

Benchmark: MMLU US Foreign Policy  
Performance Metric: Brier Score  
Correlation Metric: Pearson

