

MMLU Human Sexuality

$p_{\theta}^{\text{Choices}}$ (Correct Choice)

Per-Sample Correlation(Compute, Score) for
Pythia 300B Tokens (Param. Scaling)

