

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

