

Benchmark: MMLU Formal Logic  
Performance Metric: Target Greedily Decoded  
Correlation Metric: Spearman

