PubMedQA $p_{\theta}^{\mathsf{Choices}}(\mathsf{Correct\ Choice})$ Model Family — Cerebras (Param. and Data Scaling) — INCITE 7B Param. (Data Scaling) Correlation: spearman Correlation: pearson Correlation: kendall — LLM360 Amber 7B Tokens (Param Scaling) — OLMo 7B Param. (Data Scaling) —— Pythia 12B Param. (Data Scaling) —— Pythia 300B Tokens (Param. Scaling) 0.80.21.00 -1.001.00 -1.00Correlation Between FLOPs and Scores (Per Sample) Correlation Between FLOPs and Scores (Per Sample) Correlation Between FLOPs and Scores (Per Sample)