MMLU Moral Disputes Compute Model Family: Cerebras (Param. and Data Scaling) Model Family: Pythia 300B Tokens (Param. Scaling) Model Family: Pythia 12B Param. (Data Scaling) 10^{0} 10^{0} 10^{0} 90 10⁻¹ 10^{-1} 10^{-1} 10^{-2} 10^{-2} 10^{-3} 10^{-3} ຽ້ 10^{-4} 10^{-4} 10^{-4} $\sqrt{10^{-5}}$ 10^{-5} 10^{-5} 10^{-6} 10^{-6} 10^{-5} 10^{-3} 10^{-2} 10^{-5} 10^{-3} 10^{-2} 10^{-6} 10^{-4} 10^{-1} 10^{-6} 10^{-1} 10^{-6} 10^{-4} 10^{-1}

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

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

• 1e+18 1e + 19

1e + 201e + 21

1e + 22

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