MMLU HS Euro History $p_{\theta}^{\mathsf{Vocab}}(\mathsf{Correct\ Choice})$ Correlation: spearman Correlation: pearson Correlation: kendall 1.00 -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)

Model FamilyCerebras (Param. and Data Scaling)

— INCITE 7B Param. (Data Scaling)

LLM360 Amber 7B Tokens (Param Scaling)OLMo 7B Param. (Data Scaling)

— Pythia 12B Param. (Data Scaling)

— Pythia 300B Tokens (Param. Scaling)