## MMLU Comp Security Compute 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) $10^{-1} 10^{-2}$ $10^{-3}$ $10^{-3}$ $10^{-1}$ $10^{-1}$ $10^{-2}$ $10^{-2}$ $10^{-3}$ $10^{-3}$ ••••• ಕ<del>್ಕೆ</del> 10<sup>−4</sup> $10^{-4}$ $10^{-4}$ $\sqrt{10^{-5}}$ $10^{-5}$ $10^{-5}$ $10^{-6}$ $10^{-6}$

 $10^{-3}$ 

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

 $10^{-2}$ 

 $10^{-1}$ 

 $10^{-5}$ 

 $10^{-4}$ 

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

 $10^{-2}$ 

 $10^{-1}$ 

 $10^{-5}$ 

 $10^{-4}$ 

 $10^{-6}$ 

 $10^{-3}$ 

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

 $10^{-2}$ 

 $10^{-1}$ 

 $10^{-6}$ 

## • 1e+18 1e + 19

1e + 20

1e + 211e + 22