

Benchmark: MMLU Anatomy  
Performance Metric: Brier Score  
Correlation Metric: Kendall

