

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

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

