

Benchmark: MMLU Professional Medicine
Performance Metric: Brier Score
Correlation Metric: Kendall

