

Benchmark: MMLU Astronomy  
Performance Metric: Target Greedily Decoded  
Correlation Metric: Pearson

