

Queries and Errors per Model

Model	Count	Type		
		Both Responses	Errors	Neither Response Errors
Claude 3 Haiku	2320	0	0	0
Claude 3 Sonnet	2320	0	0	0
Claude 3.5 Sonnet	2320	0	0	0
Claude 3 Opus	2320	0	0	0
Gemini 1.5 Flash	2320	0	1	0
Gemini 1.5 Pro	2320	0	0	0
Gemma 2 27B	2320	0	0	0
Code Llama 7B	2320	15	75	0
Code Llama 13B	2320	2	4	0
Code Llama 34B	2320	0	0	0
Llama 2 Chat 7B	2320	1	2	0
Llama 2 Chat 13B	2320	4	0	0
Llama 3 Instruct 8B	2320	5	2	0
Llama 3 Instruct 70B	2320	0	0	0
Llama 3.1 Instruct 8B	2320	31	0	0
Llama 3.1 Instruct 70B	2320	9	1	0
Llama 3.1 Instruct 405B	2320	0	0	0
Mistral 7B	2320	7	20	0
Mixtral 8x7B	2320	2	6	0
Mixtral 8x22B	2320	0	1	0
Mistral Large 2	2320	0	0	0
GPT-3.5 Turbo (1106)	2320	20	10	0
GPT-3.5 Turbo (0125)	2320	80	4	0
GPT-4 (0314)	2320	0	0	0
GPT-4 (0613)	2320	1	4	0
GPT-4 Turbo (1106)	2320	0	0	0
GPT-4 Turbo (2024-04-09)	2320	0	0	0
GPT-4o mini	2320	0	0	0
GPT-4o (2024-05-13)	2320	0	0	0
o1 Mini (2024-09-12)	0	0	0	0
o1 Preview (2024-09-12)	0	0	0	0

Queries

Both Responses Errors Neither Response Errors