

Benchmark: MMLUComp Security  
Performance Metric: Target Greedily Decoded  
Correlation Metric: Pearson

