

Benchmark: MMLU Security Studies
Performance Metric: Brier Score
Correlation Metric: Kendall

