

MMLU Moral Disputes

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

