

Benchmark: MMLU Econometrics  
Performance Metric: Target Greedily Decoded  
Correlation Metric: Kendall

