

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

