

Model	Count	Queries and Errors per Model		
		Both Responses Errors	Neither Response Errors	
Claude 3 Haiku	2080	0	0	
Claude 3 Sonnet	2080	0	0	
Claude 3 Opus	2080	0	0	
Code Llama 7B	2080	19	103	
Code Llama 13B	2080	3	2	
Code Llama 34B	2080	3	0	
Gemini 1.5 Flash	2080	0	1	
Gemini 1.5 Pro	2080	0	0	
GPT-3.5 Turbo (0613)	2080	66	2	
GPT-3.5 Turbo (1106)	2080	20	10	
GPT-3.5 Turbo (0125)	2080	80	4	
GPT-4 (0314)	2080	0	0	
GPT-4 (0613)	2080	1	4	
GPT-4 Turbo (1106)	2080	0	0	
GPT-4 Turbo (2024-04-09)	2080	0	0	
GPT-4o (2024-05-13)	2080	0	0	
Llama 2 Chat 7B	2080	1	0	
Llama 2 Chat 13B	2080	0	0	
Llama 2 Chat 70B	2080	4	1	
Llama 3 Instruct 8B	2080	0	0	
Llama 3 Instruct 70B	2080	0	0	
Mistral 7B	2080	8	21	
Mixtral 8x7B	2080	1	6	
Phi-2	2080	1	8	
Yi Chat 34B	2080	243	116	
		Queries	Both Responses Errors	Neither Response Errors