MMLU Marketing 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} Choice 10^{-1} 10^{-1} 10^{-1} 10^{-2} 10^{-3} 10^{-3} 10^{-2} 10^{-2} 10^{-3} 10^{-3} $\frac{10^{-4}}{5}$ 10^{-4} 10^{-4} $\sum_{\frac{2}{5}} 10^{-5}$ 10^{-5} 10^{-5} 10^{-6} 10^{-6} 10^{-2} 10^{-1} 10^{-6} 10^{-2} 10^{-1} 10^{-5} 10^{-1} $p_{\theta}^{\mathsf{Vocab}}(\mathsf{Correct\ Choice})$ $p_{\theta}^{\mathsf{Vocab}}(\mathsf{Correct\ Choice})$ $p_{\theta}^{\mathsf{Vocab}}(\mathsf{Correct\ Choice})$

• 1e+18 1e + 19

1e + 20

1e + 211e + 22