

Benchmark: MC-TACO  
Performance Metric: Brier Score  
Correlation Metric: Spearman

