

Benchmark: MMLU Prehistory
Performance Metric: Brier Score
Correlation Metric: Pearson

