

F1-score (rank)

4

5

6

7

8

o1-mini

4.3015

gpt-4o

4.5406

cl-3.5-sonnet

4.7138

gpt-4o-mini

5.4430

cl-3.5-sonnet (rag)

5.5011

gpt-4o-mini (ftuning)

6.7774

gpt-4o-mini (rag)

6.1963

cl-3-haiku (rag)

6.1129

gpt-4o (rag)

5.7237

cl-3-haiku

5.6897

