

Benchmark: MMLU Elementary Maths
Performance Metric: Target Greedily Decoded
Correlation Metric: Kendall

