

Benchmark: Winogrande  
Performance Metric: Brier Score  
Correlation Metric: Pearson

