

Benchmark: MMLU Clinical Knowledge
Performance Metric: Target Greedily Decoded
Correlation Metric: Kendall

