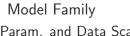
MC-TACO $p_{\theta}^{\mathsf{Vocab}}(\mathsf{Correct\ Choice})$ — Cerebras (Param. and Data Scaling) Correlation: spearman Correlation: pearson Correlation: kendall — OLMo 7B Param. (Data Scaling) 2.5— Pythia 12B Param. (Data Scaling) — Pythia 300B Tokens (Param. Scaling) 2.0 1.0 0.50.751.00 -1.000.750.751.00 1.00 -1.000.00Correlation Between FLOPs and Scores (Per Sample) Correlation Between FLOPs and Scores (Per Sample)

Correlation Between FLOPs and Scores (Per Sample)



— INCITE 7B Param. (Data Scaling)

LLM360 Amber 7B Tokens (Param Scaling)