MMLU College Comp Sci Compute • 1e+18 Model Family: Cerebras (Param. and Data Scaling) Model Family: Pythia 12B Param. (Data Scaling) Model Family: Pythia 300B Tokens (Param. Scaling) 10^{0} 10^{0} 10^{0} (lucorrect Choice) 10^{-1} 10^{-2} 10^{-3} 10^{-} 10^{-1} 10^{-2} 10^{-2} 10^{-3} 10^{-3} $\frac{5}{6} \cdot 10^{-4}$ 10^{-4} 10^{-5} 10^{-5} 10^{-6} 10^{-6} 10^{-2} 10^{-2} 10^{-6} 10^{-5} 10^{-1} 10^{0} 10^{-6} 10^{-3} 10^{-1} 10^{-5} 10^{-4} 10^{-1}

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

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

1e + 19

1e + 201e + 21

1e + 22

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