

Benchmark: MMLU Clinical Knowledge  
Performance Metric: Brier Score  
Correlation Metric: Kendall

