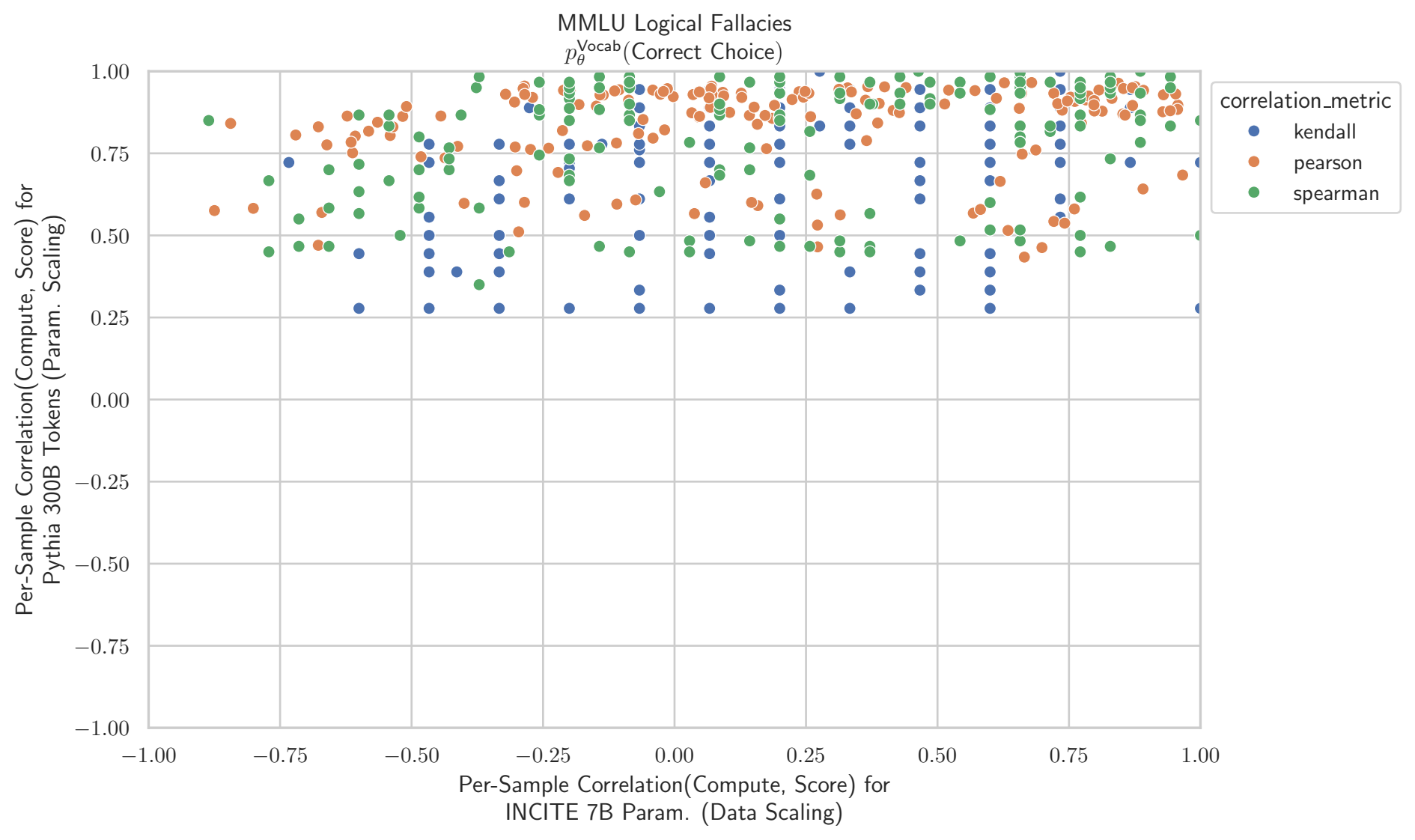
MMLU Logical Fallacies
p_θ^{Vocab}(Correct Choice)
Per-Sample Correlation(Compute, Score) for Pythia 300B Tokens (Param. Scaling)
Per-Sample Correlation(Compute, Score) for INCITE 7B Param. (Data Scaling)
correlation_metric
kendall
pearson
spearman
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00