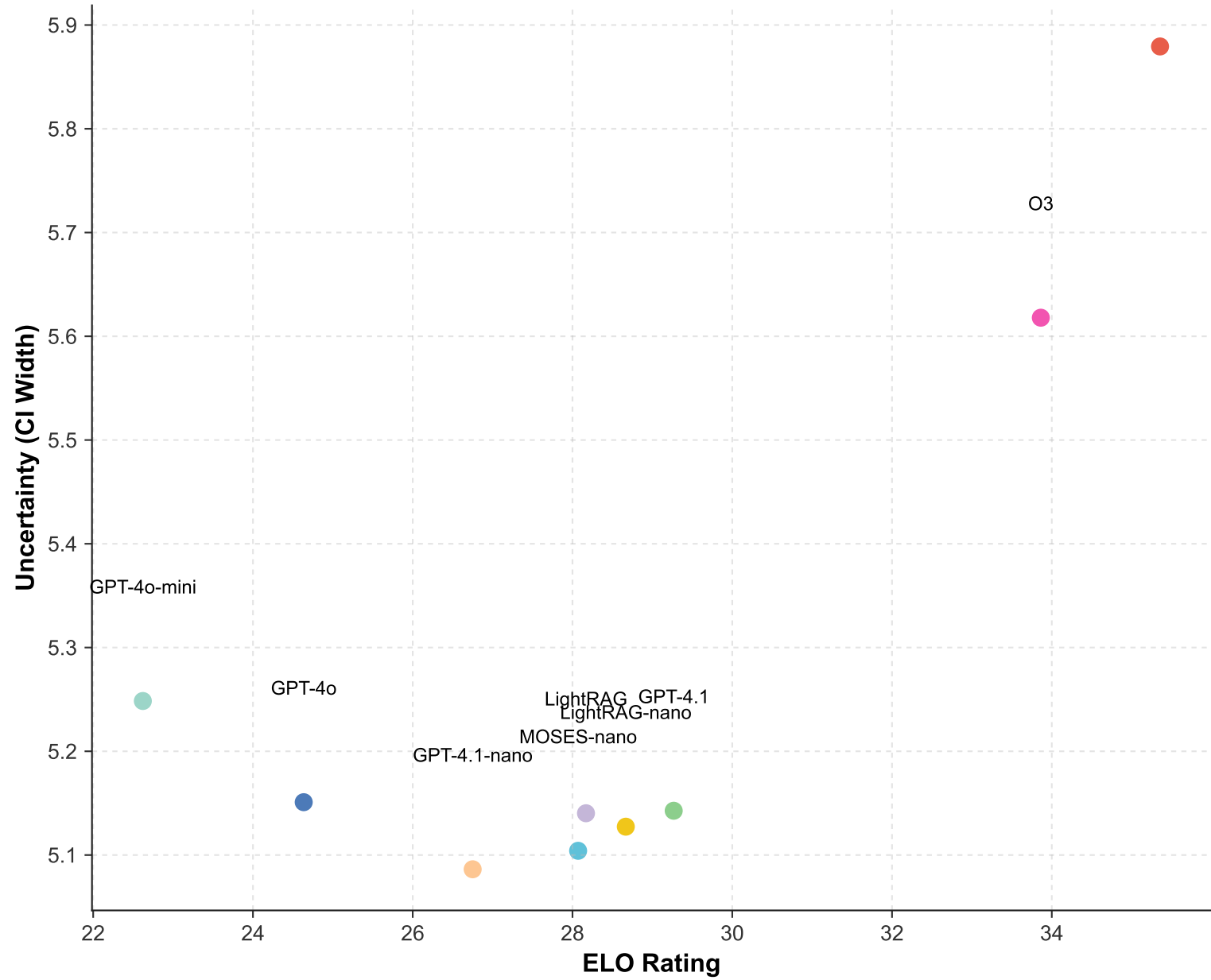


TrueSkill ELO Uncertainty Analysis  
(Model Performance Confidence Assessment)

MOSES

ELO Rating vs. Uncertainty



Model Ranking Uncertainty

