

Benchmark: MMLU Professional Medicine
Performance Metric: Target Greedily Decoded
Correlation Metric: Kendall

