

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

