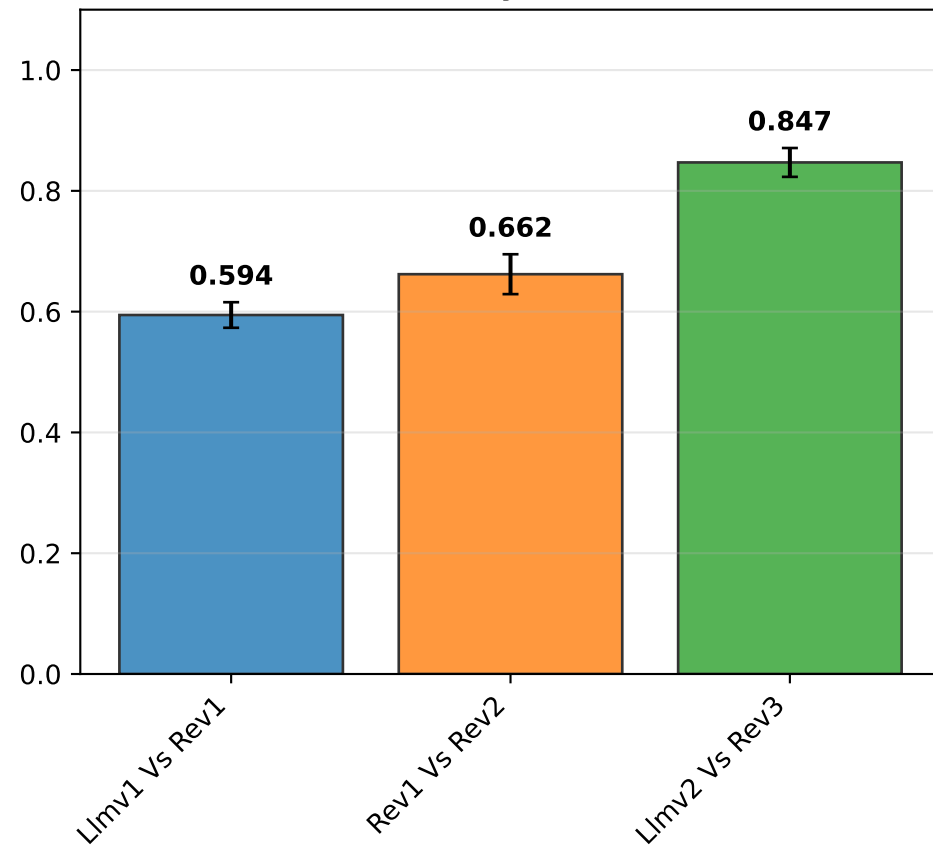
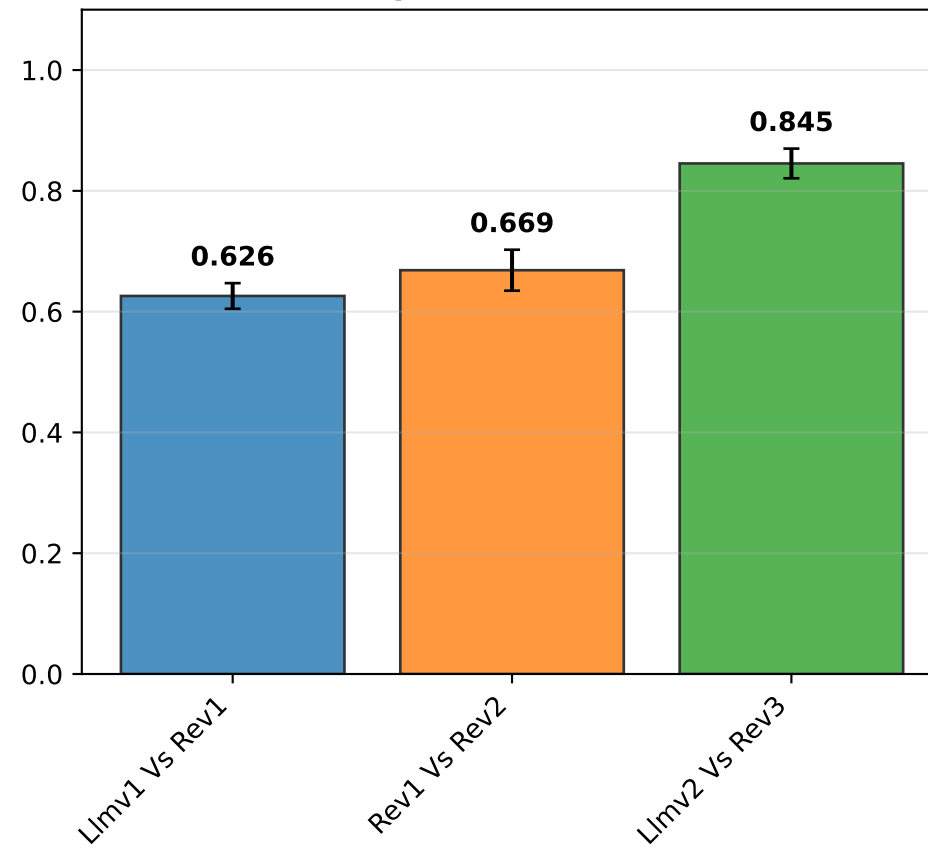


Overall Average Performance Across All Question Types

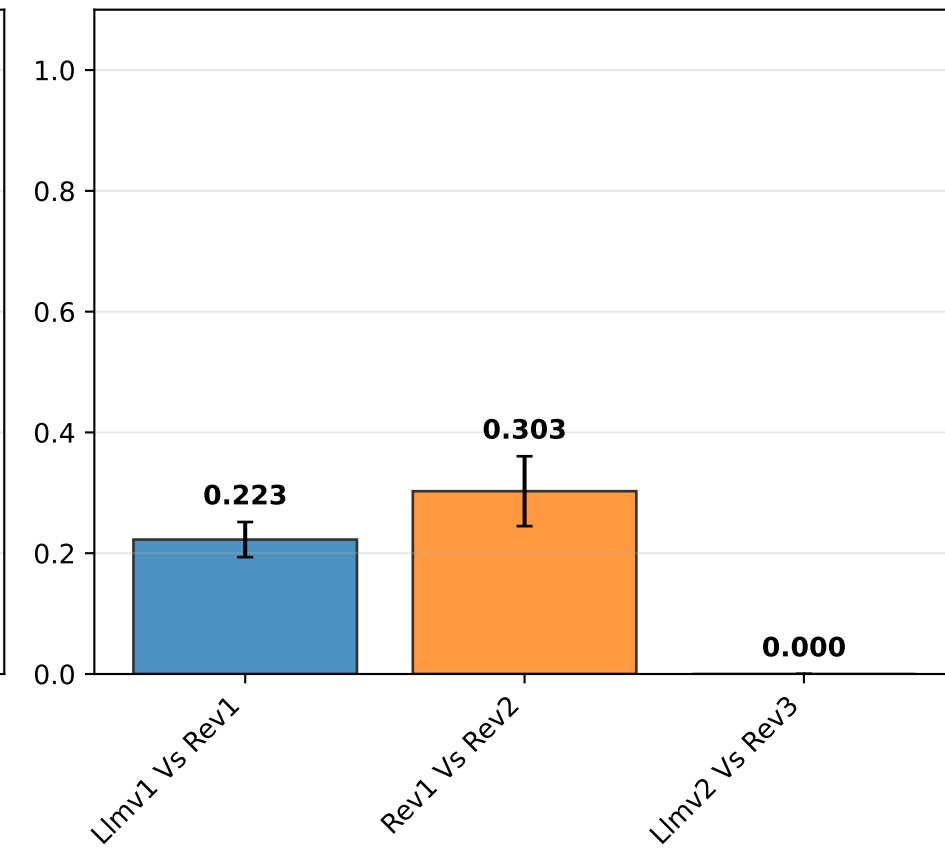
Accuracy [GEval]



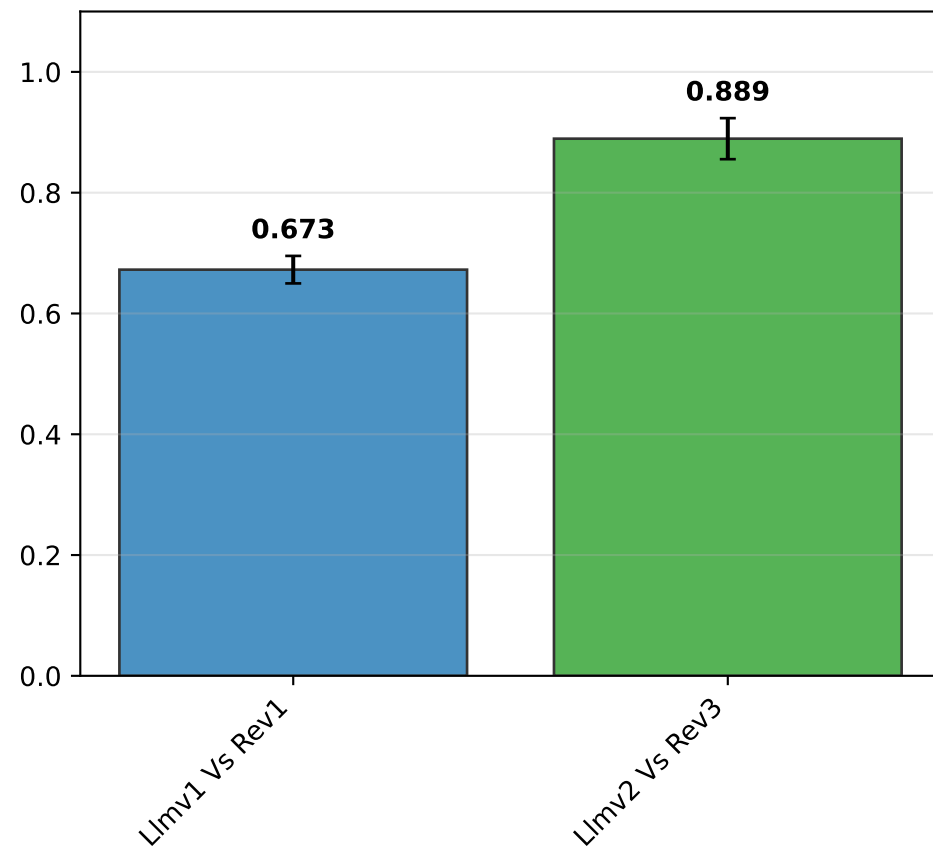
Completeness [GEval]



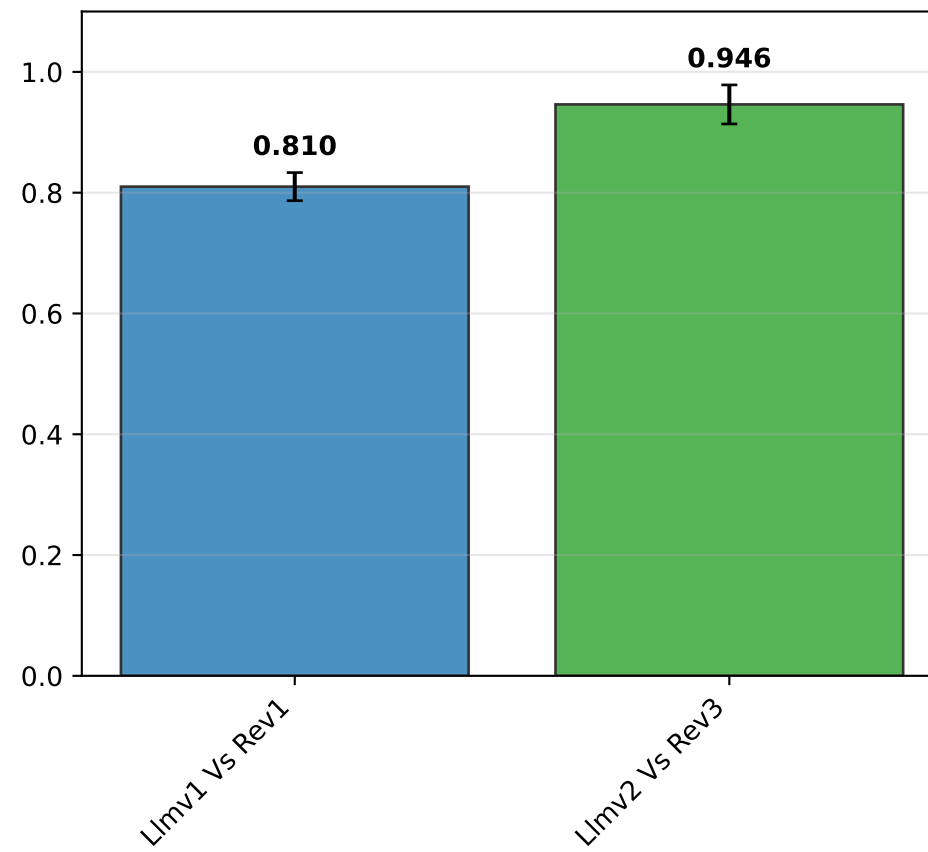
Correctness [GEval]



Contextual Precision



Contextual Recall



Faithfulness

