

F1-score (rank)

4

5

6

7

8

gpt-4o

4.7660

o1-mini

7.8601

cl-3.5-sonnet (rag)

4.8601

gpt-4o-mini (ftuning)

7.3069

gpt-4o (rag)

5.2194

cl-3.5-sonnet

6.4636

gpt-4o-mini (rag)

5.2427

gpt-4o-mini

6.4344

cl-3-haiku (rag)

5.3287

cl-3-haiku

6.3163

llama-3.1

6.2019

