

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

