Benchmark: MMLU Security Studies
Performance Metric: Target Greedily Decoded
Correlation Metric: Kendall