

Benchmark: ARC-Easy
Performance Metric: Target Greedily Decoded
Correlation Metric: Kendall

