

Table 32: **Benchmark scores for language: Russian (ru)**. The highest and second-best scores are shown in **bold** and underlined, respectively.

	Model	Multi-IF	INCLUDE	MT-AIME24	PolyMath	Average
Thinking Mode	Gemini2.5-Pro	68.1	80.4	70.0	<u>52.3</u>	67.7
	QwQ-32B	61.2	73.2	<u>76.7</u>	43.6	63.7
	Qwen3-235B-A22B	62.2	80.4	80.0	53.1	68.9
	Qwen3-32B	62.5	73.2	63.3	46.5	61.4
	Qwen3-30B-A3B	60.7	<u>76.8</u>	73.3	45.4	64.0
	Qwen3-14B	<u>63.6</u>	80.4	66.7	46.4	64.3
	Qwen3-8B	62.9	69.6	63.3	37.7	58.4
	Qwen3-4B	52.8	69.6	56.7	36.6	53.9
	Qwen3-1.7B	37.8	46.4	20.0	22.8	31.8
	Qwen3-0.6B	26.4	46.4	3.3	7.0	20.8
Non-thinking Mode	GPT-4o-2024-1120	52.0	80.4	20.0	13.7	41.5
	Gemma-3-27b-IT	57.3	71.4	23.3	21.6	43.4
	Qwen2.5-72B-Instruct	54.1	67.9	20.0	13.3	38.8
	Qwen3-235B-A22B	56.7	<u>75.0</u>	40.0	26.1	49.4
	Qwen3-32B	58.6	71.4	<u>30.0</u>	23.3	<u>45.8</u>
	Qwen3-30B-A3B	58.0	73.2	<u>30.0</u>	21.1	45.6
	Qwen3-14B	<u>60.3</u>	71.4	26.7	<u>24.2</u>	45.6
	Qwen3-8B	<u>59.3</u>	58.9	20.0	22.8	40.2
	Qwen3-4B	46.1	58.9	13.3	17.8	34.0
	Qwen3-1.7B	34.8	41.1	3.3	13.2	23.1
	Qwen3-0.6B	25.5	46.4	0.0	5.8	19.4

Table 33: **Benchmark scores for language: Vietnamese (vi)**. The highest and second-best scores are shown in **bold** and underlined, respectively.

	Model	MLogiQA	INCLUDE	MT-AIME24	PolyMath	Average
Thinking Mode	Gemini2.5-Pro	72.5	89.1	70.0	52.1	70.9
	QwQ-32B	71.2	69.1	70.0	<u>49.2</u>	64.9
	Qwen3-235B-A22B	75.0	87.3	83.3	55.1	75.2
	Qwen3-32B	67.5	81.8	83.3	44.0	69.2
	Qwen3-30B-A3B	68.8	78.2	<u>76.7</u>	46.1	67.4
	Qwen3-14B	<u>72.5</u>	72.7	73.3	45.8	66.1
	Qwen3-8B	65.0	72.7	73.3	42.9	63.5
	Qwen3-4B	68.8	63.6	60.0	42.2	58.6
	Qwen3-1.7B	52.5	61.8	30.0	26.9	42.8
	Qwen3-0.6B	33.8	38.2	6.7	9.8	22.1
Non-thinking Mode	GPT-4o-2024-1120	57.5	81.8	10.0	13.0	40.6
	Gemma-3-27b-IT	52.5	<u>74.5</u>	<u>33.3</u>	20.6	45.2
	Qwen2.5-72B-Instruct	61.3	72.7	26.7	18.6	44.8
	Qwen3-235B-A22B	70.0	83.6	36.7	27.1	54.4
	Qwen3-32B	60.0	<u>81.8</u>	23.3	21.8	<u>46.7</u>
	Qwen3-30B-A3B	52.5	<u>81.8</u>	20.0	<u>24.7</u>	44.8
	Qwen3-14B	<u>63.7</u>	67.3	20.0	21.6	43.2
	Qwen3-8B	48.8	65.5	20.0	19.1	38.4
	Qwen3-4B	48.8	65.5	20.0	19.0	38.3
	Qwen3-1.7B	36.2	60.0	3.3	10.9	27.6
	Qwen3-0.6B	30.0	36.4	3.3	3.9	18.4