Benchmark: MMLU Electrical Engineering
Performance Metric: Target Greedily Decoded
Correlation Metric: Pearson