Model	Score
openai/gpt-4o	0.4557
openai/o3-pro	0.4481
anthropic/claude-opus-4-thinking	0.4429
xAI/o3-pro	0.4312
openai/o3-2025-04-16	0.4294
openai/o3	0.4226
anthropic/claude-3-7-sonnet-20250219	0.3691
openai/gpt-4.1-mini-2025-04-14	0.3691
anthropic/claude-sonnet-4-thinking	0.3573
anthropic/claude-3-5-sonnet-20241022	0.3487
openai/o3-mini	0.3453
anthropic/claude-sonnet-4	0.3426
deepseek/deepseek-r1	0.3384
openai/gpt-4.1-nano-2025-04-14	0.3087
deepseek/deepseek-chat	0.3043
anthropic/claude-3-5-haiku-20241022	0.2963
google/gemini-ultra-preview	0.2789
xAI/grok-4	0.2447
xAI/grok-3	0.1928
google/gemini-2.5-pro-preview	0.1777
google/gemini-2.5-flash-preview	0.1341
google/gemini-2.5-flash-lite-preview	0.1032