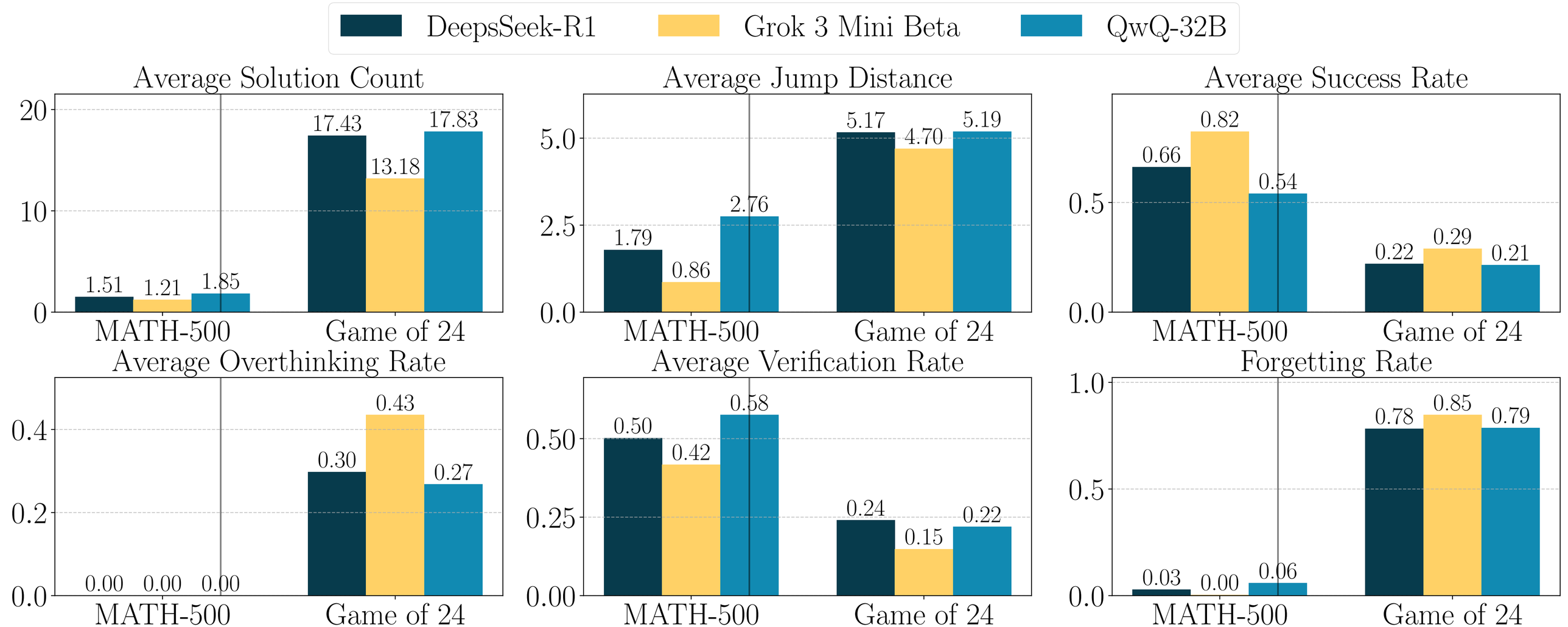
DeepsSeek-R1
Grok 3 Mini Beta
QwQ-32B
Average Solution Count
20
10
0
1.51
1.21
1.85
17.43
13.18
17.83
MATH-500
Game of 24
Average Jump Distance
5.0
2.5
0.0
1.79
0.86
2.76
5.17
4.70
5.19
MATH-500
Game of 24
Average Success Rate
0.5
0.0
0.66
0.82
0.54
0.22
0.29
0.21
MATH-500
Game of 24
Average Overthinking Rate
0.4
0.2
0.0
0.00
0.00
0.00
0.30
0.43
0.27
MATH-500
Game of 24
Average Verification Rate
0.50
0.25
0.00
0.50
0.42
0.58
0.24
0.15
0.22
MATH-500
Game of 24
Forgetting Rate
1.0
0.5
0.0
0.03
0.00
0.06
0.78
0.85
0.79
MATH-500
Game of 24