

e3 vs. open-source models

Acc. on HMMT '25

30
20
10
0

—●—

s1-32B

—●—

s1.1-32B

—●—

e3-1.7B

—●—

agentica-1.5B

—●—

qwen3-1.7B

—●—

openthinker-7B

4k

8k

16k

24k

32k

Test-time compute (tokens)

