Benchmark: MMLU Moral Scenarios
Performance Metric: Target Greedily Decoded
Correlation Metric: Pearson