

Benchmark: MMLU Logical Fallacies
Performance Metric: Accuracy
Correlation Metric: Kendall

