Benchmark Task: arc_challenge

Pearson correlation: 0.14