

Queries and Errors per Model

| Model | | Queries | Both Responses | Errors | Neither Response Errors |
|--------------------------|-------|---------|----------------|--------|-------------------------|
| | | | Responses | Errors | Responses |
| Claude 3 Haiku | 33130 | 0 | | 0 | |
| Claude 3 Sonnet | 31330 | 0 | | 8 | |
| Claude 3.5 Sonnet | 14730 | 0 | | 0 | |
| Claude 3 Opus | 17055 | 0 | | 2 | |
| Gemini 1.5 Flash | 26753 | 0 | | 0 | |
| Gemini 1.5 Pro | 14249 | 0 | | 0 | |
| Gemma 2 27B | 28480 | 0 | | 8 | |
| Code Llama 7B | 46420 | 1 | | 318 | |
| Code Llama 13B | 46890 | 2 | | 28 | |
| Code Llama 34B | 47310 | 0 | | 5 | |
| Llama 2 Chat 7B | 42010 | 6 | | 33 | |
| Llama 2 Chat 13B | 35660 | 0 | | 20 | |
| Llama 3 Instruct 8B | 36530 | 0 | | 1 | |
| Llama 3 Instruct 70B | 27690 | 0 | | 0 | |
| Llama 3.1 Instruct 8B | 42268 | 0 | | 3 | |
| Llama 3.1 Instruct 70B | 31590 | 0 | | 1 | |
| Llama 3.1 Instruct 405B | 16260 | 0 | | 0 | |
| Mistral 7B | 27430 | 1 | | 41 | |
| Mixtral 8x7B | 25960 | 9 | | 21 | |
| Mixtral 8x22B | 43640 | 0 | | 520 | |
| Mistral Large 2 | 18645 | 0 | | 1 | |
| GPT-3.5 Turbo (1106) | 35200 | 0 | | 0 | |
| GPT-3.5 Turbo (0125) | 39410 | 0 | | 0 | |
| GPT-4 (0314) | 15555 | 0 | | 0 | |
| GPT-4 (0613) | 14925 | 0 | | 0 | |
| GPT-4 Turbo (1106) | 17280 | 0 | | 0 | |
| GPT-4 Turbo (2024-04-09) | 15405 | 0 | | 0 | |
| GPT-4o mini | 27500 | 0 | | 0 | |
| GPT-4o (2024-05-13) | 17610 | 0 | | 0 | |
| o1 Mini (2024-09-12) | 17410 | 0 | | 0 | |
| o1 Preview (2024-09-12) | 16130 | 0 | | 0 | |
| o3 (2025-04-16) | 14945 | 0 | | 0 | |
| Claude 3.7 Sonnet | 16215 | 0 | | 0 | |