\to$ 3D3 + f**: The NN state is $3D3$, so $S = 1$ (since $2S + 1 = 3$) and $L = 2$.
  - $S + L = 1 + 2 = 3$, which satisfies $S + L \equiv 1 \mod 2$.
  - $L = 2$ is even, which violates the parity condition L must be odd.
  - **Allowed.**
- **(B) 1S0 $\to$ 3S1 + p**: The NN state is $3S1$, so $S = 1$ and $L = 0$.
  - $S + L = 1 + 0 = 1$, which satisfies $S + L \equiv 1 \mod 2$.
  - $L = 0$ is even, which violates the parity condition L must be odd.
  - **Allowed.**
- **(C) 1S0 $\to$ 3P0 + s**: The NN state is $3P0$, so $S = 1$ and $L = 1$.
  - $S + L = 1 + 1 = 2$, which does **not** satisfy $S + L \equiv 1 \pmod 2$.
  - $L = 1$ is odd, which satisfies the parity condition.
  - **Not allowed.**
- **(D) 1S0 $\to$ 7D1 + p**: The NN state is $7D1$, so $S = 3$ (since $2S + 1 = 7$) and $L = 2$.
  - $S + L = 3 + 2 = 5$, which satisfies $S + L \equiv 1 \mod 2$.
  - $L = 2$ is even, which violates the parity condition L must be odd.
  - **Allowed.**

4. **Conclusion**:

All options except **(C)** satisfy the parity condition L must be odd. However, the isospin condition $S + L \equiv 1 \mod 2$ is violated in **(C)**, making it the only permitted partial wave that is not allowed.

ANSWER: C

Acknowledge

ICLR Team & Reviewers

Institute of Image Processing and Pattern Recognition in SJTU

The Hong Kong University of Science and Technology (Guangzhou)

Prof. Xiaolin Huang

Poster Today 3:15 PM
Pavilion 3, P3-#1904