

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

