

AIMO 3

WHY IS HARD?

Beisenbek Nurassyl

NURIKW3

Table of Contents

IDEAS

Realisations

why is hard

RESEARCH INSIDES

IDEAS

Prompt Engineering via verify

Reasoning model + Custom
Repl + CoT
3 algo types

Sfr LoRA

Agentic

Realisation I

Prompt Engineering via verify

Solution Architecture

vLLM server - runs GPT-OSS-120B locally with OpenAI-compatible API

Parallel generation - generates 4 solution attempts simultaneously for each problem

Problem-Solving Process

Streaming with early stopping - generates solutions in real-time, stops immediately when valid answer found

Answer extraction - finds integers 0-99999 inside \boxed{} format using regex

Follow-up Question Mechanism

Smart verification - automatically asks follow-up questions if answer ≤ 10 , not in proper format, or GPU usage is low

Forced guessing - if generation exceeds 50k tokens, requests an educated guess

Voting System

Weighted voting - uses $\log(1.25 + |\text{value}|) \times \text{count}$ to penalize suspiciously small numbers

Sota's



Gpt-oss-120B

96% cuda

7 / 10

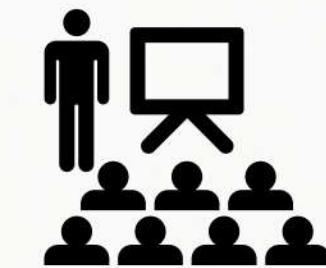


Repl + CoT

Qwen-2507-30b-thinking

424e18, a295eg

7/10



DeepseekR1-32B

8/10

— Why is HARD?

- Stochasticity
- Range of answer [AIMO2 - 0->9999]
- AIMO3 - 0->99999]
- Limits
- Blackbox

AIMO 3



Stochasticity

```
→ base5=4044120 → digits 4,0,4,4,1,2,0 → ≥3: 4,4,4 → c=3, correct.\n\nk=17  
26214 r1, 26214/5=5242 r4, 5242/5=1048 r2, 1048/5=209 r3, 209/5=41 r4, 41/  
extracted boxed answers: [62138]  
extracted python answers: []  
all extracted answers: [62138]  
answer: 62138  
processing 424e18  
score_list | 50.0 over 4 attempts  
[debug] WRONG: predicted 62138, but actual answer 21818 | score=2/4  
round 0
```

```
extracted boxed answers: [0]  
extracted python answers: []  
all extracted answers: [0]  
answer: 0  
processing 424e18  
score_list | 55.6 over 9 attempts  
[debug] WRONG: predicted 0, but actual answer 21818 | score=5/9  
round 0
```

Adding requests: 100%

5/5 [00:00<00:00, 459.52it/s]

Blackbox

```
+262+52+10+2=6553+1310=7863+262=8125+52=8177+10=8187+2=8189\n\nThen v5  
6379-16378=1. Ah, there we go, I miscalculated v5(32')  
extracted boxed answers: [1, 62140, 21818]  
extracted python answers: []  
all extracted answers: [1, 62140, 21818]  
answer: 1  
processing 424e18  
score_list | 66.7 over 6 attempts  
[debug] WRONG: predicted 1, but actual answer 21818 | score=4/9  
round 0
```

The End

THANK YOU FOR LISTENING

Beisenbek Nurassyl
NURIKW3