

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

