Benchmark: MMLU Virology Performance Metric: Accuracy Correlation Metric: Spearman Model Family Cerebras (Param. and Data Scaling) —— INCITE 7B Param. (Data Scaling) 0.8 LLM360 Amber 7B Tokens (Param Scaling) — OLMo 7B Param. (Data Scaling) —— Pythia 12B Param. (Data Scaling) —— Pythia 300B Tokens (Param. Scaling) 0.6 CDF 0.4 0.20.0 1.00 -1.00-0.75-0.50-0.250.000.250.500.751.00 Correlation Between FLOPs and Scores (Per Sample)

1.2 1.0 8.0 % of Samples % 0.6 % 0.40.2

-0.75

-0.50

-0.25

0.00

Correlation Between FLOPs and Scores (Per Sample)

0.25

0.50

0.75