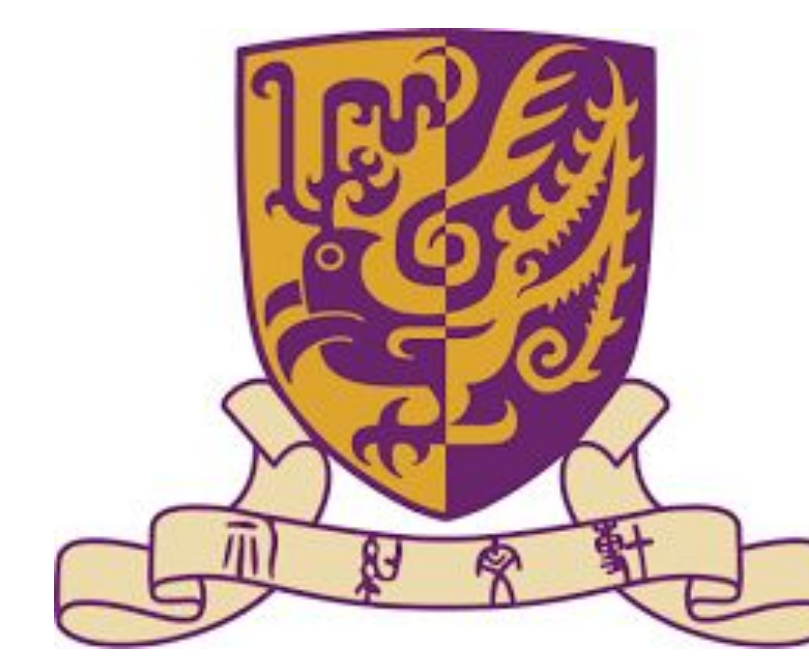




CALM Before the STORM: Unlocking Native Reasoning for Optimization Modeling

Zhengyang Tang*, Zihan Ye*, Chenyu Huang*, Xuhan Huang, Chengpeng Li, Sihang Li, Guanhua Chen, Ming Yan, Zizhuo Wang, Hongyuan Zha, Dayiheng Liu, Benyou Wang

The Chinese University of Hong Kong, Shenzhen; Shanghai University of Finance and Economics; Southern University of Science and Technology; Shenzhen Loop Area Institute; Qwen Team, Alibaba Group

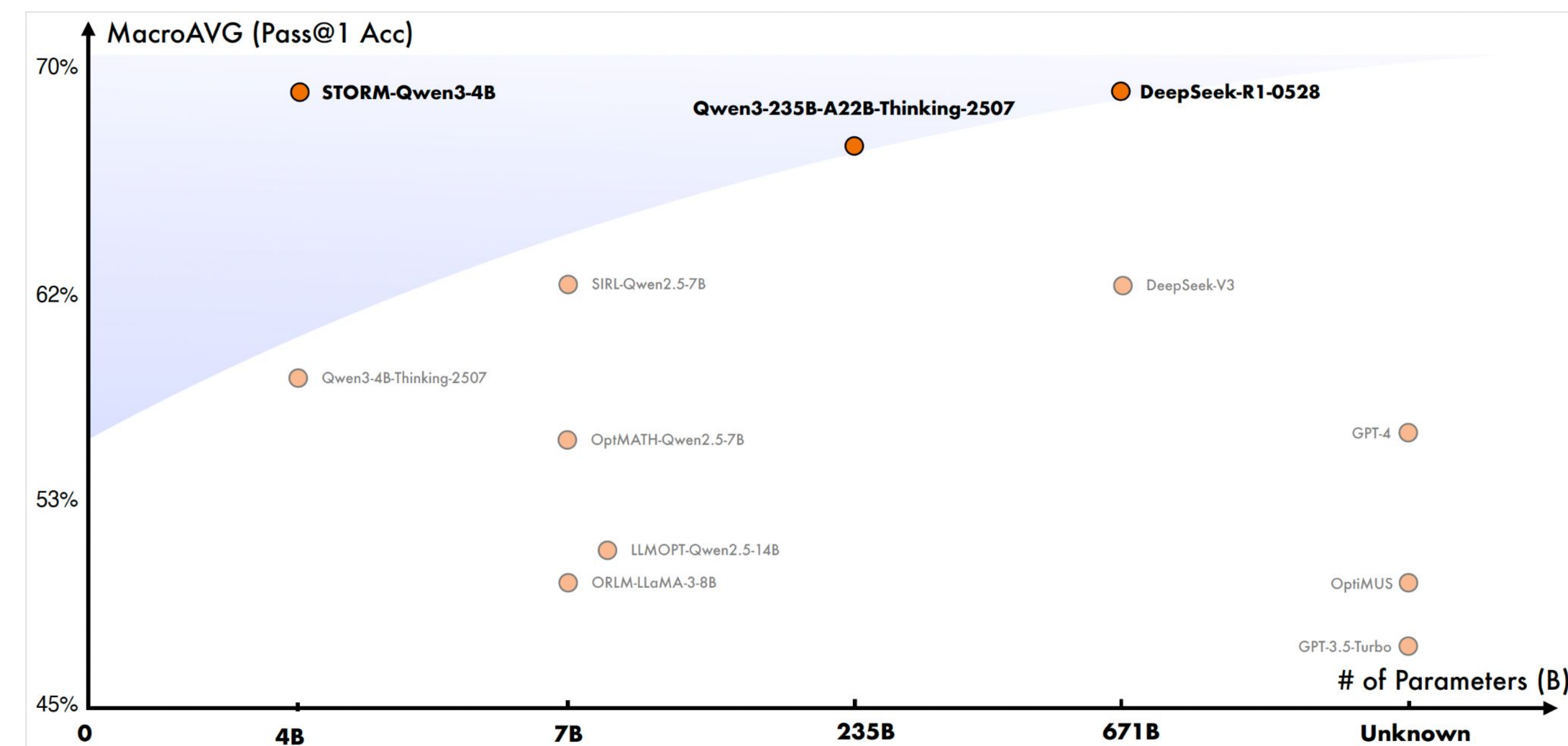


深圳河套学院
Shenzhen Loop Area Institute

Introduction

Large reasoning models create new opportunities for optimization modeling, but post-training is delicate.

Our key idea is simple: repair the model's own reasoning instead of replacing the whole solution.



Task Definition

The task is to translate a natural-language optimization problem into a mathematical model and solver code. Strong performance requires formulate, code, execute, inspect, and revise.

Problem

Transport 25 tons product via trucks/airplanes/ships.
Costs: \$100 / \$120 / \$130; caps: 10 / 20 / 30.
Truck and ship can't be chosen together...

Model

$$\begin{aligned} \min & 100y_1 + 120y_2 + 130y_3 \\ \text{s.t. } & y_1 \leq 10x_1 \\ & \dots \end{aligned}$$

Solver Code (PuLP)

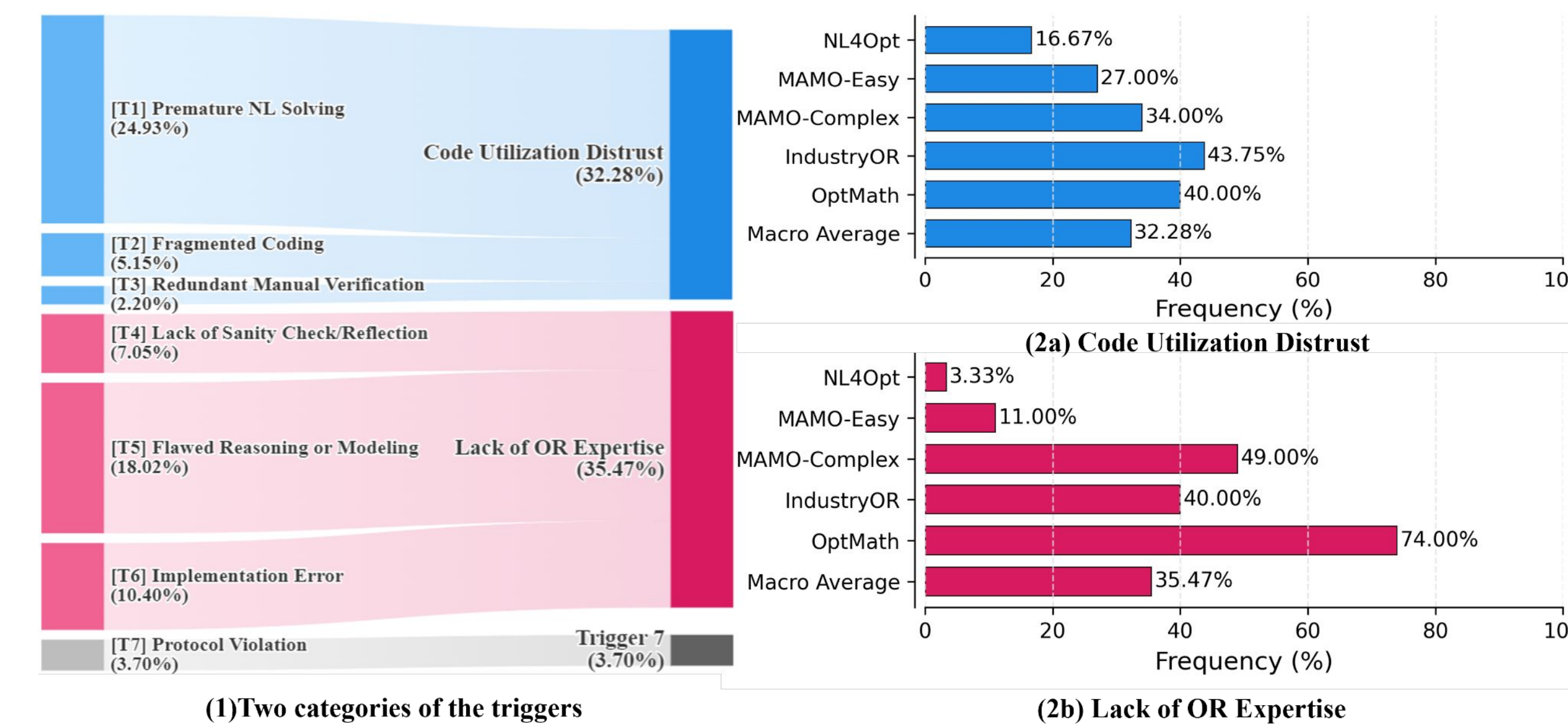
```
from pulp import LpProblem...  
m = LpProblem("OR", LpMinimize)  
x={...}; y={...}  
# ...
```

Pilot Study

Fine-tuning on complete solutions helps easy benchmarks but hurts hard ones. This suggests complete-solution supervision can interfere with the base model's native reasoning process.

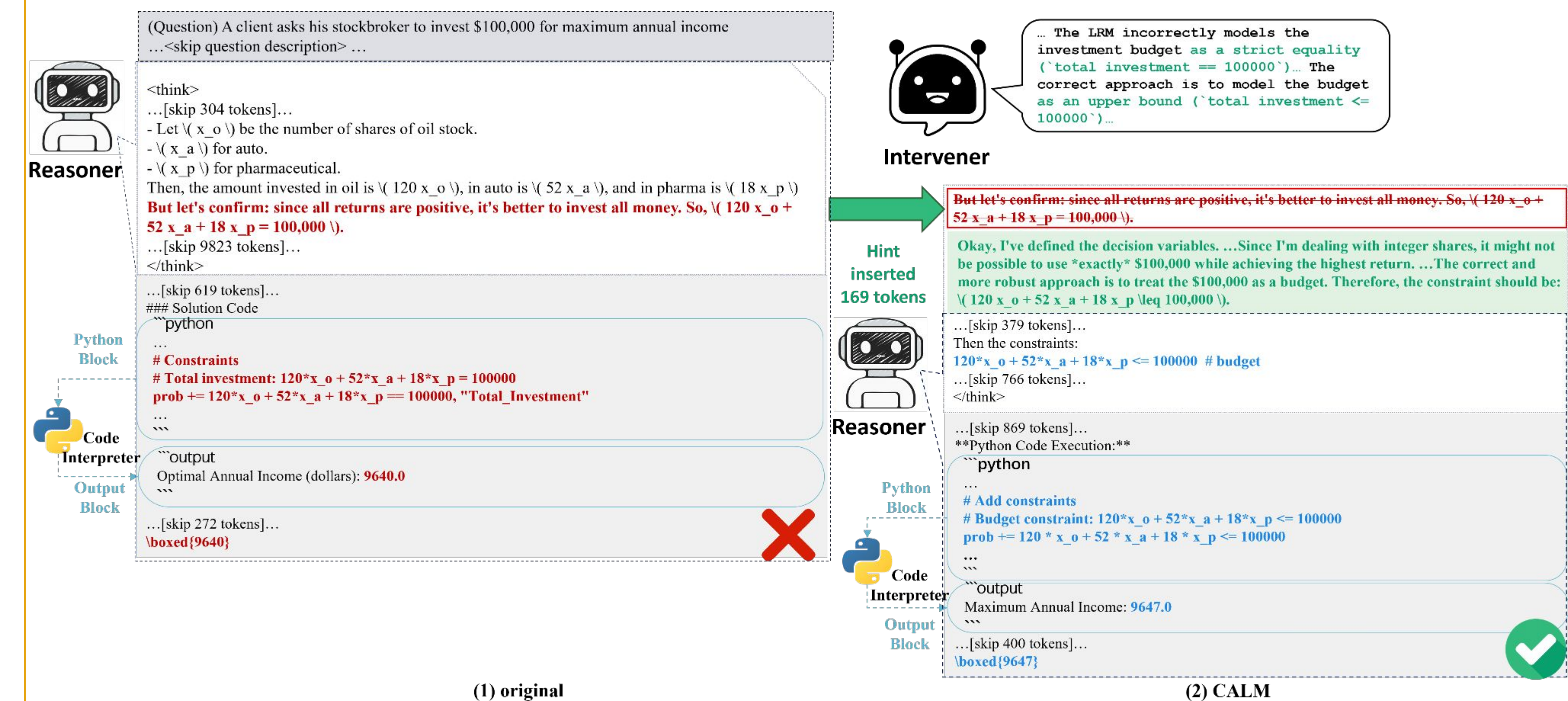
Benchmark	Base LRM	+ SFT on OR-Instruct-3K	Absolute Change
NL4OPT	85.8	92.9	+7.1
MAMO Easy	73.8	88.7	+14.9
MAMO Complex	46.5	40.5	-6.0
IndustryOR	46.2	27.5	-18.7
OptMath	33.1	6.6	-26.5
Macro AVG	57.1	51.2	-5.9

A Taxonomy of Failures in the Base LRM's Solutions:



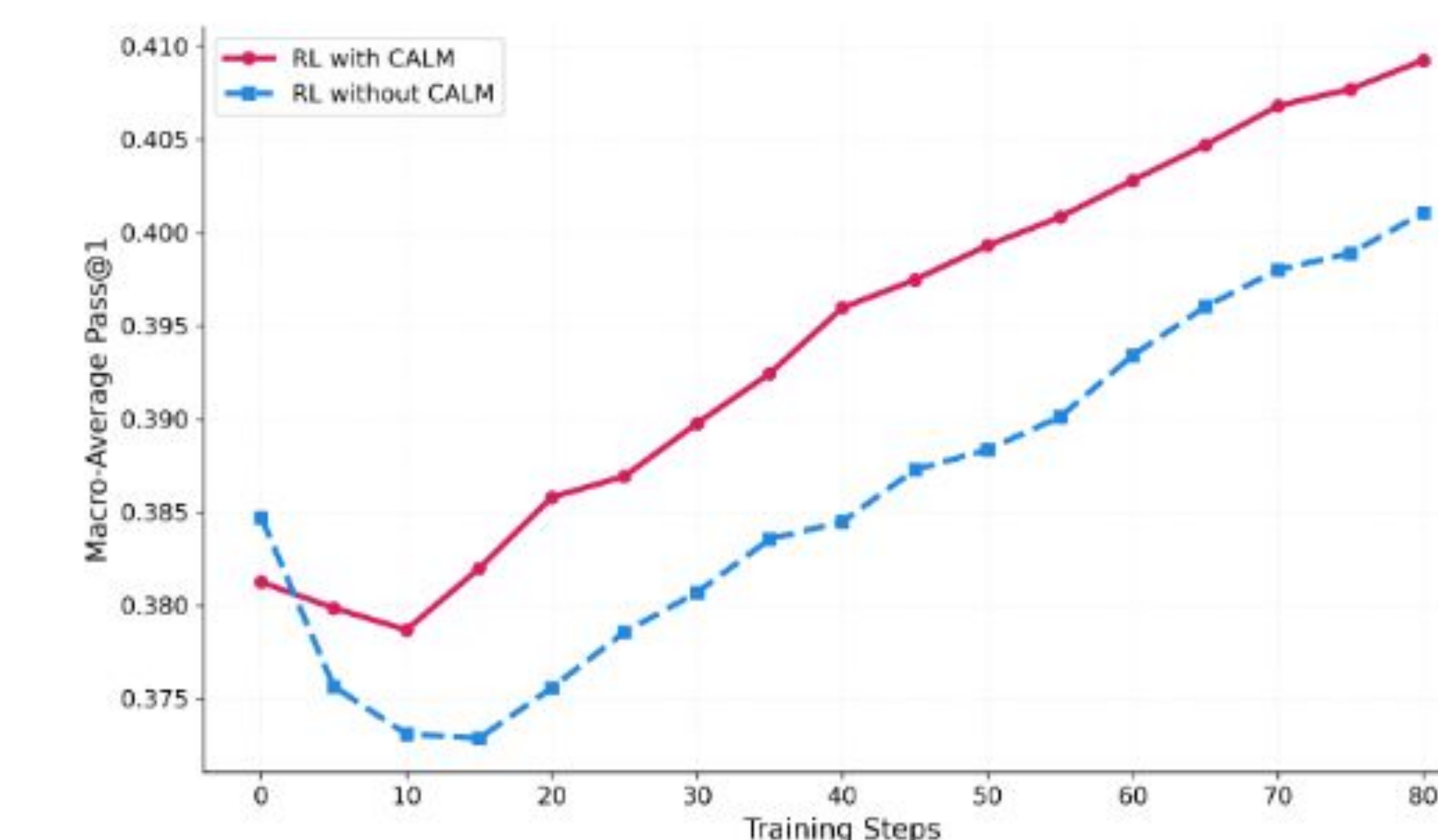
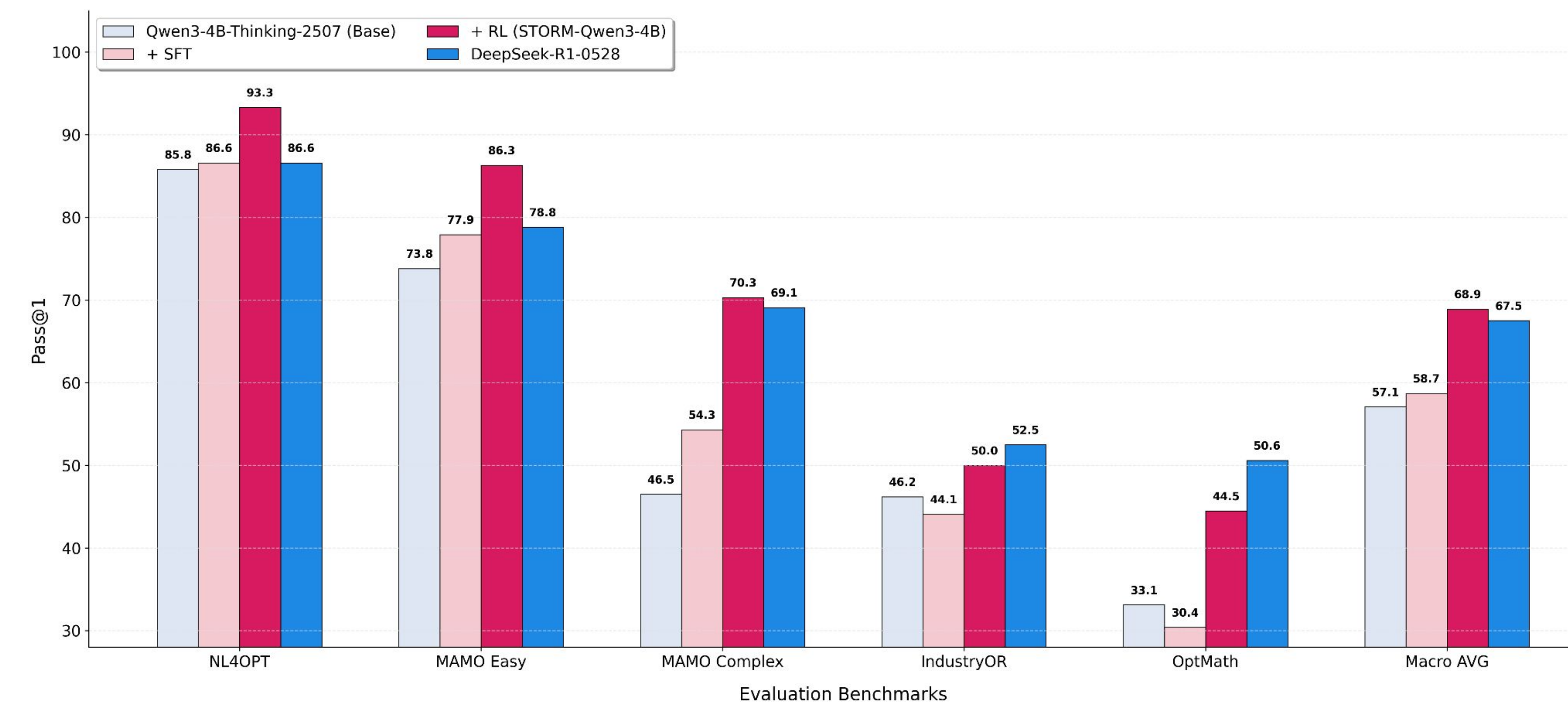
CALM Method

Let the base model try first. Detect the first mistake, insert a short hint, and continue from the repaired context.

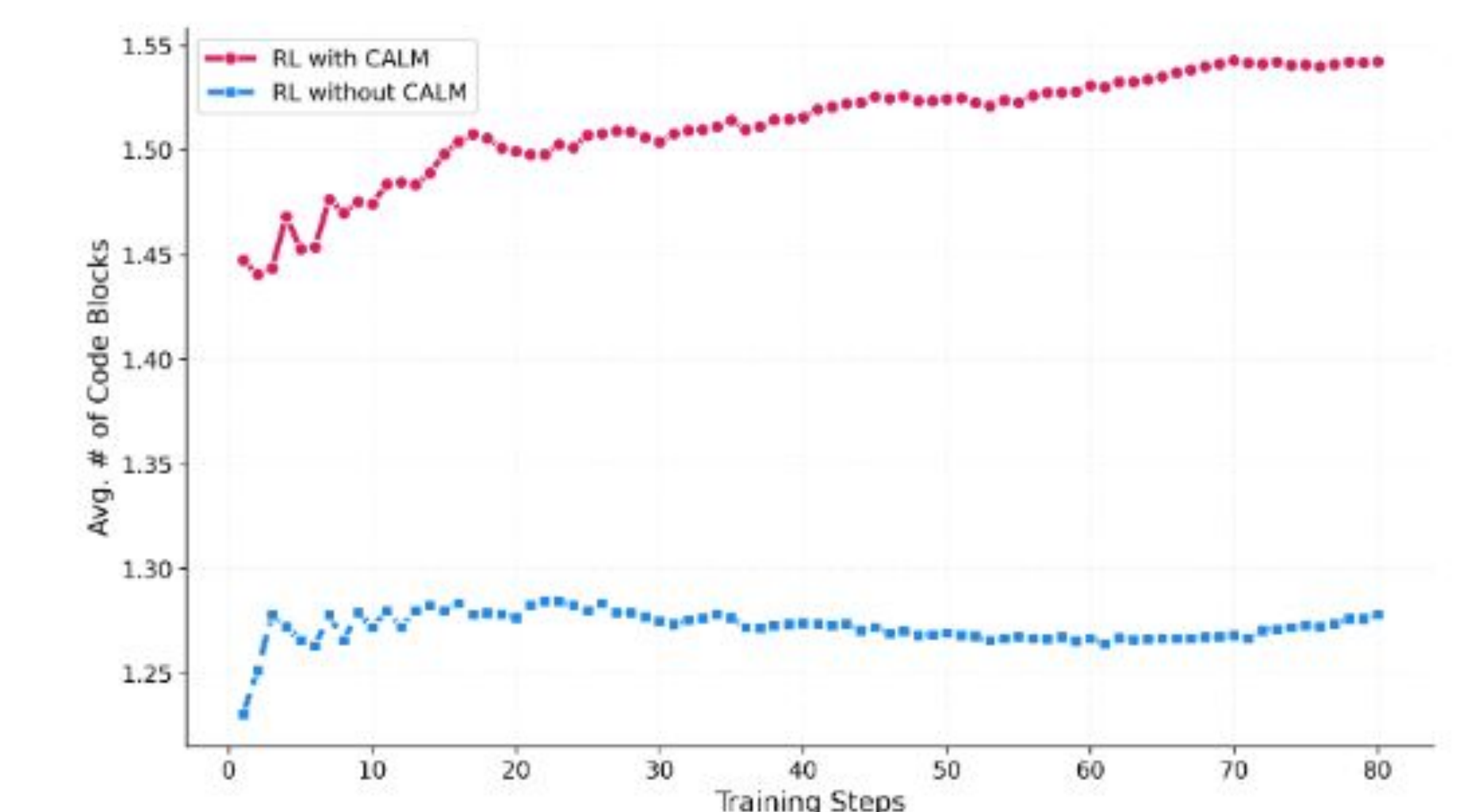


STORM Results

STORM-4B matches 671B DeepSeek-R1-0528 on five optimization modeling benchmarks. RL with CALM increases code use and improves solver reliability.



(a) RL Performance on Complex Test Sets.



(b) Average Number of Code Blocks.