

LLM-Judge Score vs TrueSkill ELO

(Pearson + Spearman + Kendall τ , per Dimension)

