

F1-score (rank)

4

5

6

7

8

gpt-4o

4.3980

o1-mini

7.1873

cl-3.5-sonnet (rag)

4.4862

gpt-4o-mini (ftuning)

6.6997

gpt-4o (rag)

4.8149

cl-3.5-sonnet

5.9424

gpt-4o-mini (rag)

4.8375

gpt-4o-mini

5.9162

cl-3-haiku (rag)

4.9125

cl-3-haiku

5.8054