Modello	Parametri	Punteggio
Cogito	32b	66.14/100
Qwen2.5	32b	65.60/100
Qwen2.5	14b	63.49/100
Gemma3	12b	62.76/100
Command-r	35b	62.26/100
Aya-expanse	32b	62.07/100
Granite3.3	8b	59.96/100
Deepseek-R1	32b	59.87/100
Openthinker	32b	59.51/100
Qwen3	32b	58.73/100
Mistral-small	24b	58.19/100
Command-r7b	7b	58.13/100
Gemma2	27b	57.81/100
Nous-hermes2-mixtral	8x7b	57.65/100
Phi4	14b	55.62/100
Phi3	14b	54.40/100
Marco-o1	7b	54.22/100
Mixtral	8x7b	52.25/100
Yi	34b	51.87/100
Phi4-reasoning	14b	50.47/100
Mistral-nemo	12b	47.08/100
Phi4-mini-reasoning	3.8b	43.90/100

Table 1: Confronto tra modelli linguistici per punteggio ottenuto