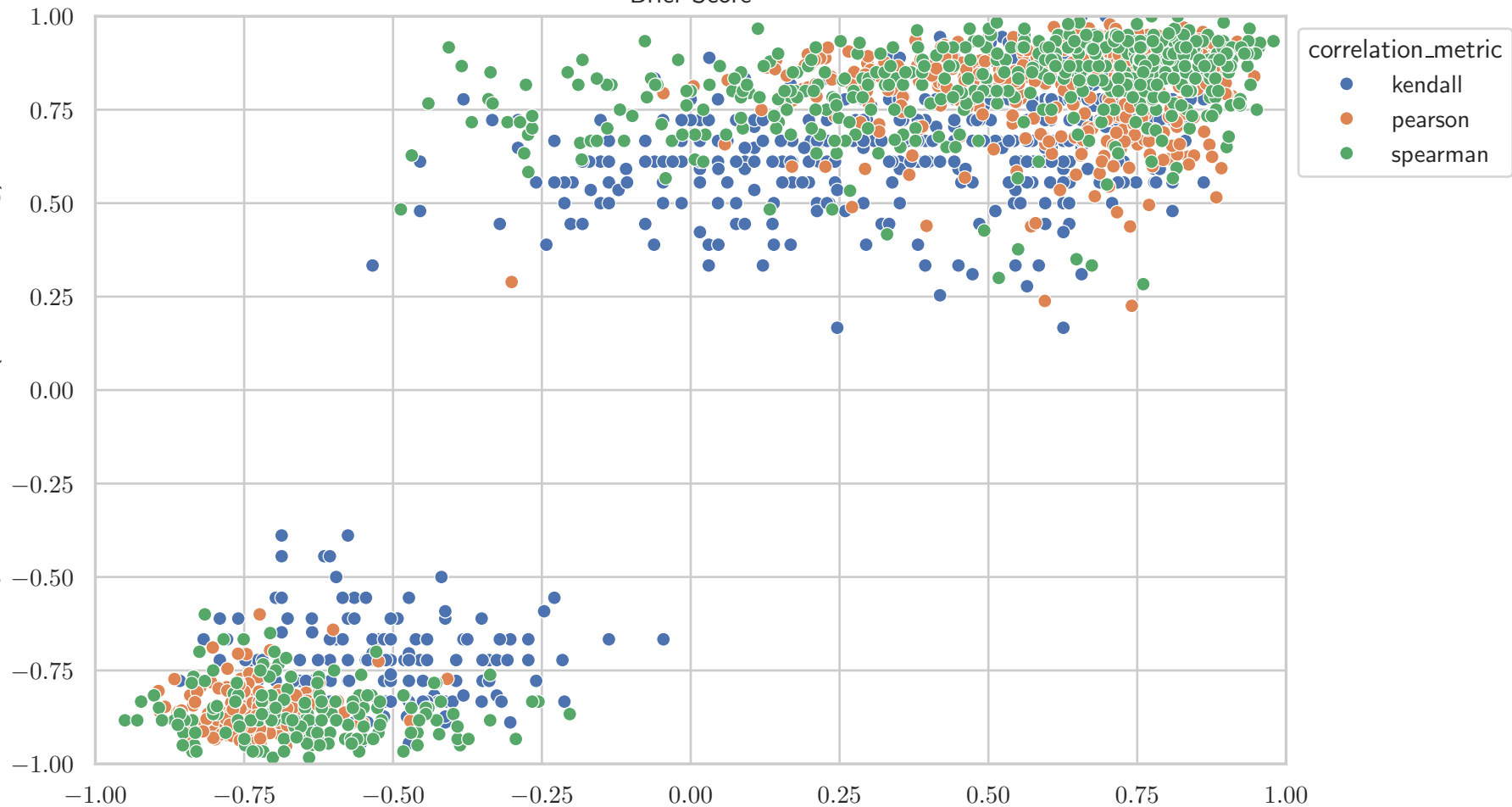


MMLU Professional Psychology

Brier Score

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



Per-Sample Correlation(Compute, Score) for
LLM360 Amber 7B Tokens (Param Scaling)