

Benchmark: MMLU Abstract Algebra
Performance Metric: Target Greedily Decoded
Correlation Metric: Pearson

