

Benchmark: MMLU US Foreign Policy
Performance Metric: Target Greedily Decoded
Correlation Metric: Kendall

