Benchmark: MMLU College Maths

Performance Metric:  $p_{\theta}^{Vocab}(Correct Choice)$ Correlation Metric: Kendall 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) 0.6CDF 0.4

