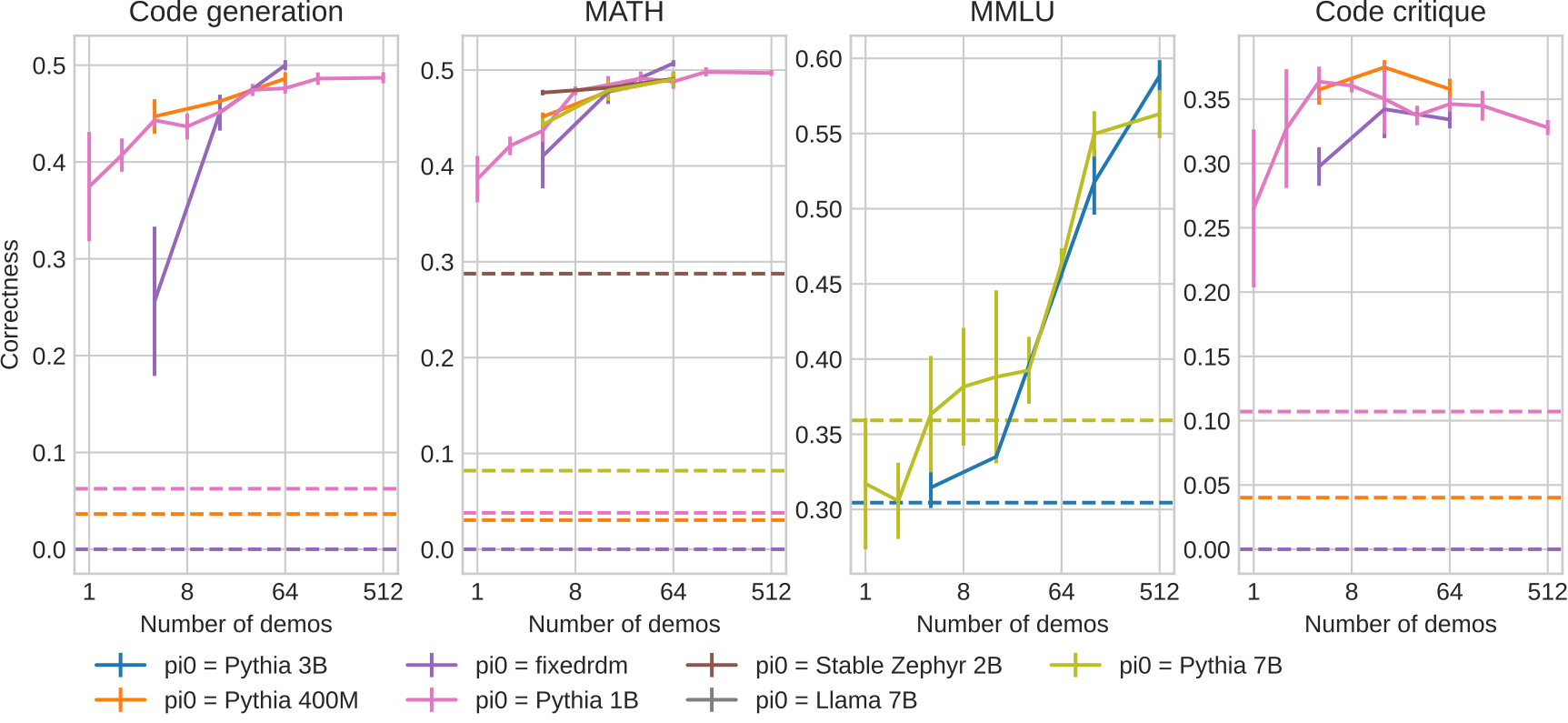
Code generation
MATH
MMLU
Code critique
Correctness
Number of demos
1
8
64
512
0.0
0.1
0.2
0.3
0.4
0.5
0.30
0.35
0.40
0.45
0.50
0.55
0.60
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
pi0 = Pythia 3B
pi0 = Pythia 400M
pi0 = fixedrdm
pi0 = Pythia 1B
pi0 = Stable Zephyr 2B
pi0 = Llama 7B
pi0 = Pythia 7B