

LLM judge score

5
4
3
2
1

gpt-4o-mini

o4-mini

