

## Supplementary Material

Table 10: **Updated: We fill addition results from Qwen3-Instruct (non-think).** Accuracy comparison of our GENERAL-REASONER with baseline methods on general reasoning benchmarks. MMLU-Pro, SuperGPQA, TheoremQA and BBEH contain multiple subfields.

Model Name Metric	Backbone	MMLU-Pro Micro	GPQA-D Acc	SuperGPQA Macro (discipline)	TheoremQA Acc	BBEH Micro
MiMo-RL	MiMo-Base	58.6	54.4	40.5	38.8	11.4
QwQ-32B	Qwen2.5-32B-Inst	52.0	54.5	43.6	48.4	22.6
GPT-4o	-	74.6	50.0	46.3	43.6	22.3
o1-mini	-	80.3	60.0	45.2	53.1	-
DeepSeek-R1	DeepSeek-V3	84.0	71.5	59.9	59.1	34.9
4B Models						
Qwen3-4B-Base	-	51.6	26.3	25.4	34.8	8.1
Qwen3-4B-Instruct (non-think)	Qwen3-4B-Base	61.8	41.7	32.1	42.0	<b>14.9</b>
GENERAL-REASONER-4B	Qwen3-4B-Base	<b>62.8</b>	<b>42.9</b>	<b>32.5</b>	<b>48.3</b>	12.2
7B Models						
Qwen2.5-7B-Base	-	47.7	29.3	26.7	29.1	8.0
Qwen2.5-7B-Instruct	Qwen2.5-7B-Base	57.0	33.8	30.7	36.6	12.2
Open-Reasoner-Zero	Qwen2.5-7B-Base	<b>59.4</b>	36.6	32.8	37.4	12.2
Nemotron-CrossThink	Qwen2.5-7B-Base	57.8	38.5	29.1	-	-
SimpleRL-Qwen2.5-7B-Zoo	Qwen2.5-7B-Base	51.5	24.2	29.9	38.0	11.9
GENERAL-REASONER-7B	Qwen2.5-7B-Base	58.9	<b>38.8</b>	<b>34.2</b>	<b>45.3</b>	<b>12.5</b>
14B Models						
Qwen2.5-14B-Base	-	53.3	32.8	30.7	33.0	10.8
Qwen2.5-14B-Instruct	Qwen2.5-14B-Base	62.7	41.4	35.8	41.9	15.2
SimpleRL-Qwen2.5-14B-Zoo	Qwen2.5-14B-Base	64.0	39.4	35.7	40.8	13.6
GENERAL-REASONER-Qw2.5-14B	Qwen2.5-14B-Base	66.6	43.4	39.5	44.3	15.2
Qwen3-14B-Base	-	64.2	45.9	36.5	44.0	13.0
Qwen3-14B-Instruct (non-think)	Qwen3-14B-Base	<b>70.9</b>	54.8	39.8	42.4	<b>19.2</b>
GENERAL-REASONER-Qw3-14B	Qwen3-14B-Base	70.3	<b>56.1</b>	<b>39.9</b>	<b>54.4</b>	17.3

Table 11: **Updated: We fill addition results from Qwen3-Instruct (non-think).** Math task accuracy across datasets. Most evaluations are based on greedy decoding except for AIME24 and AIME25 (averaged over 32 sampling runs following SimpleRL [2]).

Model Name	AVG	MATH-500	Olympiad	Minerva	GSM8K	AMC	AIME24	AIME25
4B Models								
Qwen3-4B-Base	22.5	38.8	16.4	18.4	39.6	27.5	10.3	6.7
Qwen3-4B-Instruct (non-think)	<b>54.2</b>	80.4	<b>49.0</b>	57.0	92.0	<b>62.5</b>	<b>22.5</b>	<b>16.1</b>
GENERAL-REASONER-4B	53.4	<b>80.6</b>	47.7	<b>57.7</b>	<b>92.2</b>	60.0	20.0	15.4
7B Models								
Qwen2.5-7B-Base	34.7	60.2	28.6	36.0	83.1	30.0	3.8	1.4
Qwen2.5-7B-Instruct	46.3	75.0	39.4	45.2	90.9	52.5	12.5	8.5
SimpleRL-Qwen2.5-7B-Zoo	48.4	74.0	<b>41.9</b>	49.6	90.7	<b>60.0</b>	<b>15.2</b>	7.5
GENERAL-REASONER-7B	<b>48.5</b>	<b>76.0</b>	37.9	<b>54.0</b>	<b>92.7</b>	55.0	13.8	<b>10.4</b>
14B Models								
Qwen2.5-14B-Base	37.0	65.4	33.5	24.3	91.6	37.5	3.6	2.9
Qwen2.5-14B-Instruct	49.9	77.4	44.7	52.2	<b>94.5</b>	57.5	12.2	11.0
SimpleRL-Qwen2.5-14B-Zoo	50.7	77.2	44.6	54.0	94.2	60.0	12.9	11.8
GENERAL-REASONER-Qw2.5-14B	53.9	78.6	42.1	58.1	94.2	70.0	17.5	16.9
Qwen3-14B-Base	49.9	74.6	44.3	55.9	93.2	55.0	14.7	11.4
Qwen3-14B-Instruct (non-think)	57.0	82.0	<b>52.4</b>	59.9	93.9	57.5	<b>28.5</b>	<b>25.1</b>
GENERAL-REASONER-Qw3-14B	<b>58.8</b>	<b>83.8</b>	51.9	<b>68.0</b>	<b>94.4</b>	<b>70.0</b>	24.4	19.2