

Benchmark: MMLU Prehistory  
Performance Metric: Accuracy  
Correlation Metric: Kendall

