

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

