ARC-Easy  $p_{\theta}^{\mathsf{Choices}}(\mathsf{Correct\ Choice})$ Correlation: spearman Correlation: pearson Correlation: kendall — OLMo 7B Param. (Data Scaling) 1.75— Pythia 12B Param. (Data Scaling) 1.50 1.25nsity 1.00 0.750.500.251.00 1.00 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)

Model Family

— Cerebras (Param. and Data Scaling)

— INCITE 7B Param. (Data Scaling) LLM360 Amber 7B Tokens (Param Scaling)

— Pythia 300B Tokens (Param. Scaling)