

Benchmark: ARC-Easy
Performance Metric: Brier Score
Correlation Metric: Spearman

