

Benchmark: ARC-Challenge
Performance Metric: Target Greedily Decoded
Correlation Metric: Spearman

