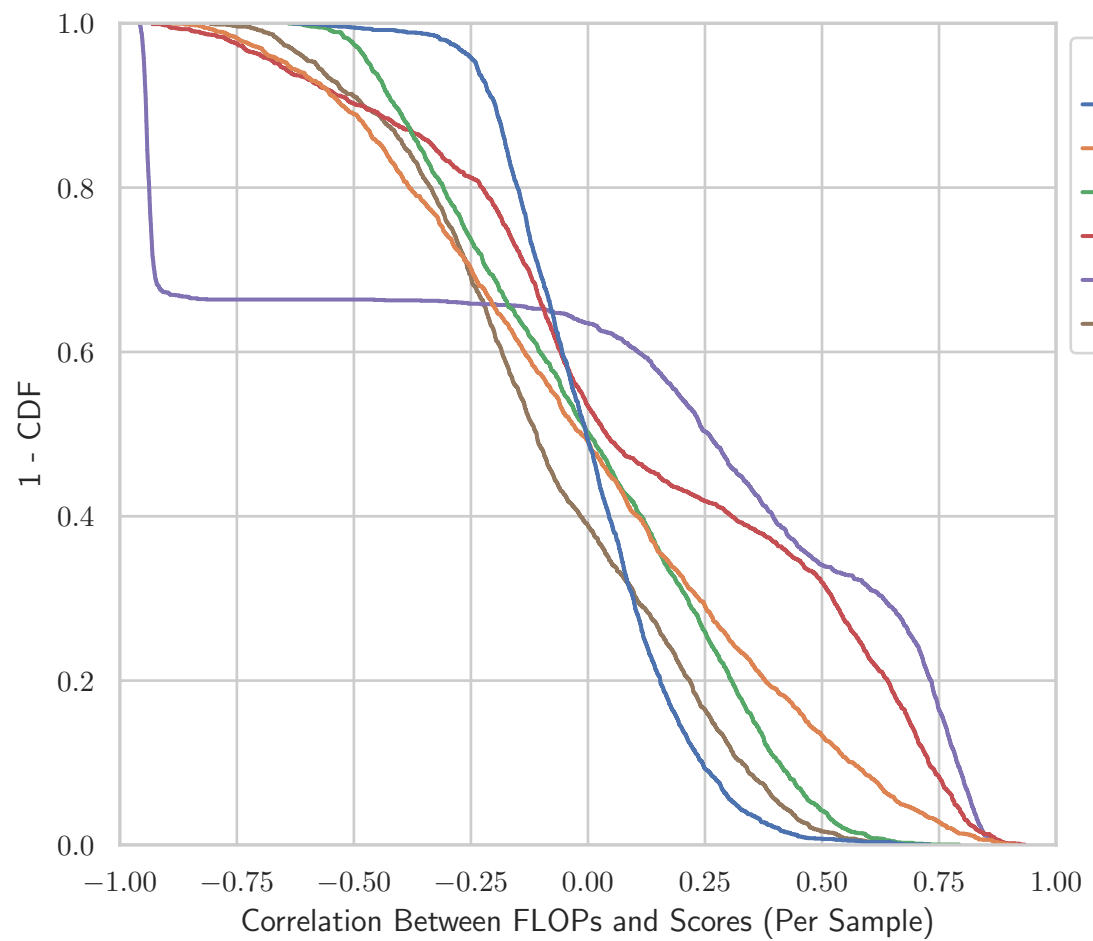
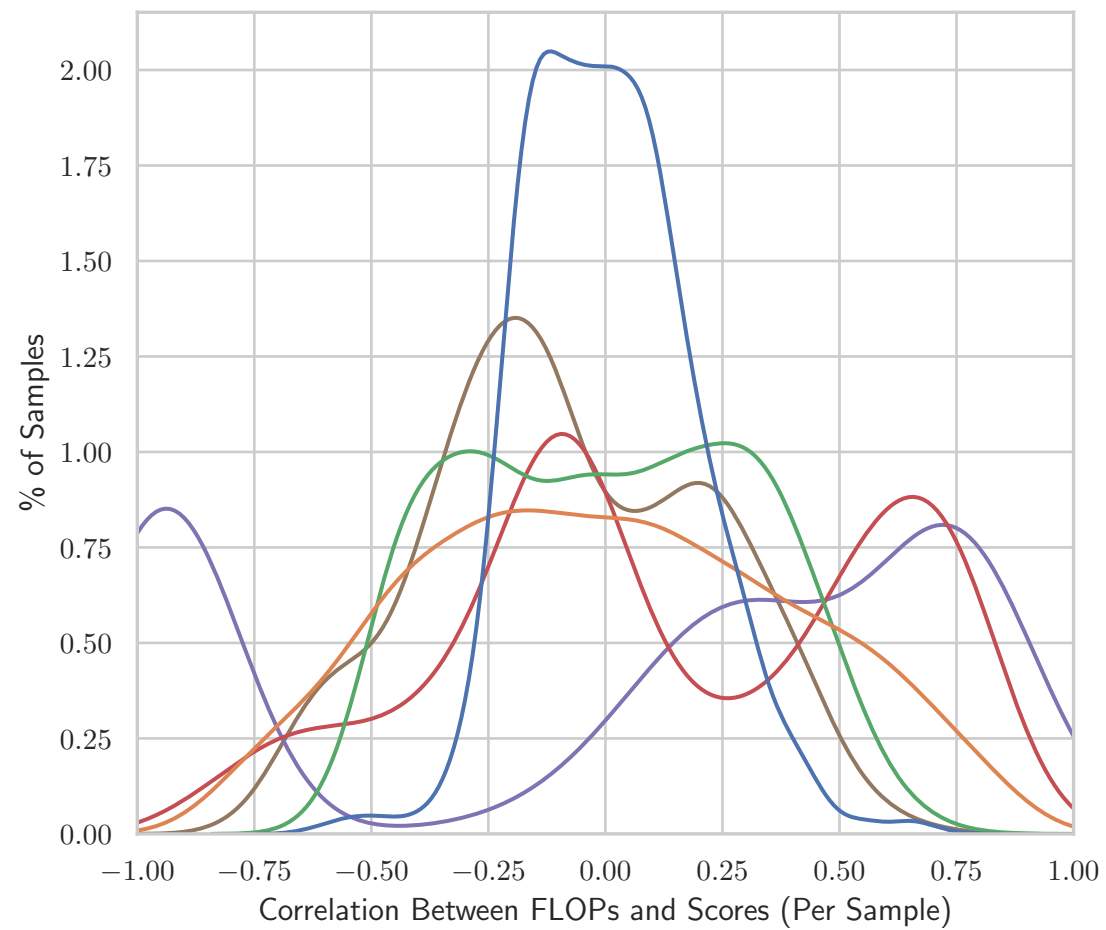


Benchmark: Social Interaction QA
Performance Metric: $p_{\theta}^{\text{Choices}}$ (Correct Choice)
Correlation Metric: Pearson



- 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)