

Benchmark: PubMedQA  
Performance Metric: Brier Score  
Correlation Metric: Spearman

