

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

