

Multi-Turn RAG Benchmark Report

Generated: 2025-06-27 19:24

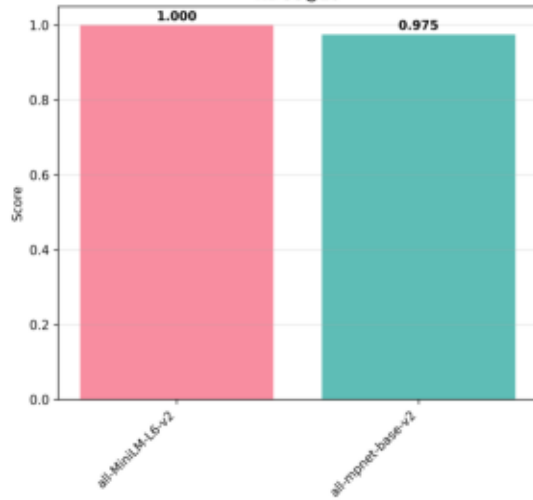
Models Tested: 2

Datasets: 1

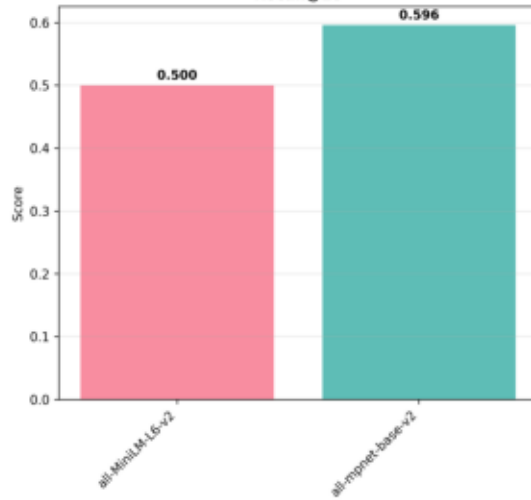
Total Evaluations: 2

Performance Overview

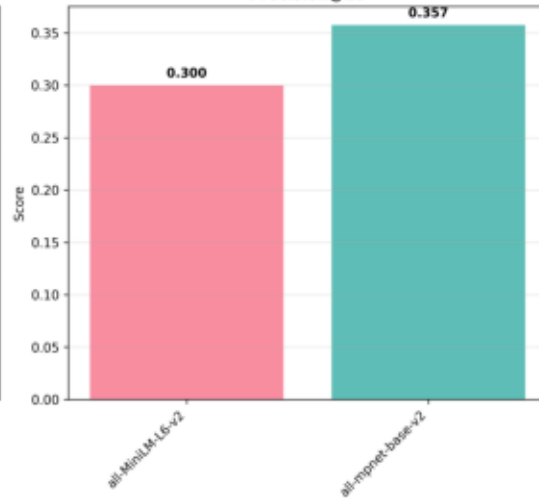
NDCG@10



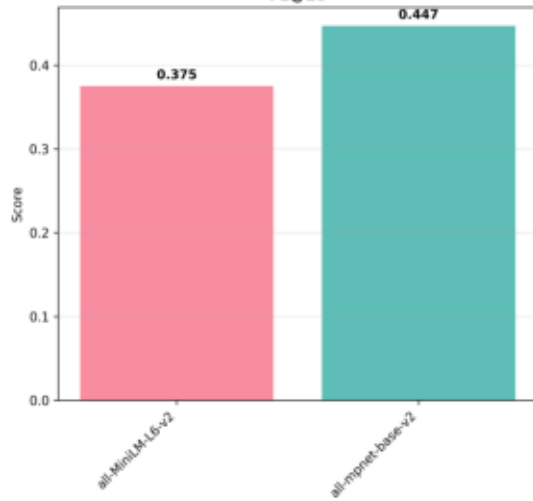
Recall@10



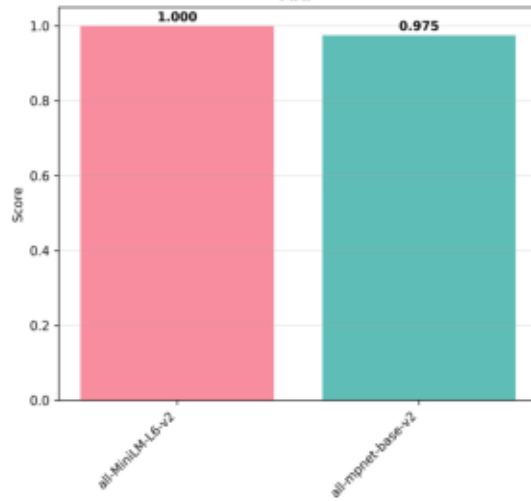
Precision@10



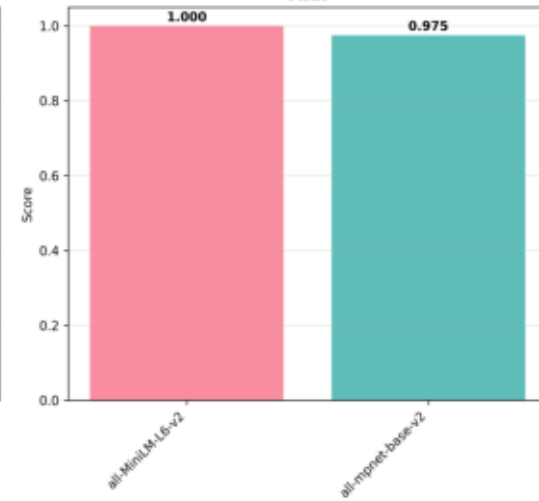
F1@10



MAP

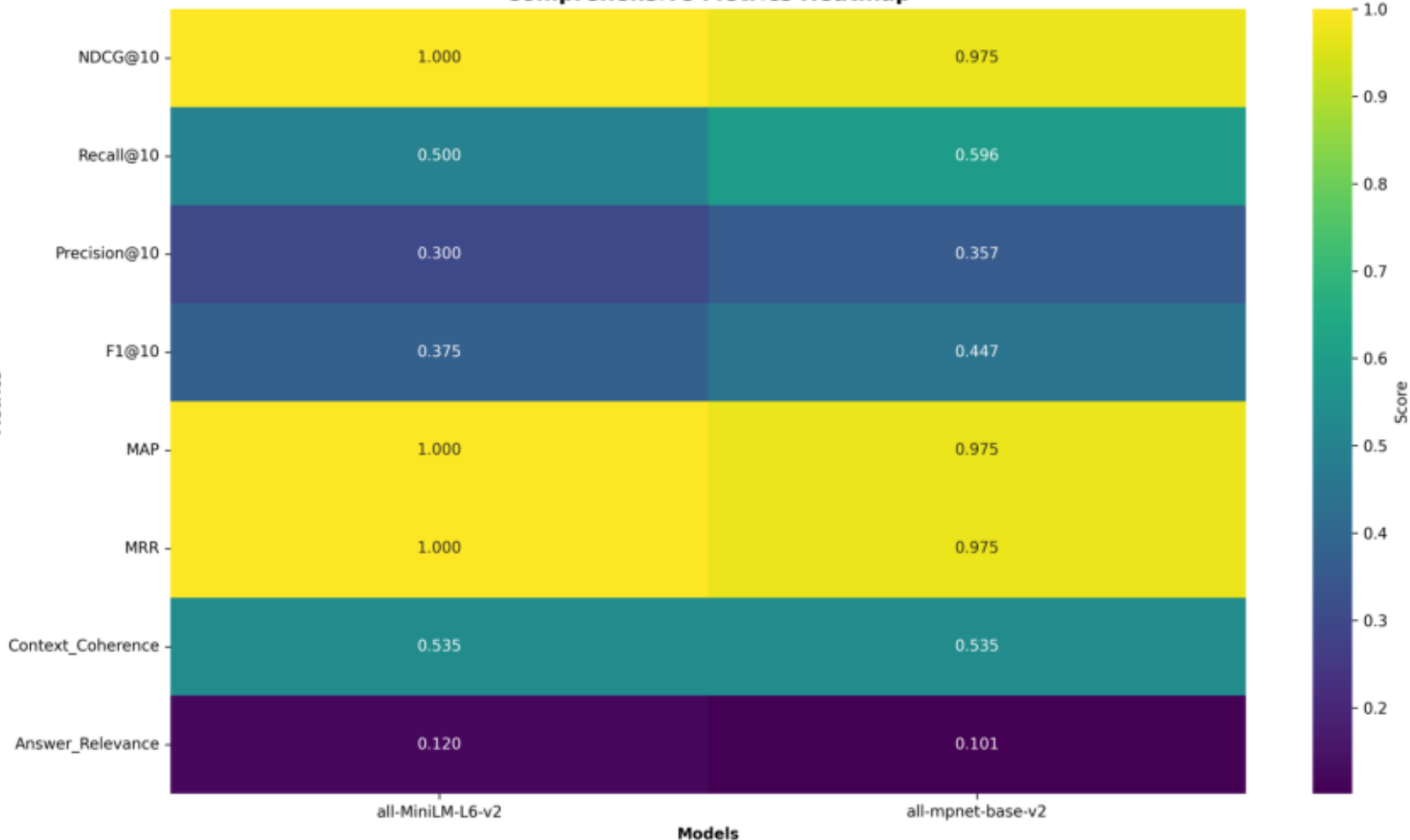


MRR

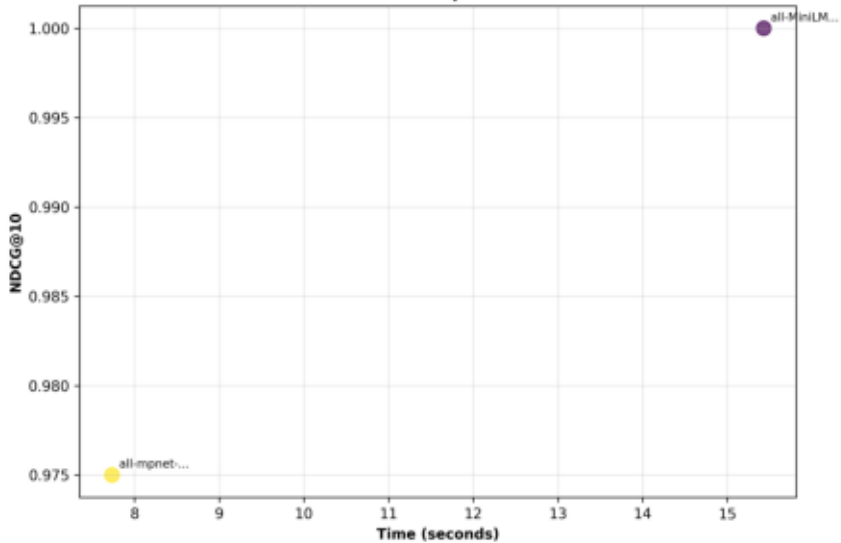


Comprehensive Metrics Heatmap

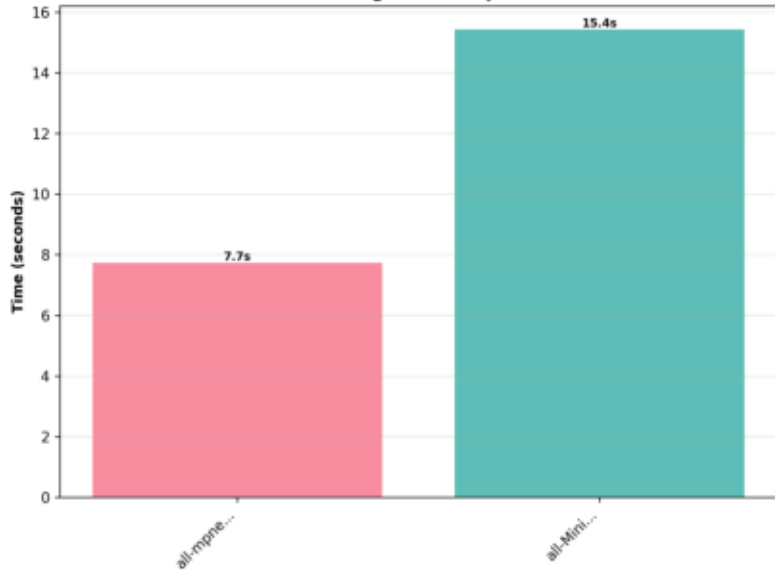
Metrics



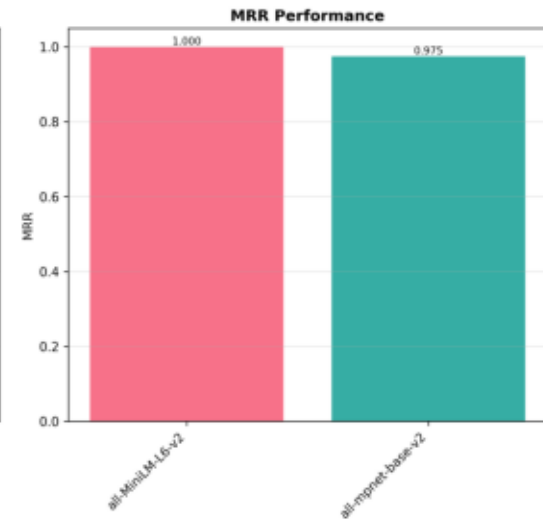
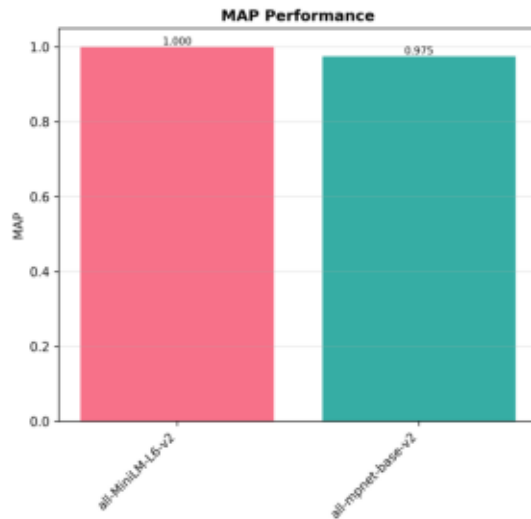
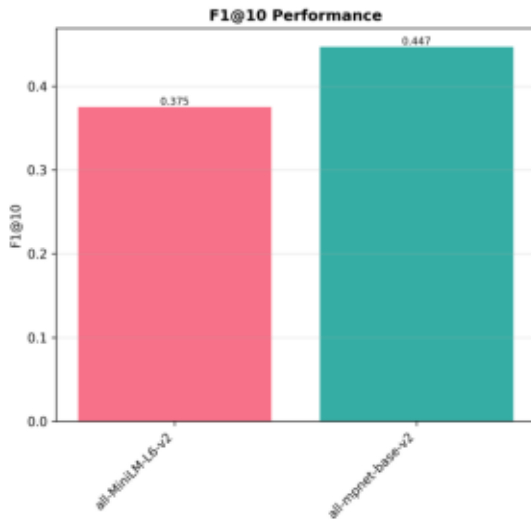
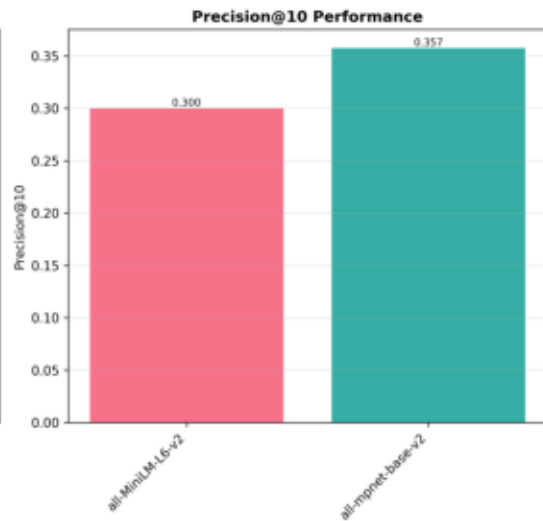
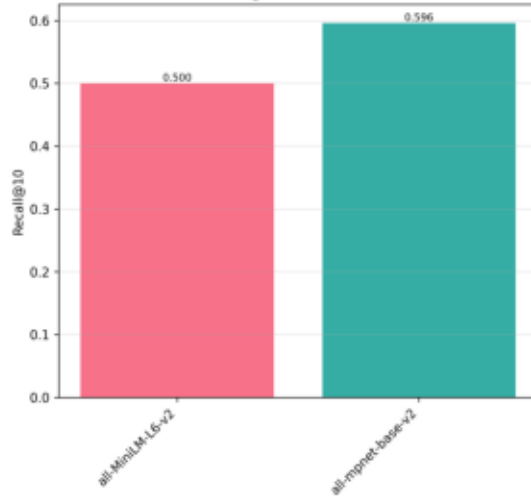
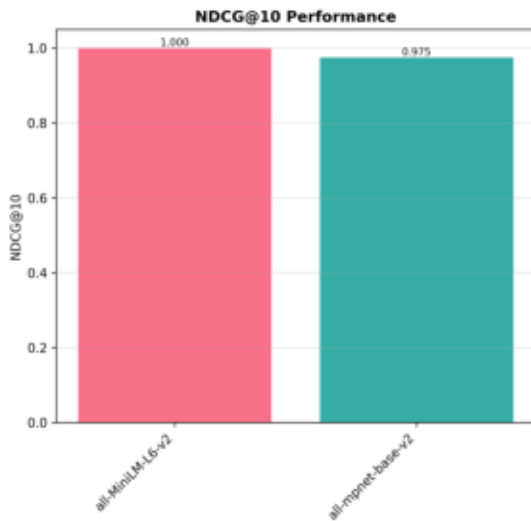
Performance vs Speed Trade-off



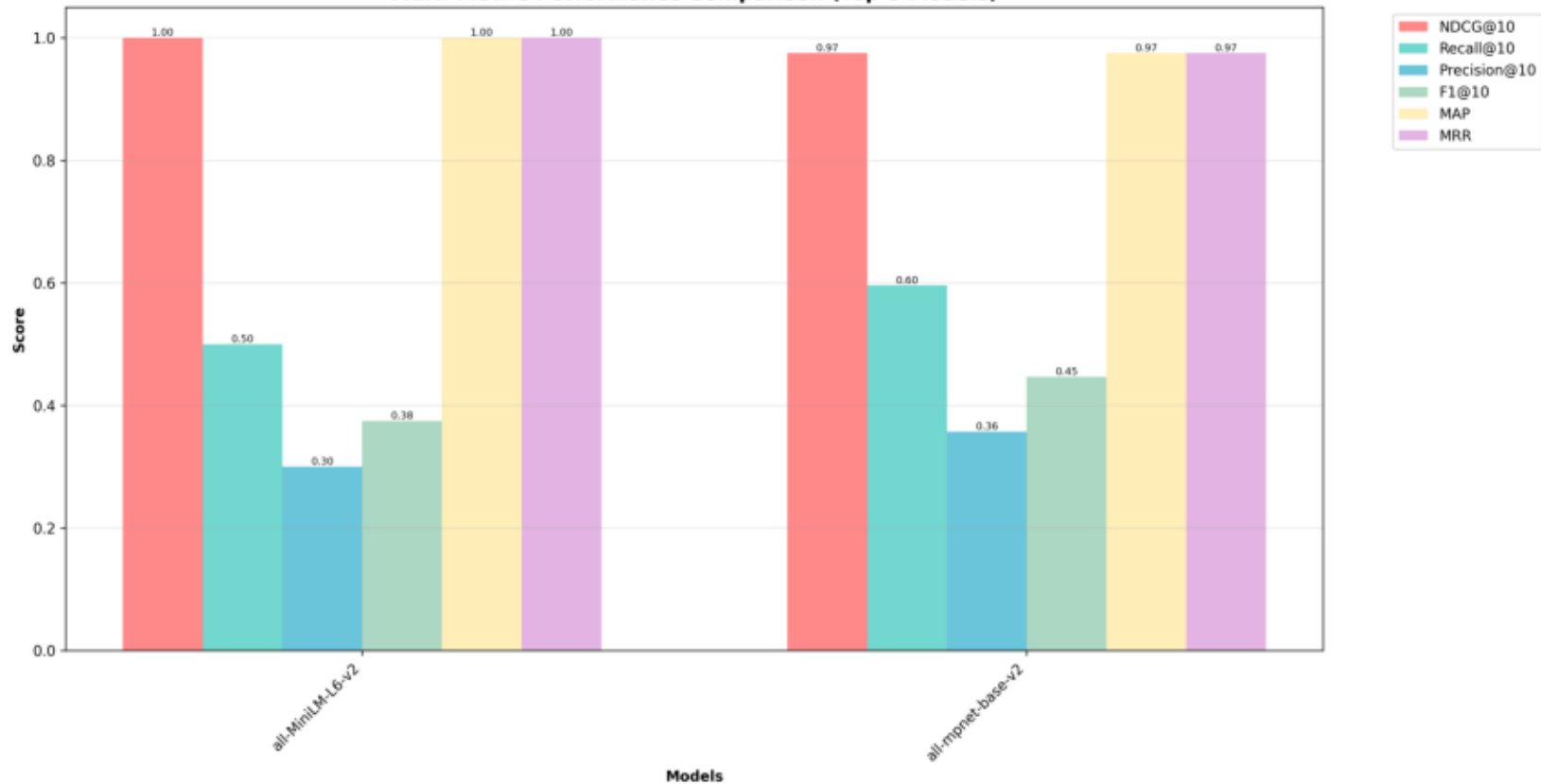
Processing Time Comparison



Individual Metric Performance Comparison



Multi-Metric Performance Comparison (Top 8 Models)



Top Performers by Metric

