

Benchmark: MMLU Philosophy
Performance Metric: Brier Score
Correlation Metric: Kendall

