

MMLU Human Sexuality
 $\log p_{\theta}^{\text{Vocab}}(\text{Correct Choice})$

Per-Sample Correlation(Compute, Score) for
LLM360 Amber 7B Tokens (Param Scaling)

