

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

