Social Interaction QA
Target Greedily Decoded

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)

Correlation: spearman | Correlation: pearson | Correlation: kendall

Density

Correlation Between FLOPs and Scores (Per Sample)