

Benchmark: MMLU Conceptual Physics  
Performance Metric: Target Greedily Decoded  
Correlation Metric: Kendall

