

Benchmark: MMLU International Law
Performance Metric: Target Greedily Decoded
Correlation Metric: Spearman

