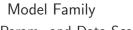
MMLU Management $p_{\theta}^{\mathsf{Choices}}(\mathsf{Correct\ Choice})$ Correlation: spearman Correlation: pearson Correlation: kendall LLM360 Amber 7B Tokens (Param Scaling) 2.00 1.751.50 .S 1.25 $\stackrel{\mathsf{o}}{\mathsf{\Box}} 1.00$ 0.750.501.00 1.00 -1.00-1.00Correlation Between FLOPs and Scores (Per Sample) Correlation Between FLOPs and Scores (Per Sample) Correlation Between FLOPs and Scores (Per Sample)



— Cerebras (Param. and Data Scaling)

— INCITE 7B Param. (Data Scaling)

— OLMo 7B Param. (Data Scaling)

— Pythia 12B Param. (Data Scaling)

— Pythia 300B Tokens (Param. Scaling)