

Syntax Accuracy vs LMArena Rank

Spearman $\rho = -0.689$, $p = 0.0045$ | Lower rank = better

