MMLU Elementary Maths $\textstyle\sum_{\mathsf{Incorrect\ Choice}} p_{\theta}^{\mathsf{Vocab}}(\mathsf{Incorrect\ Choice})$ $p_{\theta}^{\mathsf{Vocab}}(\mathsf{Correct\ Choice})$ 10^{0} 10^{-1} Probability Mass 10^{-3} M $^{-3}$ 10^{-4} 10^{-5}

 10^{17}

 10^{18}

 10^{19}

 10^{20}

Compute (6ND)

 10^{21}

 10^{22}

 10^{22}

 10^{18}

 10^{17}

 10^{19}

 10^{20}

Compute (6ND)

 10^{21}

Cerebras (Param. and Data Scaling)
Pythia 12B Param. (Data Scaling)
Pythia 300B Tokens (Param. Scaling)

Model Family