

Benchmark (and Optional Task)

Correlation: kendall

Kurtosis(Correlations) by Benchmark and Metric

Correlation: pearson

Correlation: spearman

Kurtosis(Correlations)

Kurtosis(Correlations)

Kurtosis(Correlations)

- metric
- log\_prob\_vocab\_correct
  - prob\_vocab\_correct
  - prob\_choices\_correct
  - brier\_score
  - acc
- 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)