

Benchmark: Social Interaction QA
Performance Metric: Brier Score
Correlation Metric: Kendall

