

Benchmark: MMLU Formal Logic  
Performance Metric: Brier Score  
Correlation Metric: Spearman

