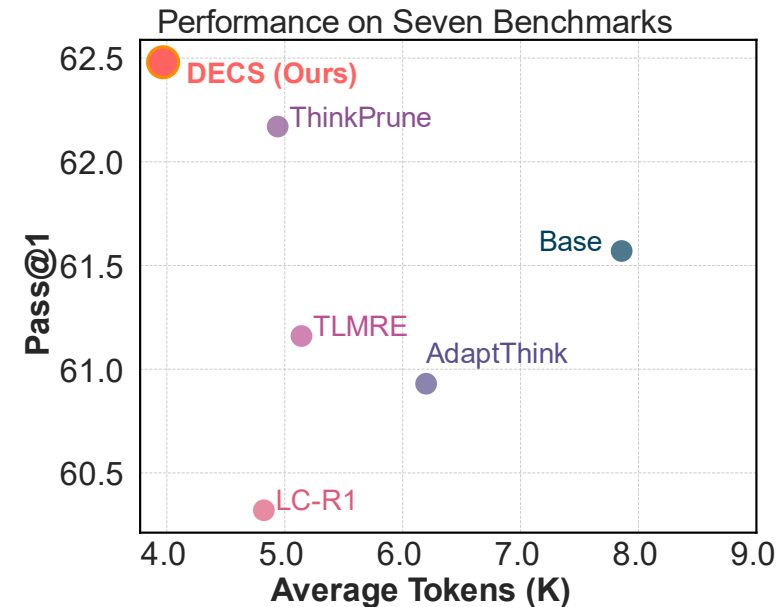


Overthinking Reduction with Decoupled Rewards and Curriculum Scheduling



Shuyang Jiang*, Yusheng Liao, Ya Zhang, Yanfeng Wang, Yu Wang✉

DECS reduces overthinking by over 50% while improving reasoning performance, as a free lunch.



Problem Setting

- ▶ **Overthinking:** generating excessively long reasoning paths without performance benefit.

Define $p = \sum_{k=1}^{\infty} \frac{1}{k^2}$ and $q = \sum_{k=1}^{\infty} \frac{1}{k^3}$. Find a way to write $\sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \frac{1}{(j+k)^3}$

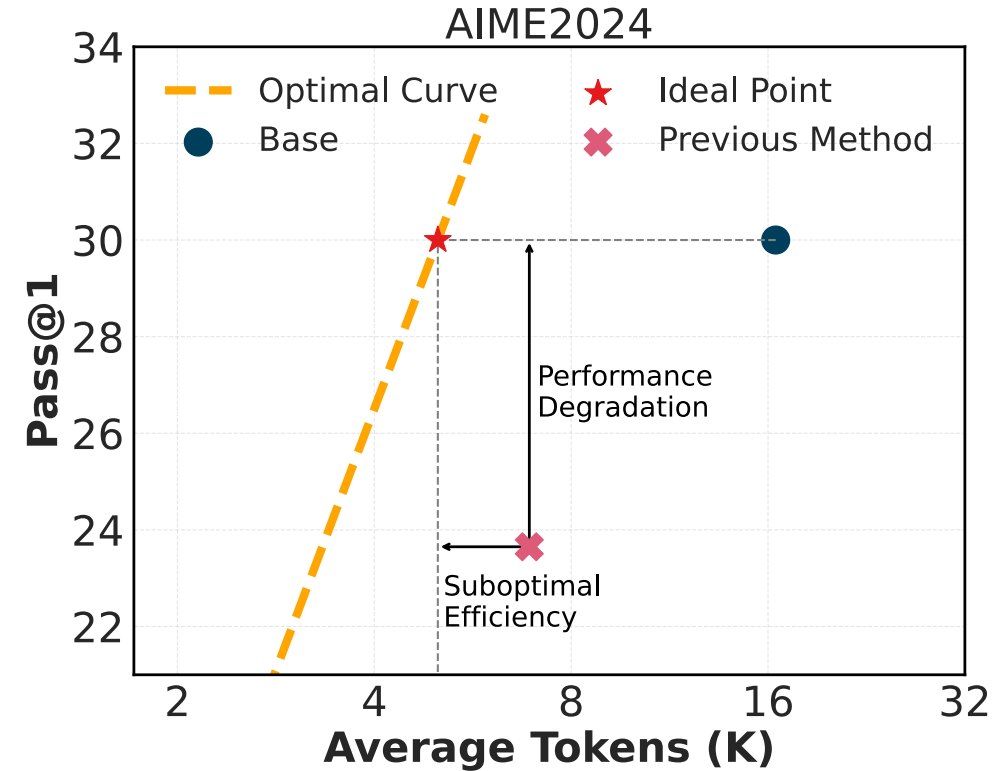
Necessary Thinking: 3283 Tokens

Okay, so I need to find a way to express the double sum $\sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \frac{1}{(j+k)^3}$ in terms of (p) and (q) , where $(p = \sum_{k=1}^{\infty} \frac{1}{k^2})$ and $(q = \sum_{k=1}^{\infty} \frac{1}{k^3})$. Hmm, let me think about how to approach this. First, I know that (p) is the well-known Basel problem, which sums to $(\frac{\pi^2}{6})$, but I don't think I need the exact value here. Similarly, (q) is the sum of reciprocals of cubes, which converges to Apéry's constant, approximately 1.202..., but again, exact value isn't necessary here. The double sum given is over all positive integers (j) and (k) , and the term inside is $(\frac{1}{(j+k)^3})$. So, it's a double series where each term depends on the sum of two variables. This seems a bit tricky because it's a two-dimensional sum, unlike single sums which are straightforward. Maybe I can change variables or find a way to express this double sum in terms of (p) and (q) . Let me consider switching the order of summation or using some symmetry. Another idea: perhaps express $(\frac{1}{(j+k)^3})$ as an integral. Since $(\frac{1}{n^3})$ can be written as an integral, maybe I can use that to interchange the sum and the integral. Recall that for positive integers (n) , we have $(\frac{1}{n^3} = \frac{1}{2} \int_0^1 \int_0^1 (xy)^{n-1} dx dy)$. Wait, is that correct? Let me check. Actually, [omitted...] So, that gives us $(p - q)$.

Overthinking: 492 Tokens

Hmm, that seems too straightforward. Wait, but is that correct? [omitted...] Which is equal to $(\sum_{m=1}^{\infty} \frac{1}{m^2} - 1) - (\sum_{m=1}^{\infty} \frac{1}{m^3} - 1) = (p - 1) - (q - 1) = p - q$. So, is that correct? Hmm. Let me verify. [omitted...] which is indeed $(p - q)$. Therefore, the double sum is equal to $(p - q)$. So, the answer is $(\boxed{p - q})$. **Final Answer** $(\boxed{p - q})$

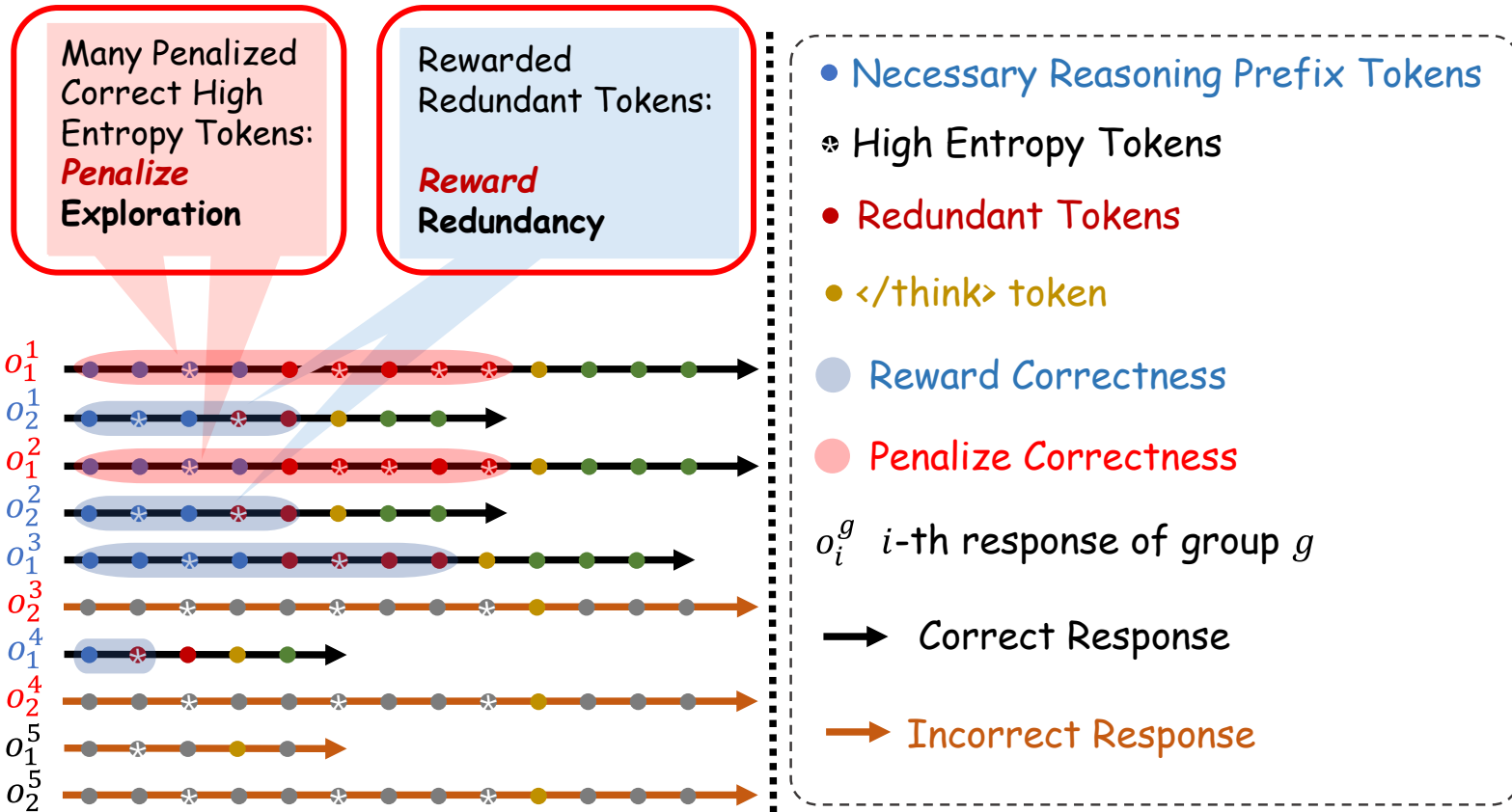
- ▶ **Goal and Challenge**



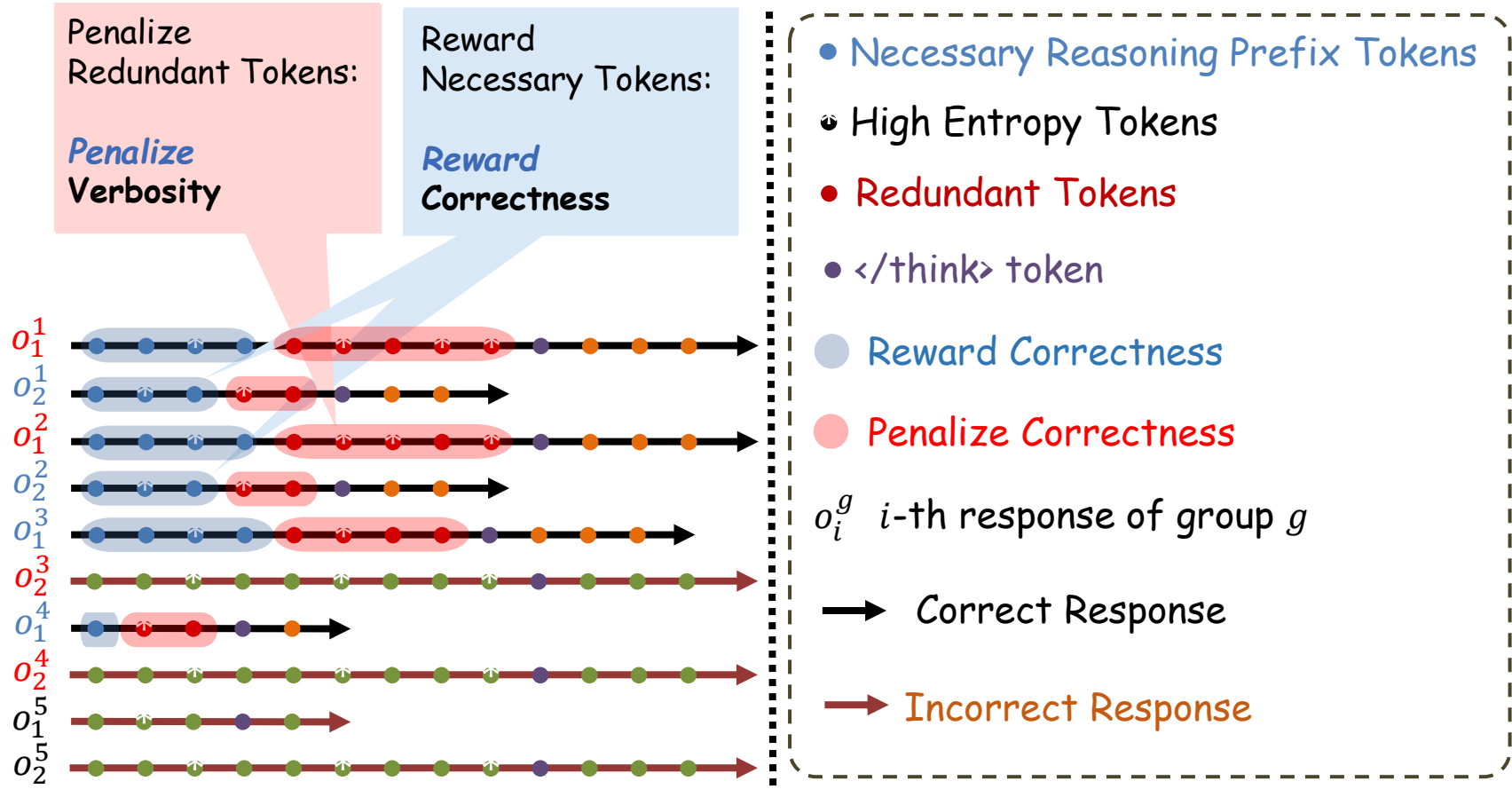
Why not just use a length penalty?

- ▶ Sequence-level length rewards are misaligned with token-level optimization.

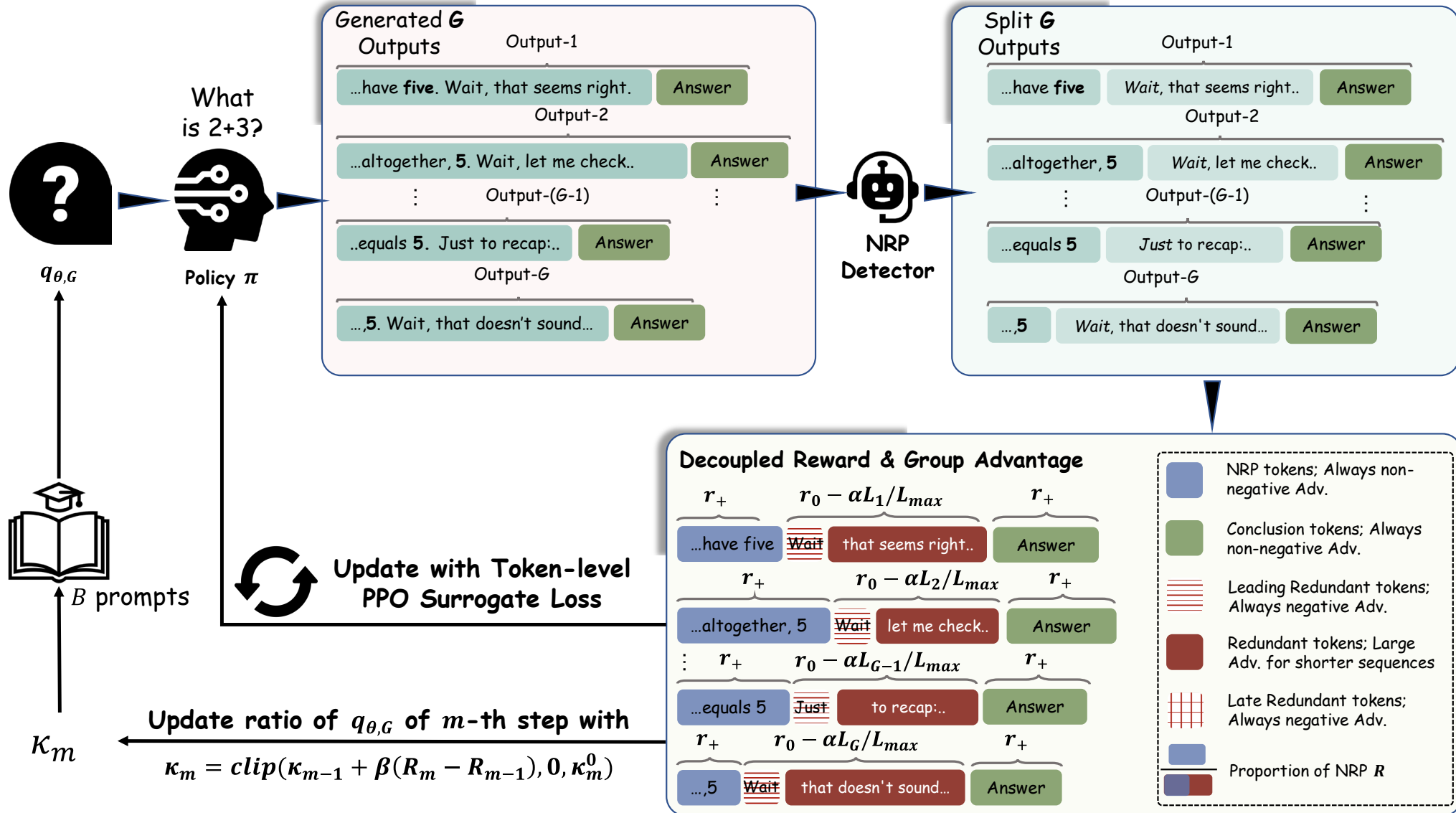
$$r'(o) = \begin{cases} r(o) - \gamma|o| & o \text{ is correct} \\ r(o) & \text{otherwise} \end{cases}$$



Where is the solution?



Method overview — DECS



Method detail 1 — NRP Detector

- ▶ Necessary Reasoning Prefix (NRP) is defined as the shortest reasoning prefix that first justifies the correct answer.

Define $p = \sum_{k=1}^{\infty} \frac{1}{k^2}$ and $q = \sum_{k=1}^{\infty} \frac{1}{k^3}$. Find a way to write $\sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \frac{1}{(j+k)^3}$

NRP: 3283 Tokens

Okay, so I need to find a way to express the double sum $\sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \frac{1}{(j+k)^3}$ in terms of (p) and (q) , where $(p = \sum_{k=1}^{\infty} \frac{1}{k^2})$ and $(q = \sum_{k=1}^{\infty} \frac{1}{k^3})$. Hmm, let me think about how to approach this. First, I know that (p) is the well-known Basel problem, which sums to $(\frac{\pi^2}{6})$, but I don't think I need the exact value here. Similarly, (q) is the sum of reciprocals of cubes, which converges to Apéry's constant, approximately 1.202..., but again, exact value isn't necessary here. The double sum given is over all positive integers (j) and (k) , and the term inside is $(\frac{1}{(j+k)^3})$. So, it's a double series where each term depends on the sum of two variables. This seems a bit tricky because it's a two-dimensional sum, unlike single sums which are straightforward. Maybe I can change variables or find a way to express this double sum in terms of (p) and (q) . Let me consider switching the order of summation or using some symmetry. Another idea: perhaps express $(1/(j+k)^3)$ as an integral. Since $(1/n^3)$ can be written as an integral, maybe I can use that to interchange the sum and the integral. Recall that for positive integers (n) , we have: $(\frac{1}{n^3} = \frac{1}{2} \int_0^1 \int_0^1 (xy)^{n-1} dx dy)$. Wait, is that correct? Let me check. Actually, [omitted...] So, that gives us $(p - q)$.

Overthinking: 492 Tokens

Hmm, that seems too straightforward. Wait, but is that correct? [omitted...] Which is equal to: $(\left(\sum_{m=1}^{\infty} \frac{1}{m^2} - 1 \right) - \left(\sum_{m=1}^{\infty} \frac{1}{m^3} - 1 \right) = (p - 1) - (q - 1) = p - q)$. So, is that correct? Hmm. Let me verify. [omitted...] which is indeed $(p - q)$. Therefore, the double sum is equal to $(p - q)$. So, the answer is $(\boxed{p - q})$. **Final Answer** $(\boxed{p - q})$

Method detail 1 — NRP Detector

- ▶ Goal: Lightweight, Precise, and Implementation-friendly
- ▶ Solution:
 - ▶ Instead of identifying necessary reasoning prefix directly, label whether a reasoning chunk contains a ground truth answer.
 - ▶ Instead of directly using a strong model, choose a small language model and distill the detection ability from a strong teacher

Method detail 1 — NRP Detector Pretraining

1. Preparing training data

- ▶ Curated 5000 cases from OpenR1-Math-220K
- ▶ Splitting reasoning trajectories into chunks using predefined separators: Wait, But, Alternatively, Hmm, However, Maybe

Necessary Thinking: 3283 Tokens

Okay, so I need to find a way to express the double sum $\sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \frac{1}{(j+k)^3}$ in terms of (p) and (q) , where $(p) = \sum_{k=1}^{\infty} \frac{1}{k^2}$ and $(q) = \sum_{k=1}^{\infty} \frac{1}{k^3}$. Hmm, let me think about how to approach this.

First, I know that (p) is the well-known Basel problem, which sums to $(\frac{\pi^2}{6})$, but I don't think I need the exact value here. Similarly, (q) is the sum of reciprocals of cubes, which converges to Apéry's constant, approximately 1.202..., but again, exact value isn't necessary here.

The double sum given is over all positive integers (j) and (k) , and the term inside is $(\frac{1}{(j+k)^3})$. So, it's a double series where each term depends on the sum of two variables. This seems a bit tricky because it's a two-dimensional sum, unlike single sums which are straightforward.

Maybe I can change variables or find a way to express this double sum in terms of (p) and (q) . Let me consider switching the order of summation or using some symmetry.

Another idea: perhaps express $(\frac{1}{(j+k)^3})$ as an integral. Since $(\frac{1}{n^3})$ can be written as an integral, maybe I can use that to interchange the sum and the integral.

Recall that for positive integers (n) , we have: $(\frac{1}{n^3}) = \int_0^1 \int_0^1 (xy)^{n-1} dx dy$

Wait, is that correct? Let me check. Actually, [omitted...] So, that gives us $(p - q)$.

CNK1: Okay, so I need to find a way to express the double sum $\sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \frac{1}{(j+k)^3}$ in terms of (p) and (q) , where $(p) = \sum_{k=1}^{\infty} \frac{1}{k^2}$ and $(q) = \sum_{k=1}^{\infty} \frac{1}{k^3}$.

CNK2: *Hmm*, let me think about how to approach this.

First, I know that (p) is the well-known Basel problem, which sums to $(\frac{\pi^2}{6})$, but I don't think I need the exact value here. Similarly, (q) is the sum of reciprocals of cubes, which converges to Apéry's constant, approximately 1.202...,

CNK3: *but* again, exact value isn't necessary here.

The double sum given is over all positive integers (j) and (k) , and the term inside is $(\frac{1}{(j+k)^3})$. So, it's a double series where each term depends on the sum of two variables. This seems a bit tricky because it's a two-dimensional sum, unlike single sums which are straightforward.

CNK4: *Maybe* I can change variables or find a way to express this double sum in terms of (p) and (q) . Let me consider switching the order of summation or using some symmetry.

Another idea: perhaps express $(\frac{1}{(j+k)^3})$ as an integral. Since $(\frac{1}{n^3})$ can be written as an integral, maybe I can use that to interchange the sum and the integral.

Recall that for positive integers (n) , we have: $(\frac{1}{n^3}) = \int_0^1 \int_0^1 (xy)^{n-1} dx dy$

CNK5: *Wait*, is that correct? Let me check. Actually, [omitted...] So, that gives us $(p - q)$.

Method detail 1 — NRP Detector Pretraining

1. Preparing training data

- ▶ Query a Large model (Qwen2.5-72B) to obtain ground truth
- ▶ Use SFT to distill the answer to the target model (Qwen2.5-1.5B-Instruct)

NRP Detector Prompt

```
Given a math problem and a segment of a long reasoning process to solve the problem, your task is to identify whether this segment has presented a correct final answer. If the segment contains information that can serve as the final answer to the problem and the answer is semantically correct when referring to the ground truth, simply explain the reason and output \boxed{yes}. Otherwise, directly output \boxed{no}.
```

```
**Problem**:
```

```
{problem}
```

```
**Reasoning segment**:
```

```
{segment}
```

```
**Ground Truth**:
```

```
{answer}
```

Method detail 1 — NRP Detector Pretraining

3. Rollout-stage Usage

- Classify each chunk and collect tokens up to the first chunk that is labeled as True

CNK-1: Okay, so I need to find a way to express the double sum $\sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \frac{1}{(j+k)^3}$ in terms of (p) and (q) , where $(p) = \sum_{k=1}^{\infty} \frac{1}{k^2}$ and $(q) = \sum_{k=1}^{\infty} \frac{1}{k^3}$.

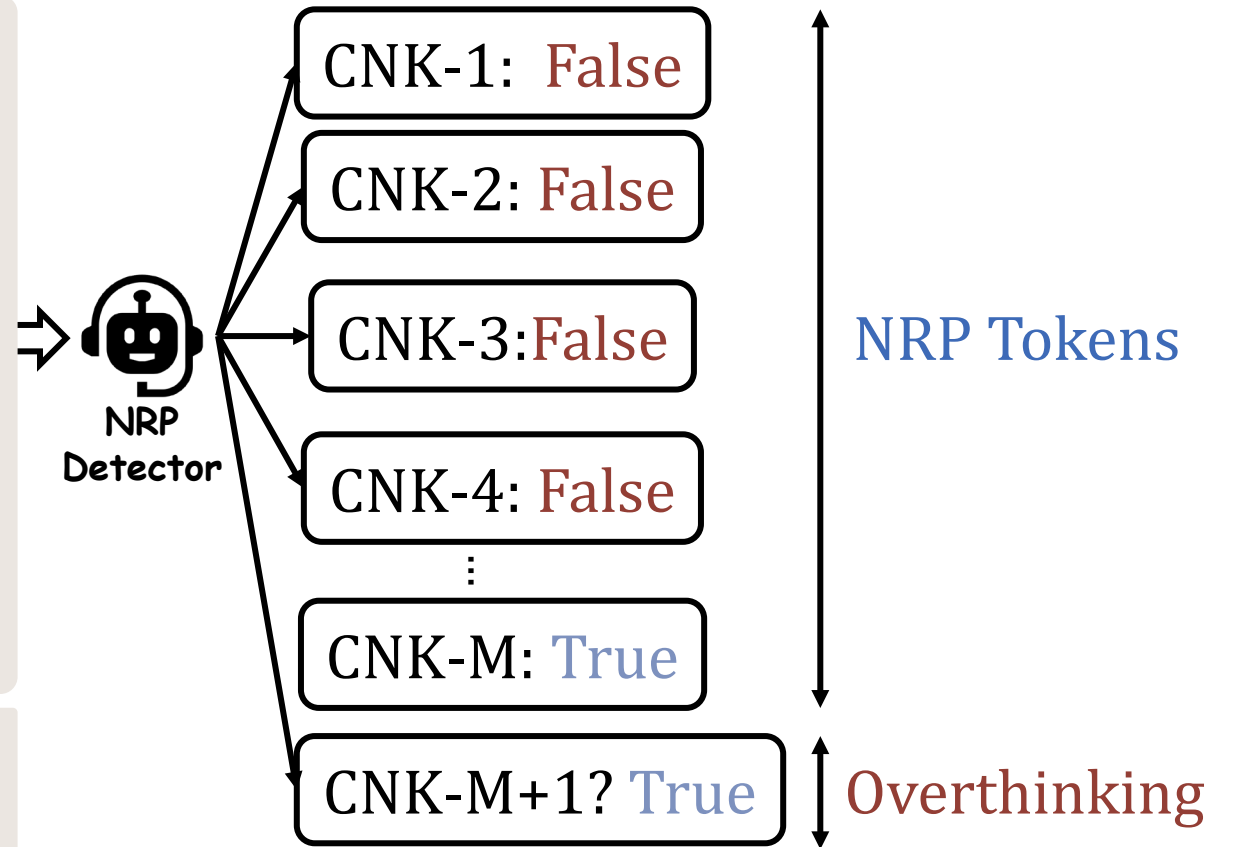
CNK-2: Hmm, let me think about how to approach this. First, I know that (p) is the well-known Basel problem, which sums to $\frac{\pi^2}{6}$, but I don't think I need the exact value here. Similarly, (q) is the sum of reciprocals of cubes, which converges to Apéry's constant, approximately 1.202..., but again, exact value isn't necessary here. The double sum given is over all positive integers (j) and (k) , and the term inside is $\frac{1}{(j+k)^3}$. So, it's a double series where each term depends on the sum of two variables. This seems a bit tricky because it's a two-dimensional sum, unlike single sums which are straightforward. Maybe I can change variables or find a way to express this double sum in terms of (p) and (q) .

CNK-3: Let me consider switching the order of summation or using some symmetry. Another idea: perhaps express $\frac{1}{(j+k)^3}$ as an integral. Since $\frac{1}{n^3}$ can be written as an integral, maybe I can use that to interchange the sum and the integral. Recall that for positive integers (n) , we have: $\frac{1}{n^3} = \frac{1}{2} \int_0^1 \int_0^1 (xy)^{n-1} dx dy$

CNK-4: Wait, is that correct? Let me check. Actually, [omitted...]

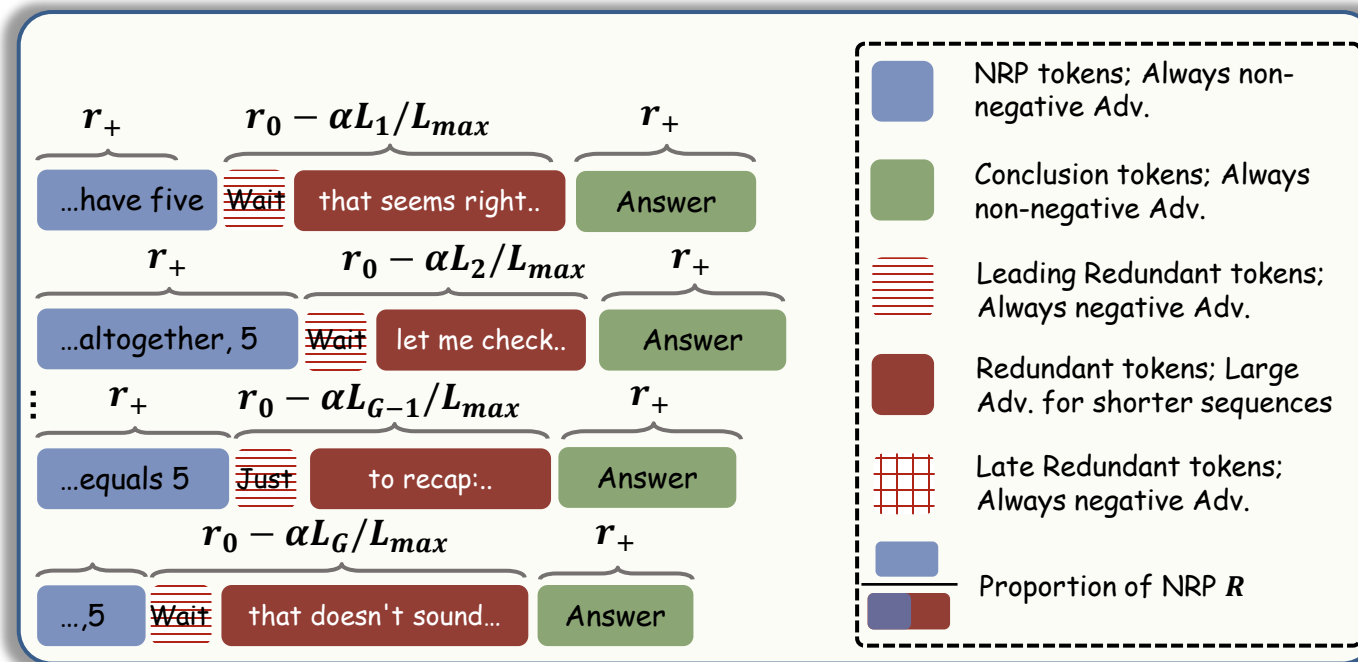
CNK-M: So, that gives us $(p - q)$.

CNK-M+1: Hmm, that seems too straightforward. Wait, but is that correct? [omitted...] Which is equal to: $\frac{1}{2} \left(\sum_{m=1}^{\infty} \frac{1}{m^2} - 1 \right) - \frac{1}{2} \left(\sum_{m=1}^{\infty} \frac{1}{m^3} - 1 \right) = (p - 1) - (q - 1) = p - q$. So, is that correct? Hmm. Let me verify. [omitted...] which is indeed $(p - q)$. Therefore, the double sum is equal to $(p - q)$. So, the answer is $(p - q)$. **Final Answer** $(p - q)$



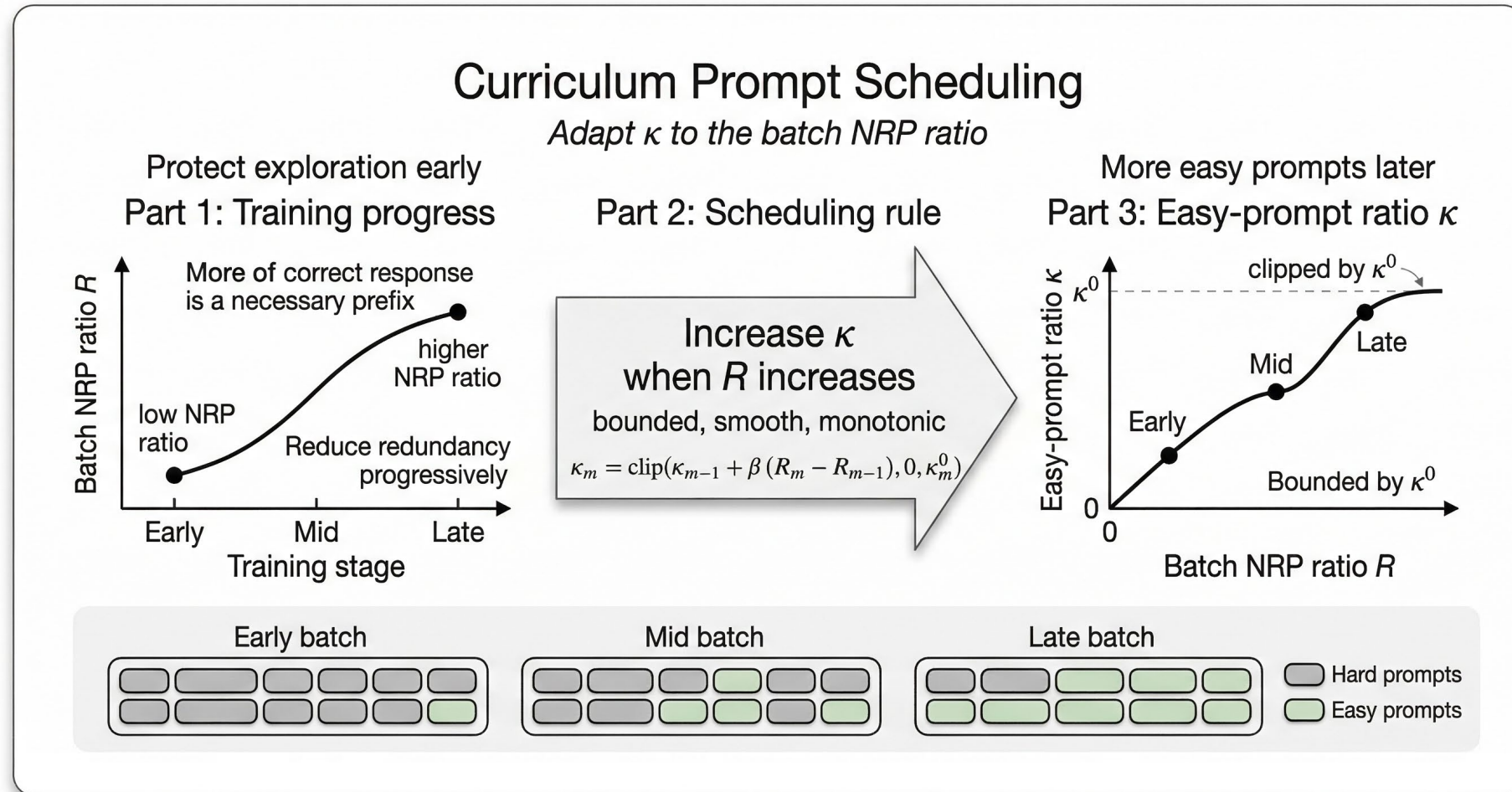
Method detail 2 — Decoupled reward

- ▶ NRP Tokens & Answer Tokens: Largest positive reward for non-negative Advantage
 - ▶ For protection of correct non-verbose reasoning logics: $r_{NRP} = 1.1$
- ▶ Overthinking Tokens: The simple linear length penalty
 - ▶ For always penalizing redundancy: $r_{overthinking} = 1 - \frac{\alpha L_i}{L_{max}}$



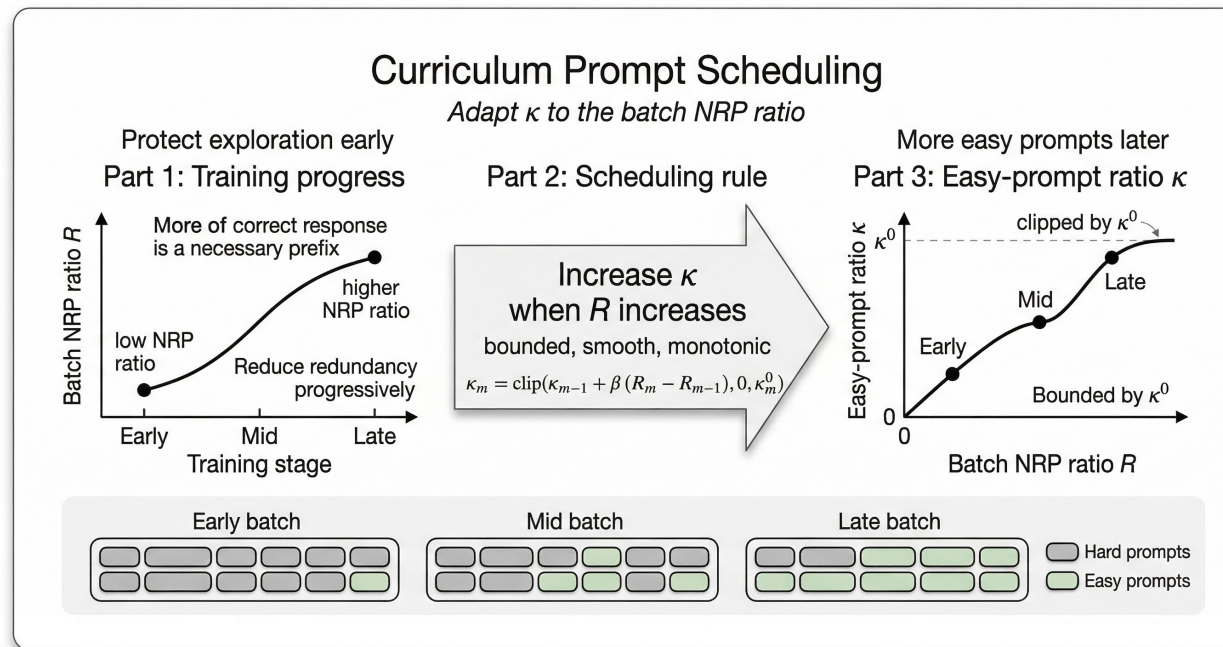
Method detail 3 — Curriculum Scheduling

Schedule easy prompts according to current redundancy level.



Method detail 3 — Curriculum Scheduling

- ▶ In the early stage, many high-entropy tokens are among the overthinking tokens.
 - ▶ Control the ratio of prompts that are optimized with DECS to avoid penalizing too many high-entropy tokens according to current redundancy level, parameterized as NRP ratio
- ▶ In the late stage, the overthinking tokens are reduced a lot
 - ▶ To speed up training progress, we allow more prompts

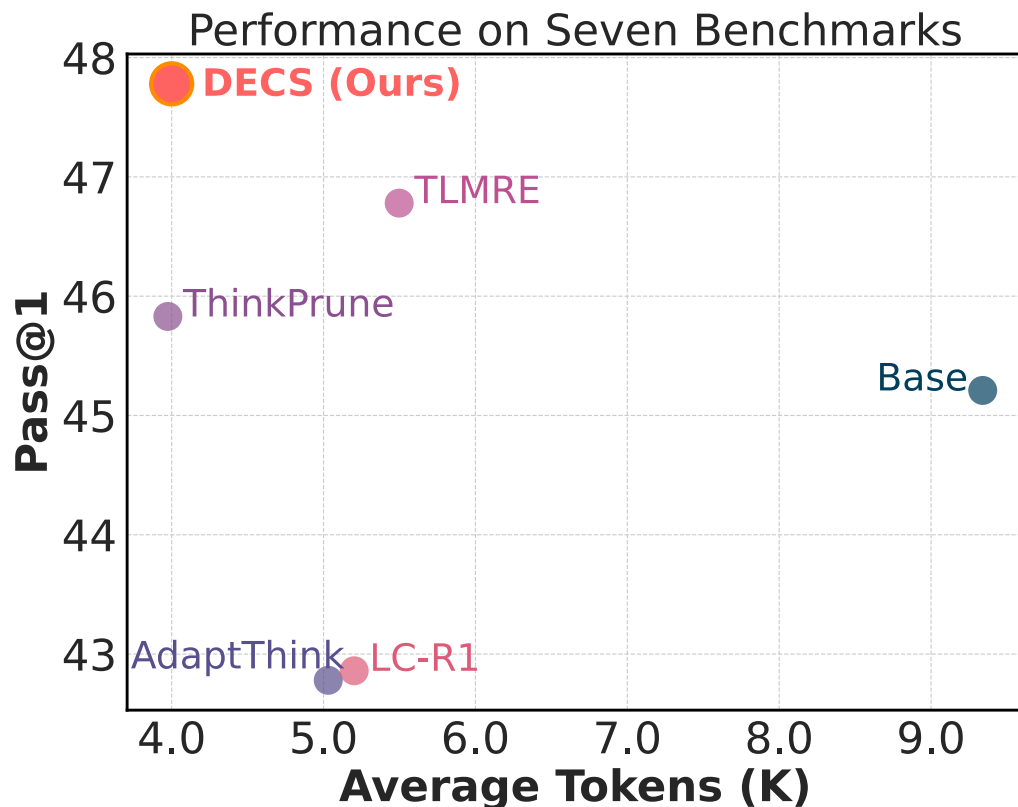


Experiments

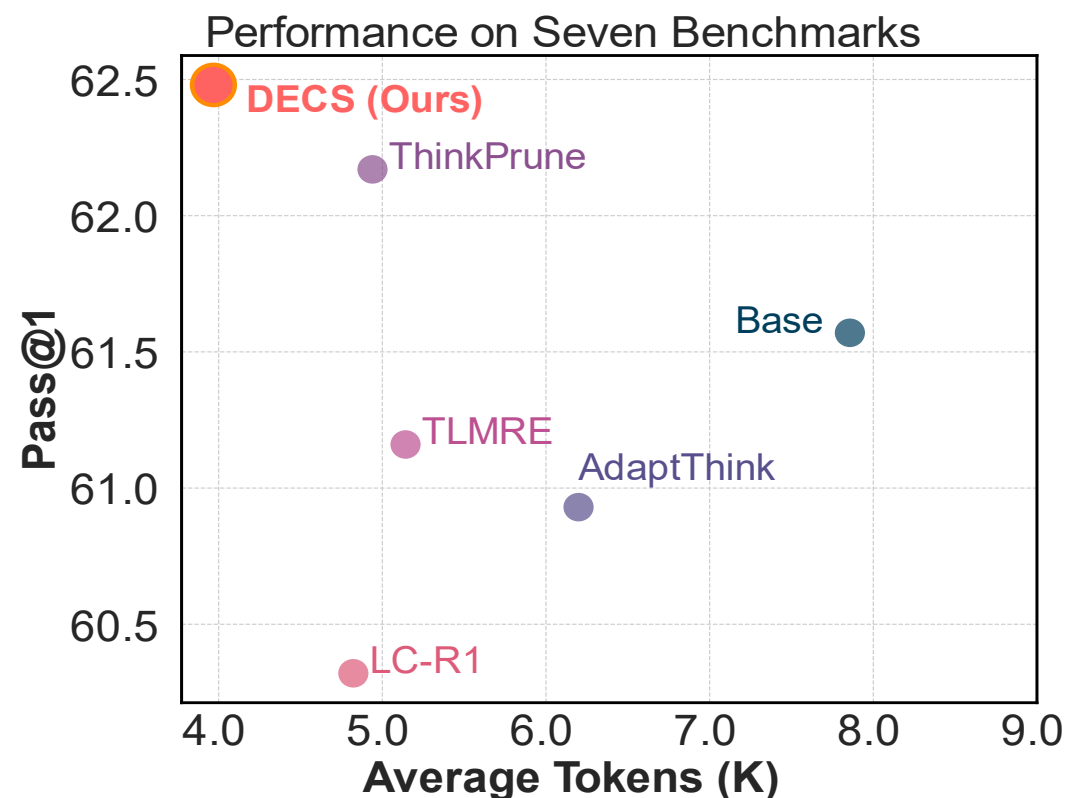
- ▶ Backbone Language models
 - ▶ DeepSeek-R1-Distilled-1.5B, 7B, Qwen3-4B
- ▶ Fine-tuning corpus
 - ▶ DeepScaleR
- ▶ Benchmark datasets
 - ▶ In-domain dataset: AIME24, AIME25, AMC23, MATH500, OlympiadBench
 - ▶ Out-of-domain dataset: GPQA-D, LiveCodeBench-v6
- ▶ Baselines:
 - ▶ ThinkPrune, TLMRE, LC-R1, AdaptThink

Experiments- DECS is optimized on a better frontier

Comparison on 1.5B

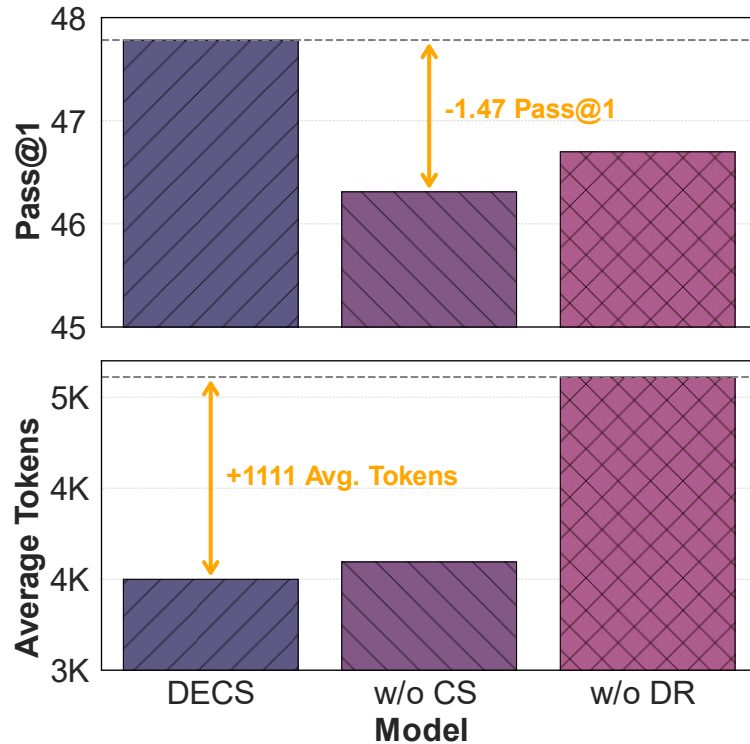


Comparison on 7B

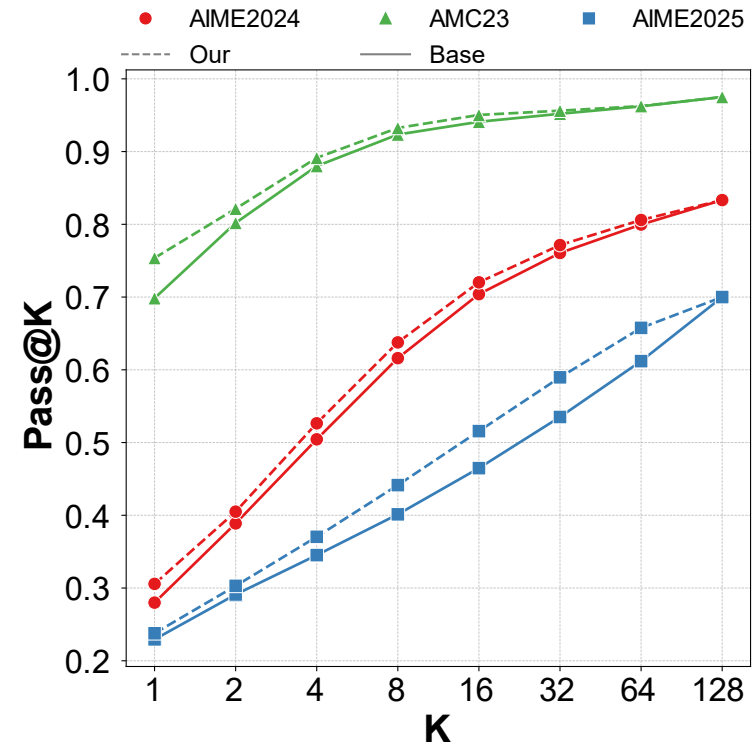


DECS reduces redundancy without hurting exploration

- ▶ Ablation: CS for correctness, DR for efficiency



- ▶ Exploration: preserving pass@K across different Ks.



Conclusion

- ▶ Efficient reasoning should penalize redundancy, not reasoning itself.
- ▶ DECS separates necessary reasoning from redundant reasoning.
- ▶ It reduces overthinking by about 50% without sacrificing capability.



Thank you!

**Paper
Project**



My Homepage



**Open-sourced
Model**



Supplementary – Full Experiment

Model	AIME2024		AIME2025		AMC23		MATH500		OlympiadB		GPQA-D		LCB		Average		AES
	Acc	#Tok.	Acc	#Tok.	Acc	#Tok.	Acc	#Tok.	Acc	#Tok.	Acc	#Tok.	Acc	#Tok.	Acc	#Tok.	
<i>DS-1.5B</i>																	
Base	27.99	12202	<u>22.94</u>	12138	69.84	7875	<u>84.55</u>	4847	53.78	9217	32.86	8540	24.53	10560	45.21	9340	0.00
+GRPO	32.76	8834	<u>25.91</u>	8431	77.09	5722	<u>87.34</u>	3577	58.73	6425	35.76	5953	26.45	8759	49.15	6814	0.53
AdaptThink	27.92	6914	21.95	7400	64.73	2644	81.57	1488	50.40	3501	25.92	4093	26.98	9181	42.78	5031	0.19
ThinkPrune	26.93	5306	20.86	4937	72.87	<u>2869</u>	84.27	1879	55.04	<u>3477</u>	<u>35.51</u>	<u>3839</u>	<u>25.36</u>	5515	45.83	3975	<u>0.62</u>
TLMRE	<u>29.87</u>	7550	22.24	7151	<u>74.51</u>	<u>3943</u>	84.86	2376	<u>56.08</u>	<u>4833</u>	<u>33.74</u>	<u>4896</u>	26.13	7737	<u>46.78</u>	5498	<u>0.52</u>
LC-R1	<u>23.65</u>	6904	19.64	6681	<u>68.69</u>	3715	82.02	2277	<u>51.57</u>	4519	30.93	5377	23.54	6940	<u>42.86</u>	5202	0.18
DECS	31.25	<u>5550</u>	23.78	<u>4965</u>	75.37	2988	84.40	<u>1817</u>	56.10	3396	35.92	3255	27.66	<u>6026</u>	47.78	<u>4000</u>	0.74
<i>DS-7B</i>																	
Base	50.65	10508	36.67	11096	88.77	5764	93.25	3654	69.22	7507	46.46	7502	45.95	8966	61.57	7857	0.00
+GRPO	52.50	9011	38.54	9670	91.88	5205	94.21	3520	72.59	6425	49.62	6101	47.71	8569	63.86	6929	0.23
AdaptThink	53.31	8884	<u>36.48</u>	9525	86.66	3675	91.06	1824	67.98	5528	43.91	5746	47.09	8209	60.93	6199	0.16
ThinkPrune	51.15	<u>6625</u>	<u>36.46</u>	<u>7127</u>	88.28	3193	<u>92.98</u>	2105	<u>70.03</u>	4154	48.42	<u>4498</u>	<u>47.90</u>	6881	<u>62.17</u>	4940	<u>0.40</u>
TLMRE	50.11	<u>7023</u>	34.24	<u>7501</u>	87.07	3329	<u>91.83</u>	2073	<u>68.84</u>	4382	<u>49.02</u>	<u>4913</u>	<u>47.03</u>	6772	<u>61.16</u>	5142	<u>0.31</u>
LC-R1	50.52	6891	32.50	7387	85.74	<u>2802</u>	90.28	1473	67.76	<u>3983</u>	48.58	4672	46.83	<u>6554</u>	60.32	<u>4823</u>	0.28
DECS	<u>51.33</u>	5277	36.43	5516	89.04	2772	92.96	1728	70.28	3283	49.27	3276	48.05	5921	62.48	3968	0.54