Summary of performance on some simple inferences
Temperature: None, Condition: zero-shot_log_probs

Model	Average correct answer frequency
GPT-4 Turbo (2024-04-09)	87%
GPT-4 Turbo (1106)	83%
GPT-4 (0613)	83%
GPT-4 (0314)	80%
GPT-4o (2024-05-13)	79%
GPT-3.5 Turbo (0125)	57%
GPT-3.5 Turbo (1106)	55%