

Benchmark: MMLU Medical Genetics
Performance Metric: Target Greedily Decoded
Correlation Metric: Kendall

