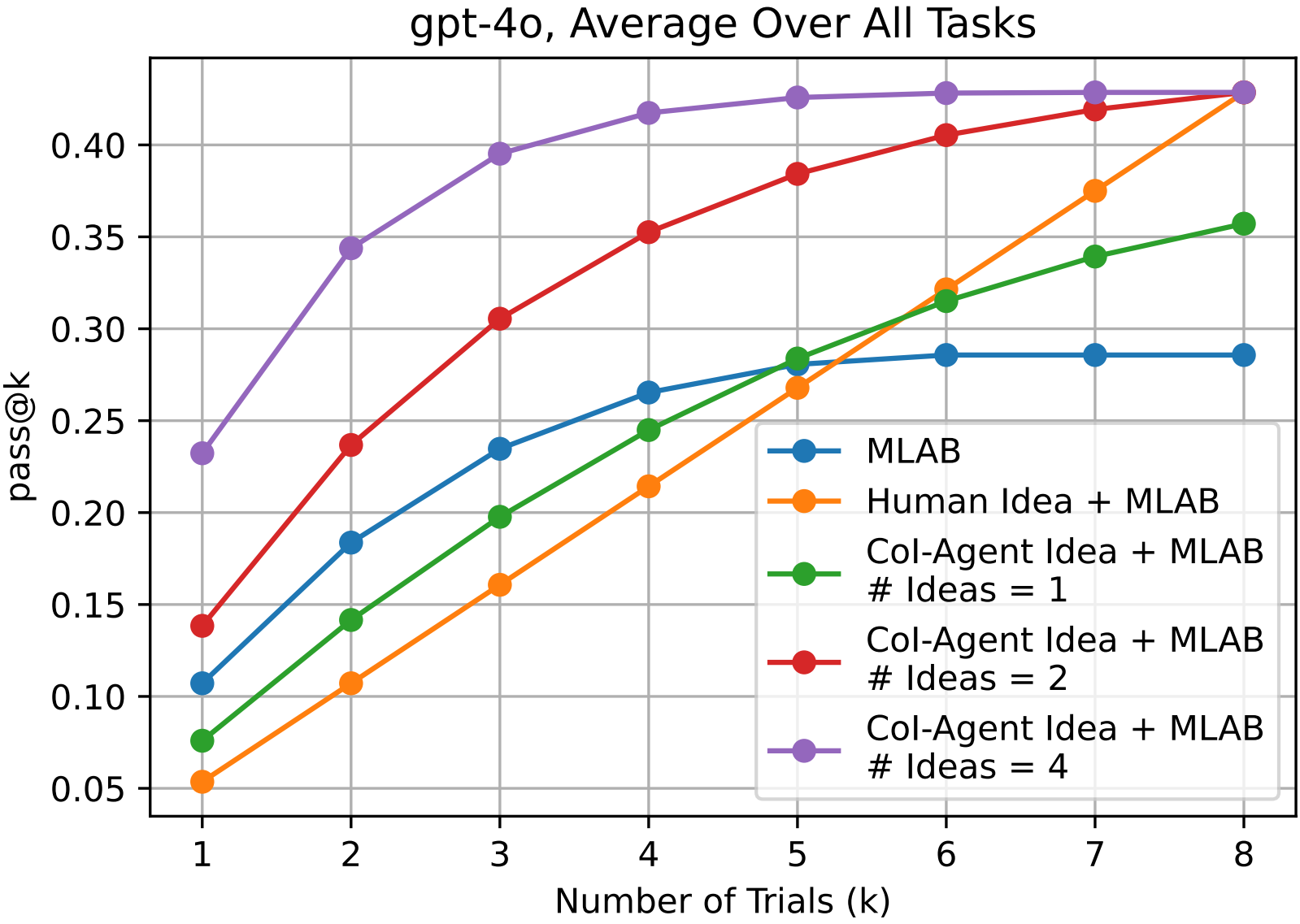

gpt-4o, Average Over All Tasks
pass@k
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
1
2
3
4
5
6
7
8
Number of Trials (k)
MLAB
Human Idea + MLAB
CoI-Agent Idea + MLAB
Ideas = 1
CoI-Agent Idea + MLAB
Ideas = 2
CoI-Agent Idea + MLAB
Ideas = 4