

Benchmark: MMLUHS US History  
Performance Metric: Brier Score  
Correlation Metric: Kendall

