

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

