LLM-Judge Score vs TrueSkill ELO
(Pearson + Spearman; ELO from rep20 overall)