

Table 24: **Benchmark scores for language: Spanish (es).** The highest and second-best scores are shown in **bold** and underlined, respectively.

	Model	Multi-IF	MLogiQA	INCLUDE	MMMLU	MT-AIME24	PolyMath	Average
Thinking Mode	Gemini2.5-Pro	80.1	70.0	96.4	88.7	90.0	54.4	79.9
	QwQ-32B	70.0	<u>75.0</u>	81.8	84.5	76.7	52.2	73.4
	Qwen3-235B-A22B	74.2	76.2	89.1	<u>86.7</u>	<u>86.7</u>	57.3	78.4
	Qwen3-32B	74.7	68.8	<u>90.9</u>	82.8	76.7	51.8	74.3
	Qwen3-30B-A3B	74.9	71.2	80.0	81.9	76.7	48.5	72.2
	Qwen3-14B	<u>76.2</u>	67.5	83.6	81.1	73.3	50.3	72.0
	Qwen3-8B	74.1	70.0	78.2	79.2	70.0	43.7	69.2
	Qwen3-4B	69.1	68.8	72.7	75.7	66.7	42.3	65.9
	Qwen3-1.7B	56.0	55.0	72.7	64.5	46.7	30.2	54.2
	Qwen3-0.6B	39.2	42.5	54.5	48.8	13.3	14.3	35.4
Non-thinking Mode	GPT-4o-2024-1120	67.5	52.5	89.1	80.6	10.0	15.5	52.5
	Gemma-3-27b-IT	<u>73.5</u>	57.5	89.1	<u>77.7</u>	30.0	22.4	58.4
	Qwen2.5-72B-Instruct	66.7	61.3	80.0	80.1	20.0	18.8	54.5
	Qwen3-235B-A22B	71.7	66.2	83.6	83.7	<u>33.3</u>	29.5	61.3
	Qwen3-32B	72.1	<u>65.0</u>	83.6	80.4	26.7	24.7	58.8
	Qwen3-30B-A3B	72.1	53.8	<u>85.5</u>	78.3	<u>33.3</u>	<u>25.0</u>	58.0
	Qwen3-14B	<u>76.2</u>	63.7	78.2	77.4	40.0	<u>25.0</u>	60.1
	Qwen3-8B	73.1	50.0	80.0	73.7	16.7	21.3	52.5
	Qwen3-4B	65.8	50.0	60.0	68.3	13.3	17.3	45.8
	Qwen3-1.7B	47.9	43.8	50.9	54.3	10.0	11.6	36.4
	Qwen3-0.6B	35.5	37.5	43.6	39.5	3.3	5.8	27.5

Table 25: **Benchmark scores for language: French (fr).** The highest and second-best scores are shown in **bold** and underlined, respectively.

	Model	Multi-IF	MLogiQA	INCLUDE	MMMLU	MT-AIME24	PolyMath	Average
Thinking Mode	Gemini2.5-Pro	80.5	73.8	85.7	88.3	80.0	52.8	76.8
	QwQ-32B	72.4	<u>78.8</u>	76.2	84.0	80.0	49.4	73.5
	Qwen3-235B-A22B	77.3	<u>78.8</u>	85.7	<u>86.6</u>	86.7	57.4	78.8
	Qwen3-32B	76.7	81.2	76.2	82.1	<u>83.3</u>	47.1	74.4
	Qwen3-30B-A3B	75.2	67.5	<u>83.3</u>	81.0	76.7	46.9	71.8
	Qwen3-14B	<u>77.6</u>	71.2	73.8	80.4	73.3	44.2	70.1
	Qwen3-8B	73.8	66.2	85.7	77.9	70.0	45.3	69.8
	Qwen3-4B	71.3	63.7	71.4	74.5	66.7	40.2	64.6
	Qwen3-1.7B	52.6	56.2	54.8	64.8	60.0	28.7	52.8
	Qwen3-0.6B	36.1	48.8	47.6	48.4	6.7	14.0	33.6
Non-thinking Mode	GPT-4o-2024-1120	67.8	56.2	85.7	81.8	10.0	15.3	52.8
	Gemma-3-27b-IT	73.9	57.5	73.8	78.3	23.3	21.5	54.7
	Qwen2.5-72B-Instruct	72.1	55.0	81.0	80.2	26.7	15.7	55.1
	Qwen3-235B-A22B	73.2	<u>65.0</u>	88.1	<u>81.1</u>	36.7	28.1	62.0
	Qwen3-32B	<u>75.8</u>	60.0	73.8	79.5	30.0	23.0	57.0
	Qwen3-30B-A3B	75.6	52.5	69.0	77.9	26.7	<u>27.3</u>	54.8
	Qwen3-14B	78.4	<u>63.7</u>	73.8	75.1	<u>33.3</u>	24.4	58.1
	Qwen3-8B	71.9	52.5	71.4	71.7	20.0	21.4	51.5
	Qwen3-4B	64.2	47.5	61.9	67.6	20.0	19.2	46.7
	Qwen3-1.7B	46.1	43.8	64.3	53.2	3.3	11.6	37.0
	Qwen3-0.6B	32.8	35.0	38.1	39.4	6.7	4.6	26.1