

Benchmark: MMLU HS Geography  
Performance Metric:  $\log p_{\theta}^{\text{Vocab}}$  (Correct Choice)  
Correlation Metric: Kendall

