

MATH-500

Tree Similarity

DeepSeek-R1

1.00

0.72

Grok 3 mini beta

0.72

1.00

Walk Similarity

1.00

0.87

0.87

1.00

Game of 24

DeepSeek-R1

1.00

0.45

Grok 3 mini beta

0.45

1.00

1.00

0.92

0.92

1.00

DeepSeek-R1

Grok 3 mini beta

DeepSeek-R1

Grok 3 mini beta