

Benchmark: MMLU HS Euro History
Performance Metric: Target Greedily Decoded
Correlation Metric: Spearman

