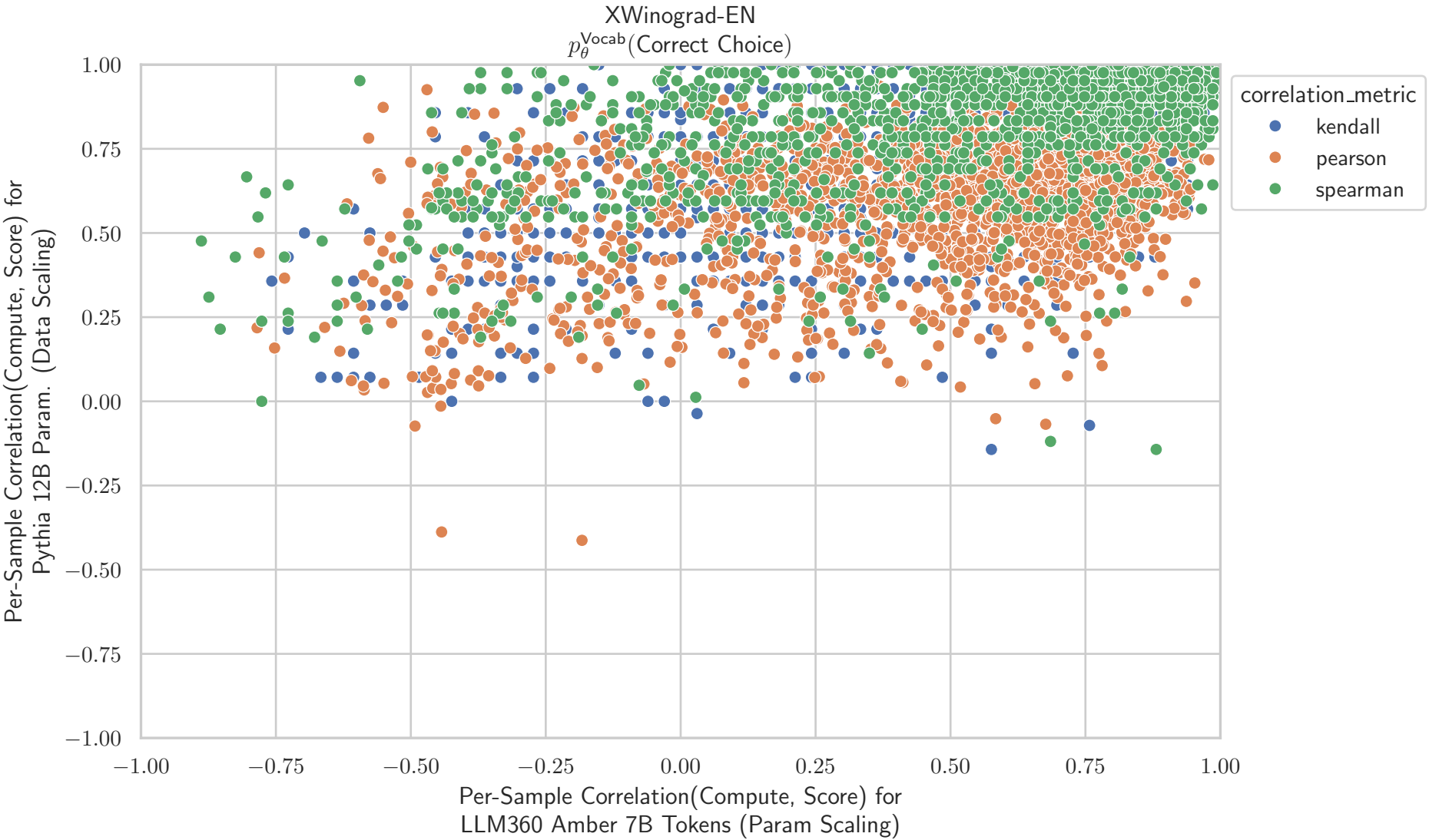
XWinograd-EN
p_θ^{Vocab}(Correct Choice)
Per-Sample Correlation(Compute, Score) for Pythia 12B Param. (Data Scaling)
Per-Sample Correlation(Compute, Score) for LLM360 Amber 7B Tokens (Param Scaling)
correlation_metric
kendall
pearson
spearman
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00