

Benchmark: ARC-Challenge
Performance Metric: Brier Score
Correlation Metric: Pearson

