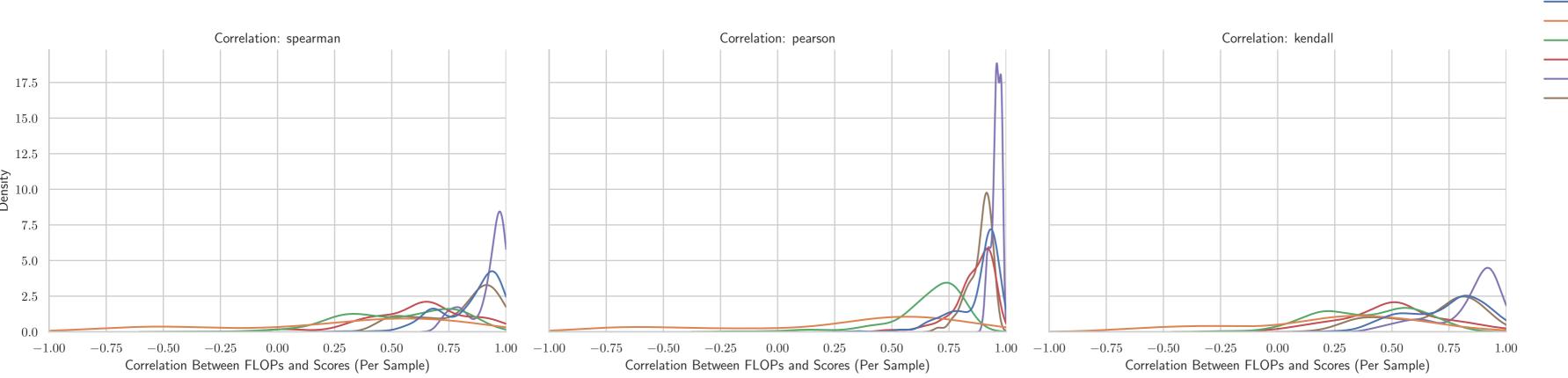
$\begin{array}{c} \mathsf{MMLU} \; \mathsf{Marketing} \\ \log p_{\theta}^{\mathsf{Vocab}}(\mathsf{Correct} \; \mathsf{Choice}) \end{array}$ 



Model Family

Cerebras (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)