

Benchmark: MMLU Moral Scenarios
Performance Metric: Brier Score
Correlation Metric: Kendall

