

Benchmark: MMLU Management
Performance Metric: Target Greedily Decoded
Correlation Metric: Spearman

