

e3 vs. open-source models

Acc. on HMMT '25

30
20
10

4k 8k 16k 24k 32k

Test-time compute (tokens)

agentica-1.5B
qwen3-1.7B
openthinker-7B
s1.1-32B
r1-distilled-1.5B
e3-1.7B

