

Benchmark: MMLUCollege Medicine  
Performance Metric: Target Greedily Decoded  
Correlation Metric: Spearman

