Table 28: **Benchmark scores for language: Arabic (ar)**. The highest and second-best scores are shown in **bold** and underlined, respectively.

	Model	MLogiQA	INCLUDE	MMMLU	MT-AIME24	PolyMath	Average
Thinking Mode	Gemini2.5-Pro	<u>75.0</u>	**89.3**	**87.8**	76.7	<u>52.6</u>	**76.3**
	QwQ-32B	<u>75.0</u>	67.9	81.8	80.0	41.3	69.2
	Qwen3-235B-A22B	**80.0**	71.4	<u>83.6</u>	76.7	**53.7**	<u>73.1</u>
	Qwen3-32B	66.2	<u>73.2</u>	80.1	**86.7**	47.0	70.6
	Qwen3-30B-A3B	66.2	66.1	77.2	<u>83.3</u>	47.3	68.0
	Qwen3-14B	71.2	67.9	77.4	<u>83.3</u>	46.6	69.3
	Qwen3-8B	65.0	67.9	74.4	76.7	44.9	65.8
	Qwen3-4B	62.5	55.4	67.7	66.7	41.2	58.7
	Qwen3-1.7B	55.0	44.6	53.2	36.7	25.8	43.1
	Qwen3-0.6B	40.0	41.1	38.9	10.0	11.7	28.3
Non-thinking Mode	GPT-4o-2024-1120	51.2	**78.6**	**80.9**	13.3	12.9	47.4
	Gemma-3-27b-IT	<u>56.2</u>	62.5	74.4	26.7	22.8	48.5
	Qwen2.5-72B-Instruct	<u>56.2</u>	66.1	77.2	6.7	14.7	44.2
	Qwen3-235B-A22B	**66.2**	67.9	<u>79.5</u>	**40.0**	**28.2**	**56.4**
	Qwen3-32B	55.0	<u>69.6</u>	75.7	23.3	<u>25.4</u>	<u>49.8</u>
	Qwen3-30B-A3B	48.8	64.3	71.6	<u>30.0</u>	22.6	47.5
	Qwen3-14B	52.5	60.7	69.5	23.3	23.5	45.9
	Qwen3-8B	45.0	58.9	64.6	13.3	16.4	39.6
	Qwen3-4B	52.5	42.9	56.7	13.3	15.3	36.1
	Qwen3-1.7B	31.2	37.5	43.6	3.3	9.4	25.0
	Qwen3-0.6B	40.0	39.3	35.4	0.0	3.8	23.7

Table 29: **Benchmark scores for language: Japanese (ja)**. The highest and second-best scores are shown in **bold** and underlined, respectively.

	Model	MLogiQA	INCLUDE	MMMLU	MT-AIME24	PolyMath	Average
Thinking Mode	Gemini2.5-Pro	72.5	74.5	<u>83.8</u>	**83.3**	**55.4**	<u>73.9</u>
	QwQ-32B	<u>73.8</u>	86.3	82.3	53.3	39.9	67.1
	Qwen3-235B-A22B	**75.0**	**94.1**	**84.8**	73.3	<u>52.7</u>	**76.0**
	Qwen3-32B	70.0	<u>90.2</u>	80.2	<u>76.7</u>	47.7	73.0
	Qwen3-30B-A3B	66.2	88.2	79.9	73.3	47.4	71.0
	Qwen3-14B	68.8	88.2	79.4	66.7	45.7	69.8
	Qwen3-8B	71.2	86.3	74.9	73.3	44.7	70.1
	Qwen3-4B	63.7	80.4	72.5	53.3	40.7	62.1
	Qwen3-1.7B	53.8	74.5	61.8	36.7	28.5	51.1
	Qwen3-0.6B	47.5	47.1	45.1	13.3	14.5	33.5
Non-thinking Mode	GPT-4o-2024-1120	60.0	<u>92.2</u>	**81.9**	10.0	12.5	51.3
	Gemma-3-27b-IT	<u>66.2</u>	86.3	76.5	20.0	17.3	53.3
	Qwen2.5-72B-Instruct	55.0	**94.1**	77.7	16.7	17.7	52.2
	Qwen3-235B-A22B	**67.5**	<u>92.2</u>	<u>80.9</u>	26.7	**26.9**	**58.8**
	Qwen3-32B	58.8	<u>92.2</u>	78.0	20.0	20.5	<u>53.9</u>
	Qwen3-30B-A3B	51.2	82.4	74.9	<u>30.0</u>	<u>20.6</u>	51.8
	Qwen3-14B	55.0	84.3	73.8	**33.3**	19.8	53.2
	Qwen3-8B	47.5	82.4	69.9	20.0	18.5	47.7
	Qwen3-4B	46.2	76.5	64.8	13.3	15.1	43.2
	Qwen3-1.7B	40.0	68.6	46.3	3.3	11.6	34.0
	Qwen3-0.6B	37.5	37.3	37.9	3.3	3.7	23.9