AUC of Correlations' Complementary Cumulative Distribution Function by Benchmark and Metric log_prob_vocab_correct Correlation: pearson Correlation: spearman prob_vocab_correct prob_choices_correct brier_score



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)