Benchmark: MMLU HS Comp Sci
Performance Metric: Target Greedily Decoded
Correlation Metric: Pearson