

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

