

MMLU Moral Disputes

Brier Score

Per-Sample Correlation(Compute, Score) for
Pythia 300B Tokens (Param. Scaling)

