

Benchmark: MMLU Machine Learning
Performance Metric: Brier Score
Correlation Metric: Kendall

