

Benchmark: MMLU Abstract Algebra
Performance Metric: Brier Score
Correlation Metric: Kendall

