

Do NOT Think That Much for 2+3=? On the Overthinking of Long Reasoning Models

Xingyu Chen^{1,2}, Jiahao Xu², Tian Liang², Zhiwei He^{1,2}, Jianhui Pang, Dian Yu², Linfeng Song², Qiuzhi Liu², Mengfei Zhou¹, Zhuosheng Zhang¹, Rui Wang¹, Zhaopeng Tu², Haitao Mi², Dong Yu²

¹Shanghai Jiao Tong University

²Tencent

OVERVIEW

- **Long Reasoning Models**, such as OpenAI o1 and DeepSeek-R1, have attracted much attention for their ability to demonstrate human-like deep thinking.
- However, the reasoning process might be **inefficient**.
- We presents the first comprehensive study on the prevalent issue of **overthinking** in these models.
- **Overthinking**: The long reasoning model generates *redundant, homogeneous solutions* to a single question, and the subsequent solutions contribute less to the accuracy and diversity.

EXAMPLE

Question: *What's the answer of 2+3?*

QwQ-32B-Preview (901 tokens)

Solution1
39 tokens

2 plus 3... that's a pretty straightforward arithmetic problem. I think I can handle this. So, let's see, 2 plus 3 means I'm adding two numbers together. I know that when you add two and three, you get five.

Solution2
109 tokens

But maybe I should think about it step by step to make sure I understand the process. First, what is addition? Addition is a mathematical operation where you combine two or more numbers to get a total or sum. In this case, I'm dealing with two numbers: 2 and 3. So, starting with 2, if I add 3 to it, I'm essentially moving three steps forward on the number line. Let me visualize that. If I start at 2 and move one step to the right, I reach 3. Another step takes me to 4, and one more step takes me to 5. So, 2 plus 3 equals 5.

Solution3
39 tokens

Alternatively, I can think of it in terms of objects. Suppose I have two apples and then I get three more apples. How many apples do I have in total? Well, two apples plus three apples equals five apples.

Solution4
34 tokens

Another way to look at it is using fingers. I can hold up two fingers on one hand and three on the other, then count all the fingers together, which gives me five fingers.

Solution5
42 tokens

I can also break it down using smaller numbers. For example, 2 plus 2 equals 4, and then adding one more makes it 5. So, 2 plus 2 is 4, plus 1 is 5, which is the same as 2 plus 3.

Solution6
26 tokens

Moreover, if I consider the number properties, 2 and 3 are both positive integers, and their sum should also be a positive integer, which is 5.

Solution8
29 tokens

Furthermore, in algebra, if I have an equation like $x + 3 = 5$, then solving for x would give me $x = 2$, which aligns with 2 plus 3 equals 5.

Solution10
19 tokens

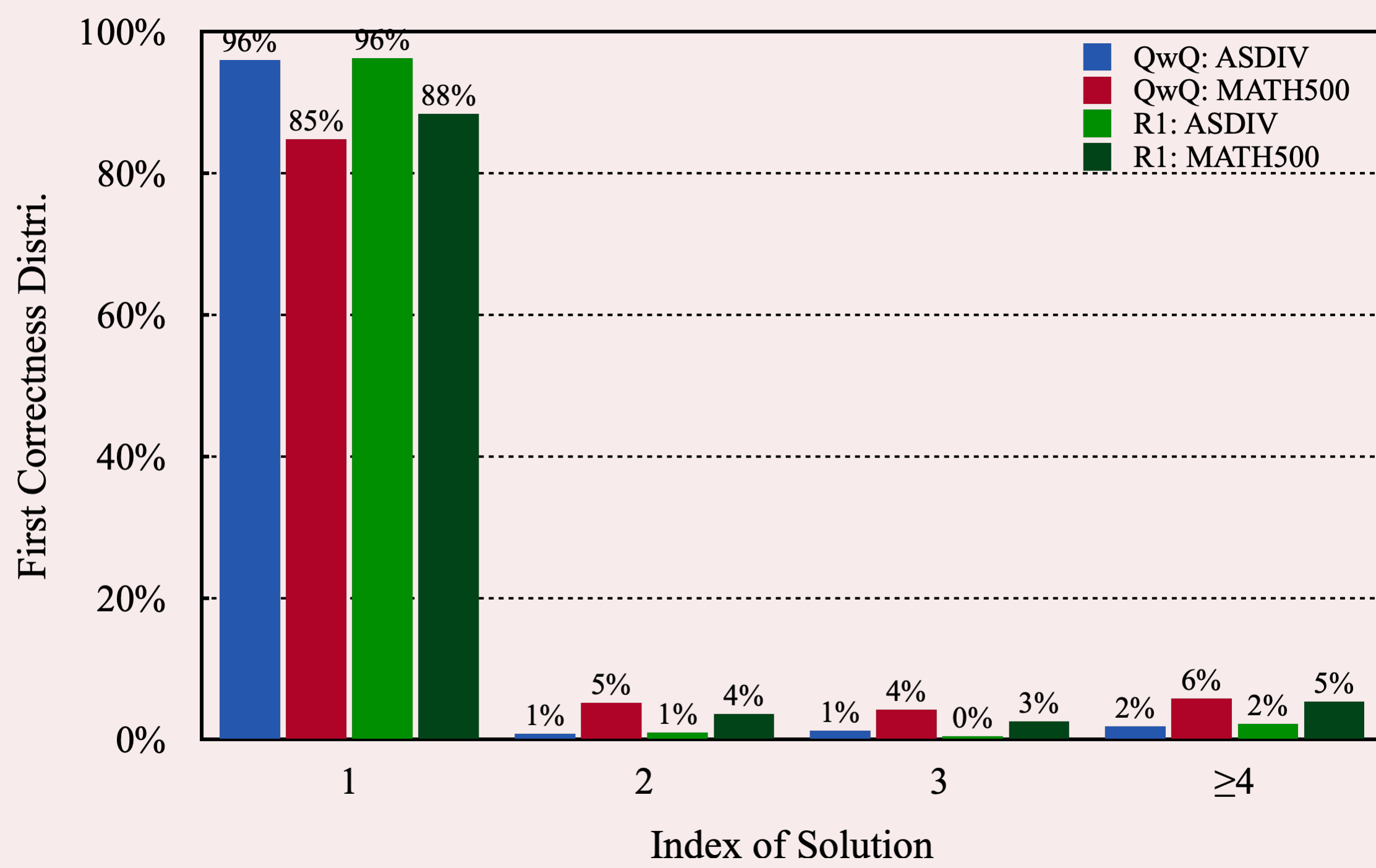
Alternatively, using Roman numerals, 2 is II and 3 is III. Adding them together gives V, which is 5.

In conclusion, the answer to 2 plus 3 is 5.

QwQ-32B-Preview generates **13 solutions!**

FINDINGS

Finding 1: Redundant solutions contribute less to the accuracy

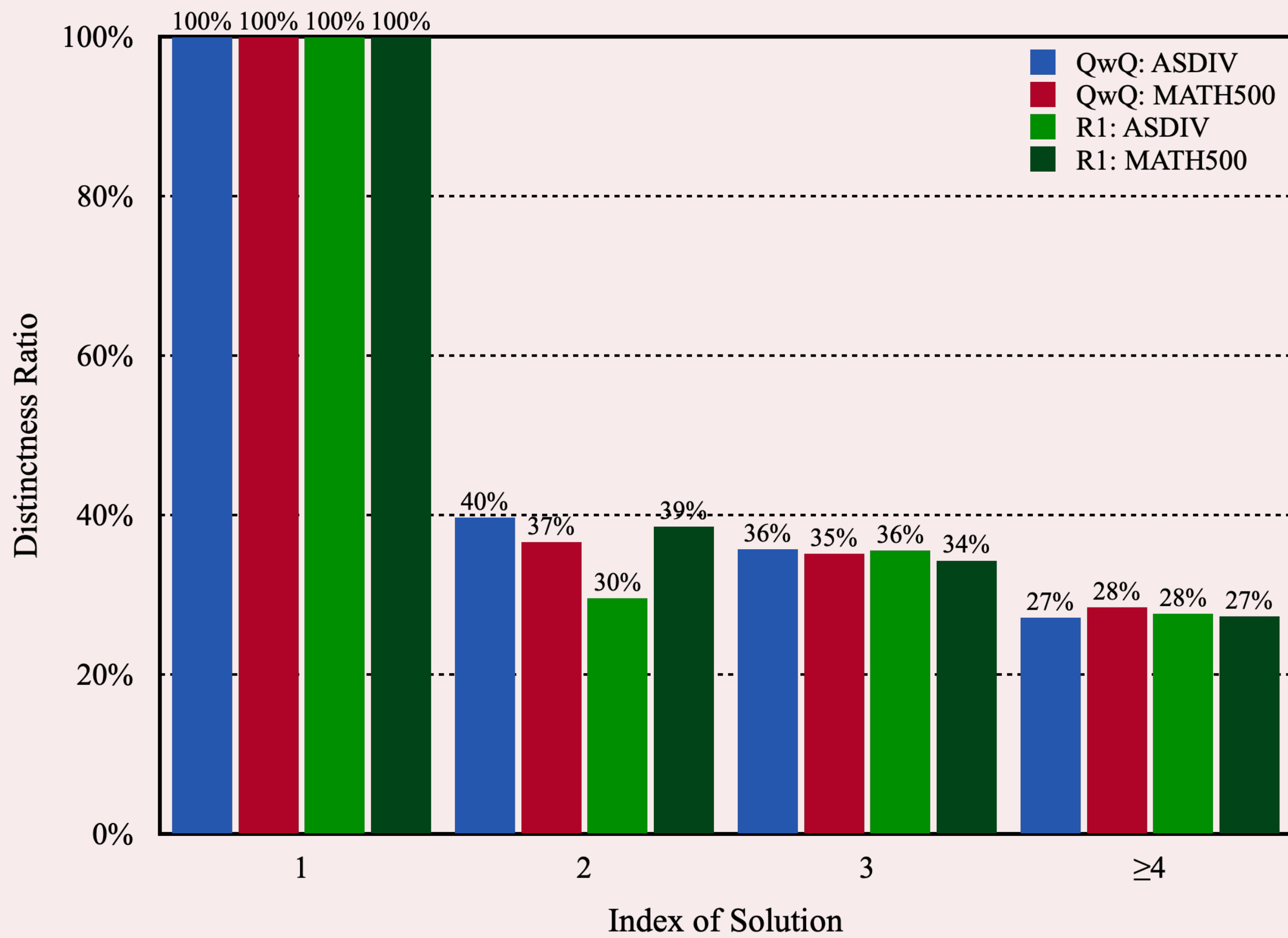


- More than 85% of the time the first answer is already correct.
- Subsequent solutions mainly verify previous solutions.

(Figure: The distribution of First Correct Solution)

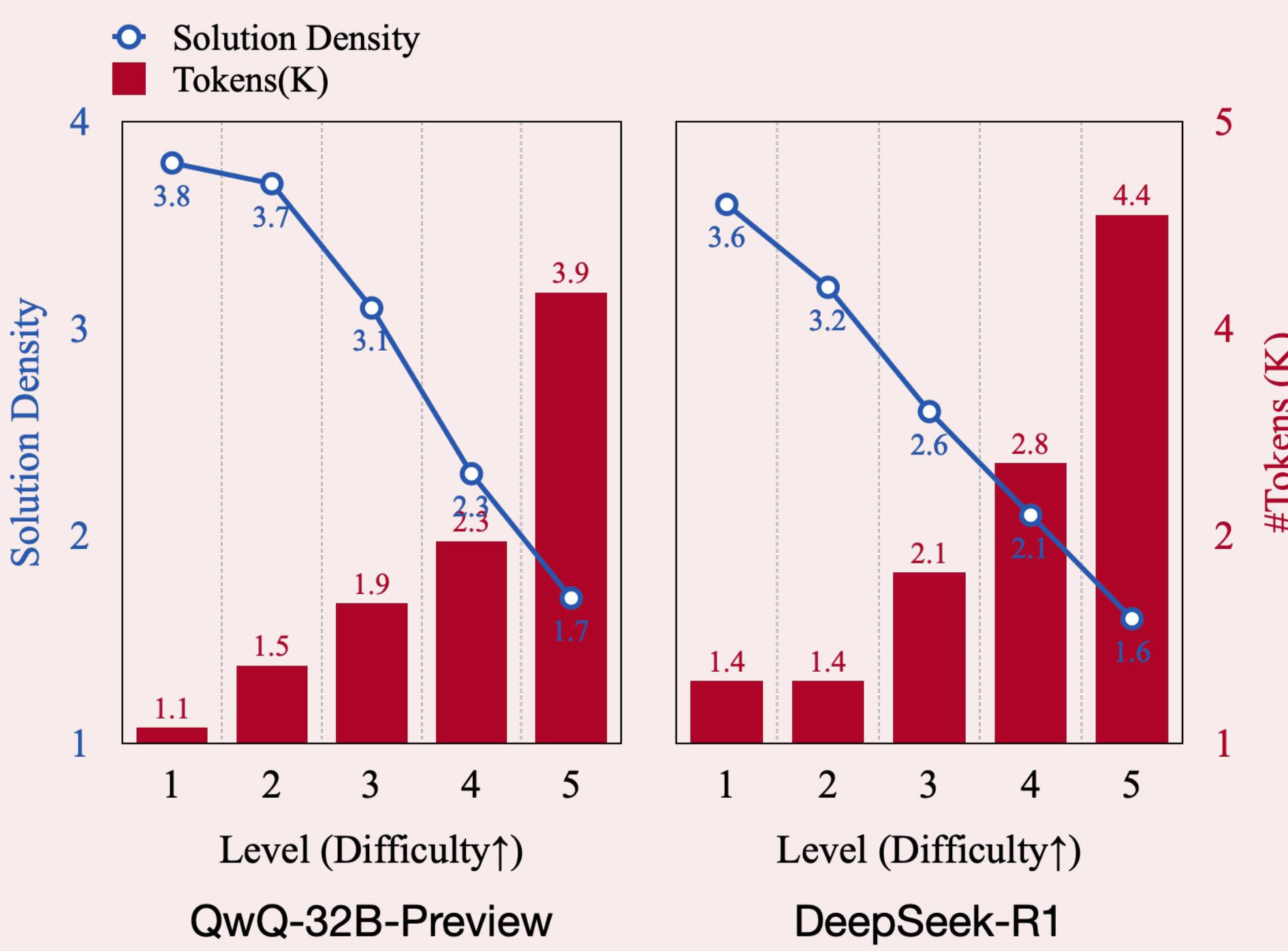
Finding 2: Redundant solutions lack diversity

- The second solution has only a 50% chance of introducing a new reasoning strategy.
- As the number of solutions increases, the possibility decreases.



(Figure: The probability that the solution of each position introduces a new reasoning strategy)

Finding 3: Overthinking is more severe in simple questions



- Solution Density: the number of solutions per 1000 tokens
- Overthinking is more prominent in simple questions

(Figure: The Solution Density in different difficult level of problems (MATH500))

METRICS

- **Outcome Efficiency:** The ratio of the tokens in *first correct solution* to the total tokens

$$\xi_o = \frac{1}{N} \sum_{i=1}^N \sigma_i \frac{\hat{T}_i}{T_i}$$

σ_i : The correctness of the i -th response.
 \hat{T}_i : The tokens in the first correct solutions
 T_i : The tokens in the full response

- **Process Efficiency:** The ratio of the tokens in different reasoning strategy to the total tokens

$$\xi_p = \frac{1}{N} \sum_{i=1}^N \frac{D_i}{T_i}$$

D_i : The tokens in the different reasoning strategies
 T_i : The tokens in the full response

METHODS

Length Preference Optimization

- **Step 1:** Sampling on training set
- **Step 2:** Streamline sampling results
- **Step 3:** Construct Preference Pair
 - Positive Example: Keep the *first correct solution* and another round of *verification (reflection)*
 - Negative Example: The longest response in sampling results

EXPERIMENT RESULTS

Methods	Accuracy	Response		Efficiency	
		#Solution	#Token	Outcome	Process
ASDIV					
QwQ-32B-Preview	96.5	3.5	713.7	53.7%	69.0%
+SimPO _{FCS} +Reflection	96.6	1.9	381.5	82.5%	87.9%
GSM8K					
QwQ-32B-Preview	94.7	2.9	756.1	11.8%	75.2%
+SimPO _{FCS} +Reflection	95.9	1.8	416.5	86.0%	91.0%
MATH500					
QwQ-32B-Preview	91.2	4.3	2398.5	51.4%	70.3%
+SFT _{Shortest Response}	92.6	4.4	2359.0	59.7%	72.8%
+DPO _{Shortest Response}	93.2	3.4	1928.8	64.3%	77.8%
+RPO _{Shortest Response}	90.2	3.5	2015.2	64.7%	76.6%
+SimPO _{Shortest Response}	91.0	3.5	1871.5	64.7%	78.1%
+SimPO _{First-Correct Solution}	90.4	1.3	1015.6	85.5%	96.3%
+SimPO _{FCS} +Reflection (Ours)	91.4	2.4	1330.3	79.1%	88.9%
+SimPO _{Greedily Diverse Solutions}	91.2	1.7	1285.8	80.1%	90.2%

- Our proposed method maintains comparable math reasoning performance.
- Also greatly reduces generated tokens and improves efficiency.