

F1-score (rank)

4

5

6

7

8

o1-mini

4.7675

gpt-4o

5.0285

llama-3.1

5.0855

cl-3.5-sonnet

5.2270

gpt-4o-mini

6.0241

cl-3.5-sonnet (rag)

6.1009

7.5044

gpt-4o-mini (ftuning)

6.8607

gpt-4o-mini (rag)

6.7632

cl-3-haiku (rag)

6.3388

gpt-4o (rag)

6.2993

cl-3-haiku