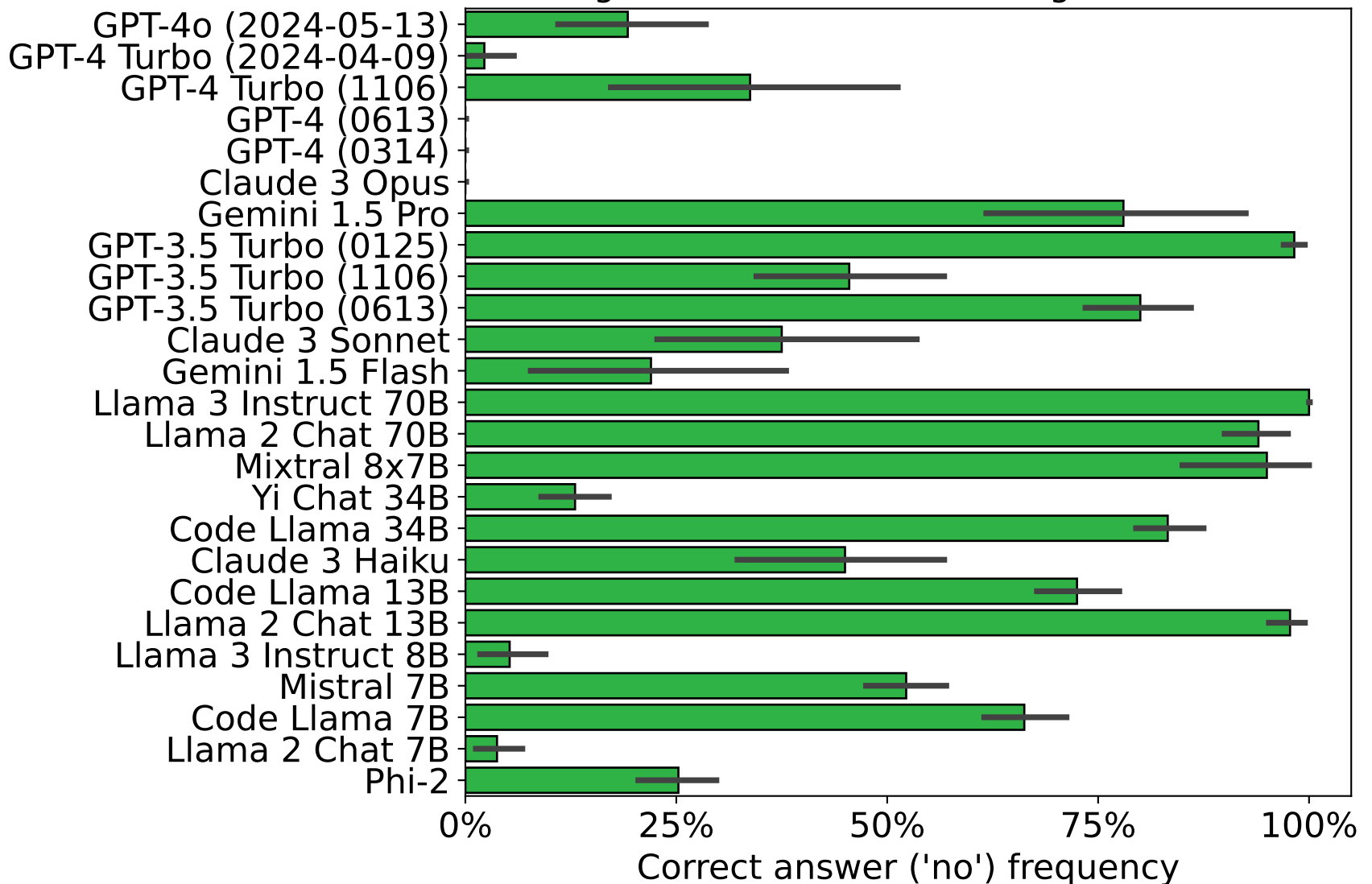DSmiox example: From 'The flugel might not be zargled' together with 'Either the flugel is blimmed or it must be zargled', can we infer 'The flugel is blimmed'?

Temperature: 1, Condition: few-shot

Correct answer ('no') frequency

Models (top to bottom):
- GPT-4o (2024-05-13)
- GPT-4 Turbo (2024-04-09)
- GPT-4 Turbo (1106)
- GPT-4 (0613)
- GPT-4 (0314)
- Claude 3 Opus
- Gemini 1.5 Pro
- GPT-3.5 Turbo (0125)
- GPT-3.5 Turbo (1106)
- GPT-3.5 Turbo (0613)
- Claude 3 Sonnet
- Gemini 1.5 Flash
- Llama 3 Instruct 70B
- Llama 2 Chat 70B
- Mixtral 8x7B
- Yi Chat 34B
- Code Llama 34B
- Claude 3 Haiku
- Code Llama 13B
- Llama 2 Chat 13B
- Llama 3 Instruct 8B
- Mistral 7B
- Code Llama 7B
- Llama 2 Chat 7B
- Phi-2