

Benchmark: MMLU Philosophy  
Performance Metric: Target Greedily Decoded  
Correlation Metric: Spearman

