

Benchmark: ARC-Challenge  
Performance Metric: Accuracy (Normalized)  
Correlation Metric: Kendall

