

Benchmark: MMLU Conceptual Physics
Performance Metric: Brier Score
Correlation Metric: Kendall

