

## Syntax Accuracy vs LMArena Rank

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

