Ultimate RAG Benchmark Report

Generated: 2025-06-27 19:30

Models Tested: 2

Datasets: 2

Corpora: 1

Total Evaluations: 4

Key Results:
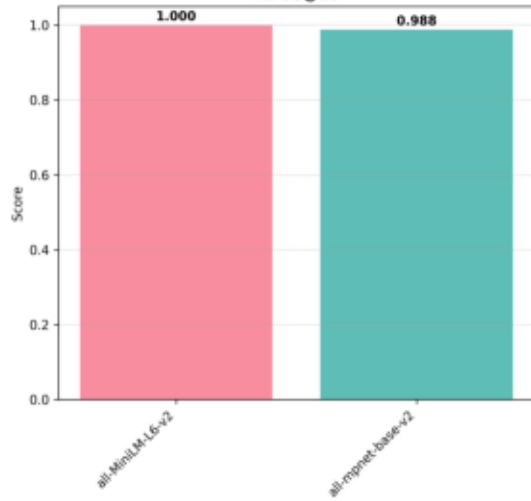
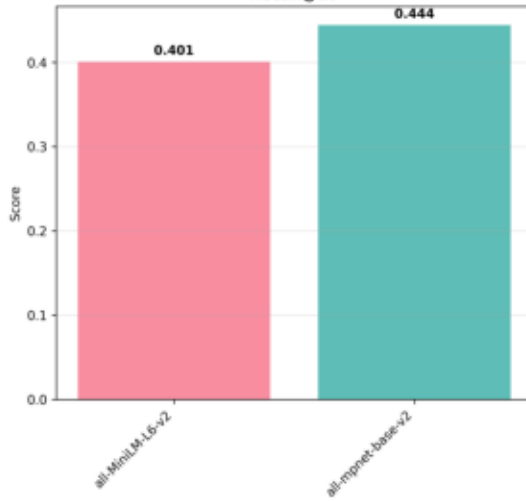Best F1@10: 0.447 (all-mpnet-base-v2)

Best NDCG@10: 1.000
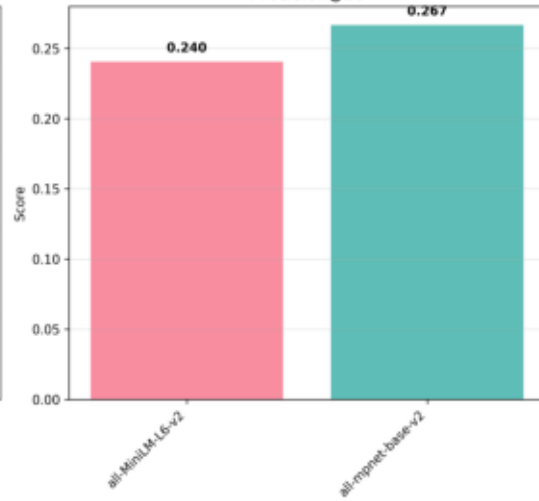
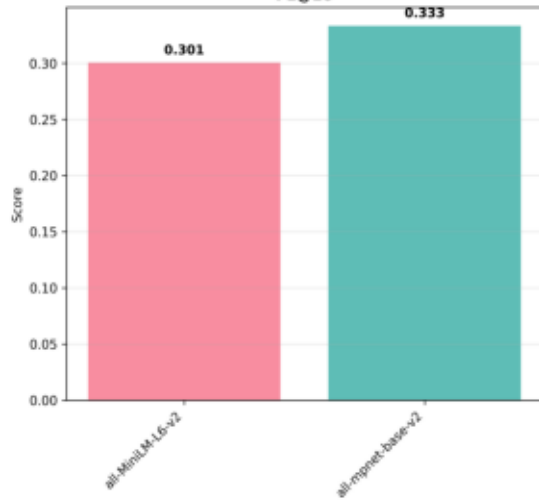Avg Processing Time: 10.0s

Performance Overview

NDCG@10

Model	Score
all-MiniLM-L6-v2	1.000
all-mpnet-base-v2	0.988

Recall@10

Model	Score
all-MiniLM-L6-v2	0.401
all-mpnet-base-v2	0.444

Precision@10

Model	Score
all-MiniLM-L6-v2	0.240
all-mpnet-base-v2	0.267

F1@10

Model	Score
all-MiniLM-L6-v2	0.301
all-mpnet-base-v2	0.333

MAP

Model	Score
all-MiniLM-L6-v2	1.000
all-mpnet-base-v2	0.988

MRR

Model	Score
all-MiniLM-L6-v2	1.000
all-mpnet-base-v2	0.988

Comprehensive Metrics Heatmap

Metrics	all-MiniLM-L6-v2	all-mpnet-base-v2
NDCG@10	1.000	0.988
Recall@10	0.401	0.444
Precision@10	0.240	0.267
F1@10	0.301	0.333
MAP	1.000	0.988
MRR	1.000	0.988
Context_Coherence	0.711	0.711
Answer_Relevance	0.187	0.188

Models
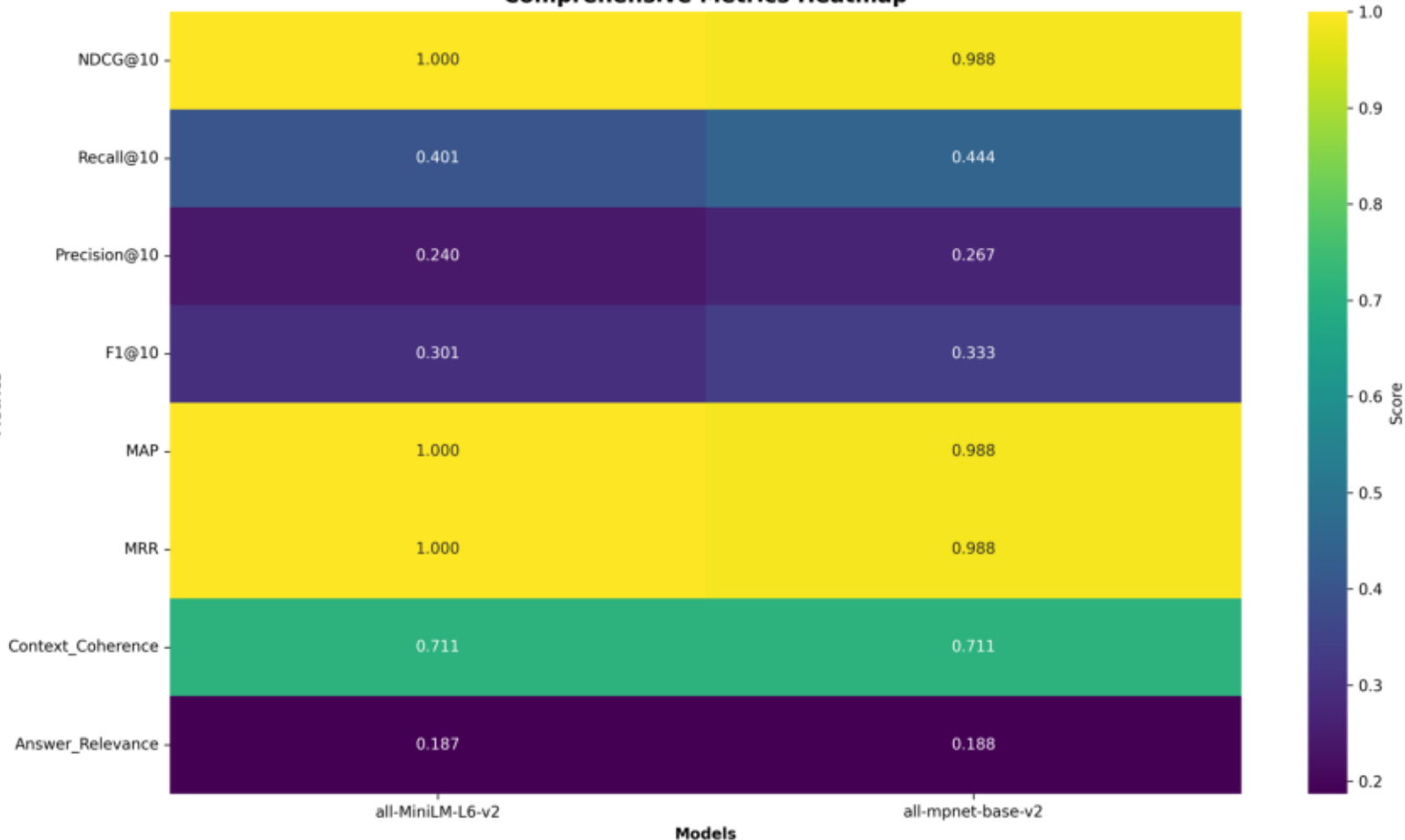
Performance vs Speed Trade-off

NDCG@10 vs Time (seconds)

all-MiniLM...

all-mpnet-...
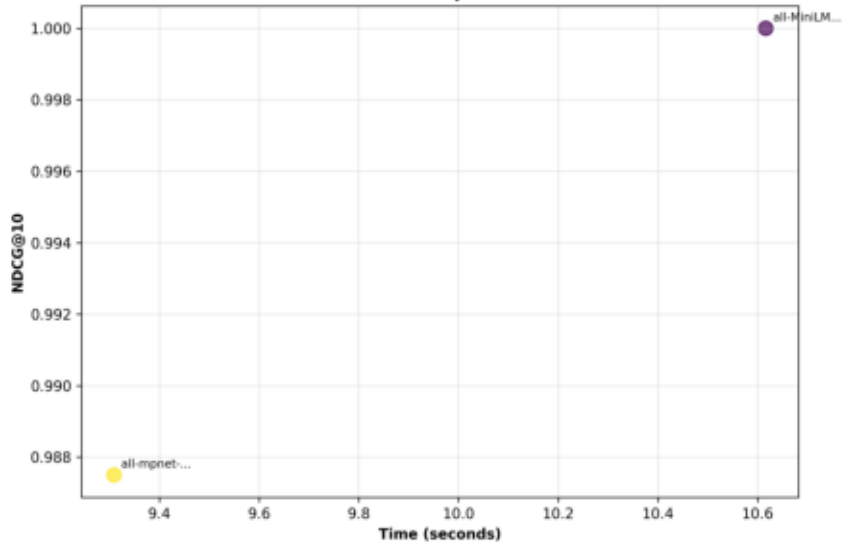
Processing Time Comparison

Time (seconds)

9.3s — all-mpne...

10.6s — all-Mini...
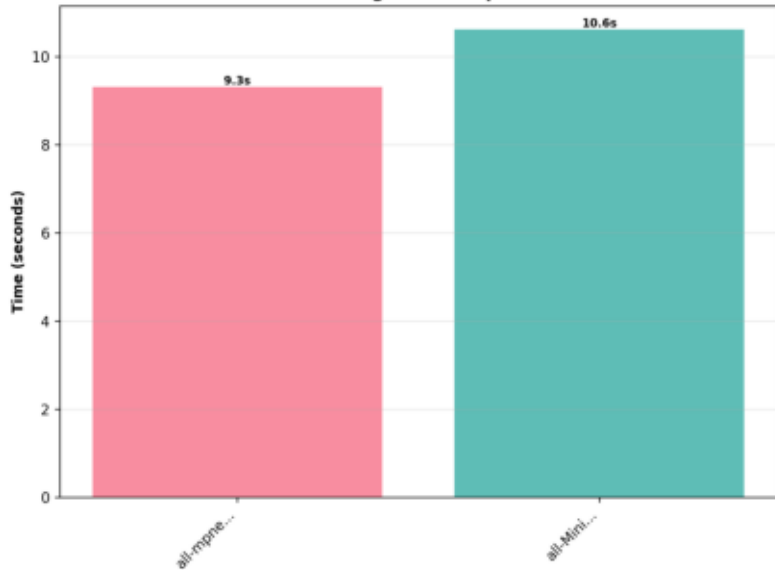
Individual Metric Performance Comparison

NDCG@10 Performance
- all-MiniLM-L6-v2: 1.000
- all-mpnet-base-v2: 0.988

Recall@10 Performance
- all-MiniLM-L6-v2: 0.401
- all-mpnet-base-v2: 0.444
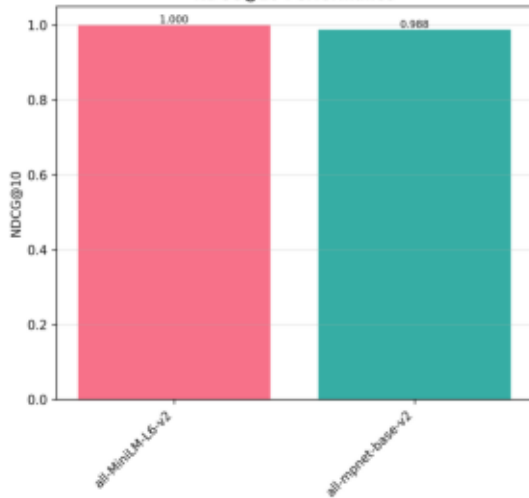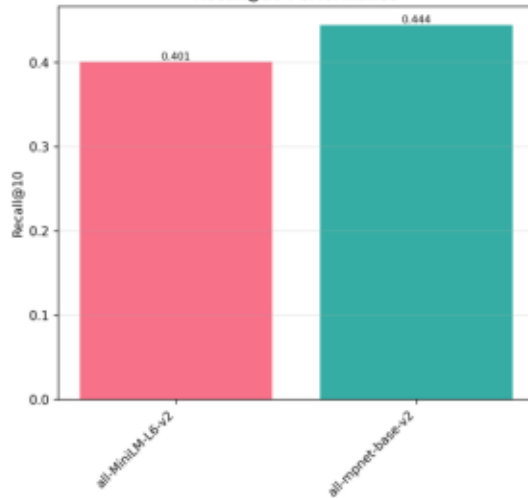
Precision@10 Performance
- all-MiniLM-L6-v2: 0.240
- all-mpnet-base-v2: 0.267

F1@10 Performance
- all-MiniLM-L6-v2: 0.301
- all-mpnet-base-v2: 0.333

MAP Performance
- all-MiniLM-L6-v2: 1.000
- all-mpnet-base-v2: 0.988
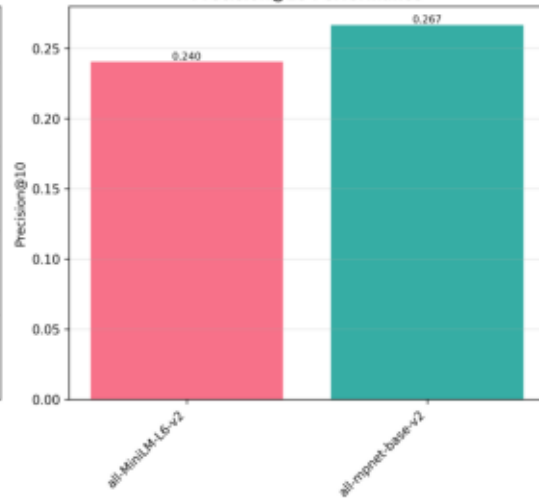
MRR Performance
- all-MiniLM-L6-v2: 1.000
- all-mpnet-base-v2: 0.988
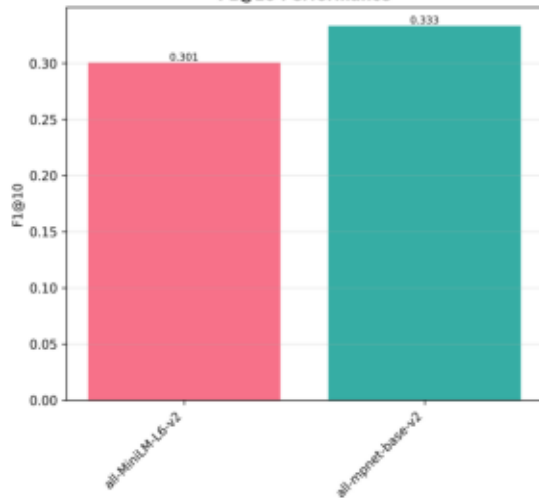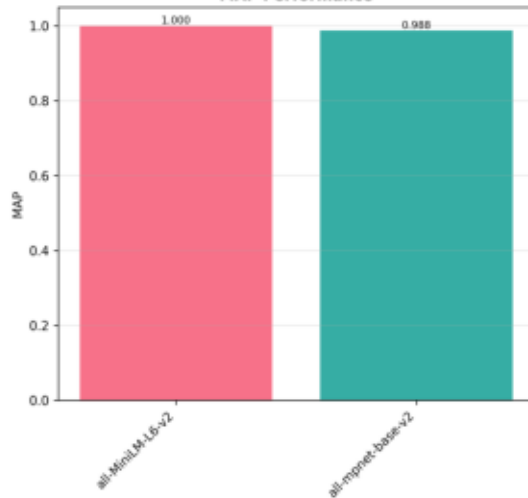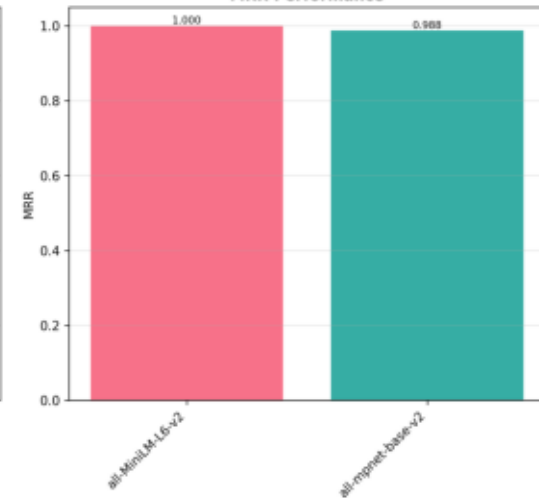
Multi-Metric Performance Comparison (Top 8 Models)

Legend:
- NDCG@10
- Recall@10
- Precision@10
- F1@10
- MAP
- MRR

all-MiniLM-L6-v2:
- NDCG@10: 1.00
- Recall@10: 0.40
- Precision@10: 0.24
- F1@10: 0.30
- MAP: 1.00
- MRR: 1.00

all-mpnet-base-v2:
- NDCG@10: 0.99
- Recall@10: 0.44
- Precision@10: 0.27
- F1@10: 0.33
- MAP: 0.99
- MRR: 0.99

Score (y-axis)
Models (x-axis)
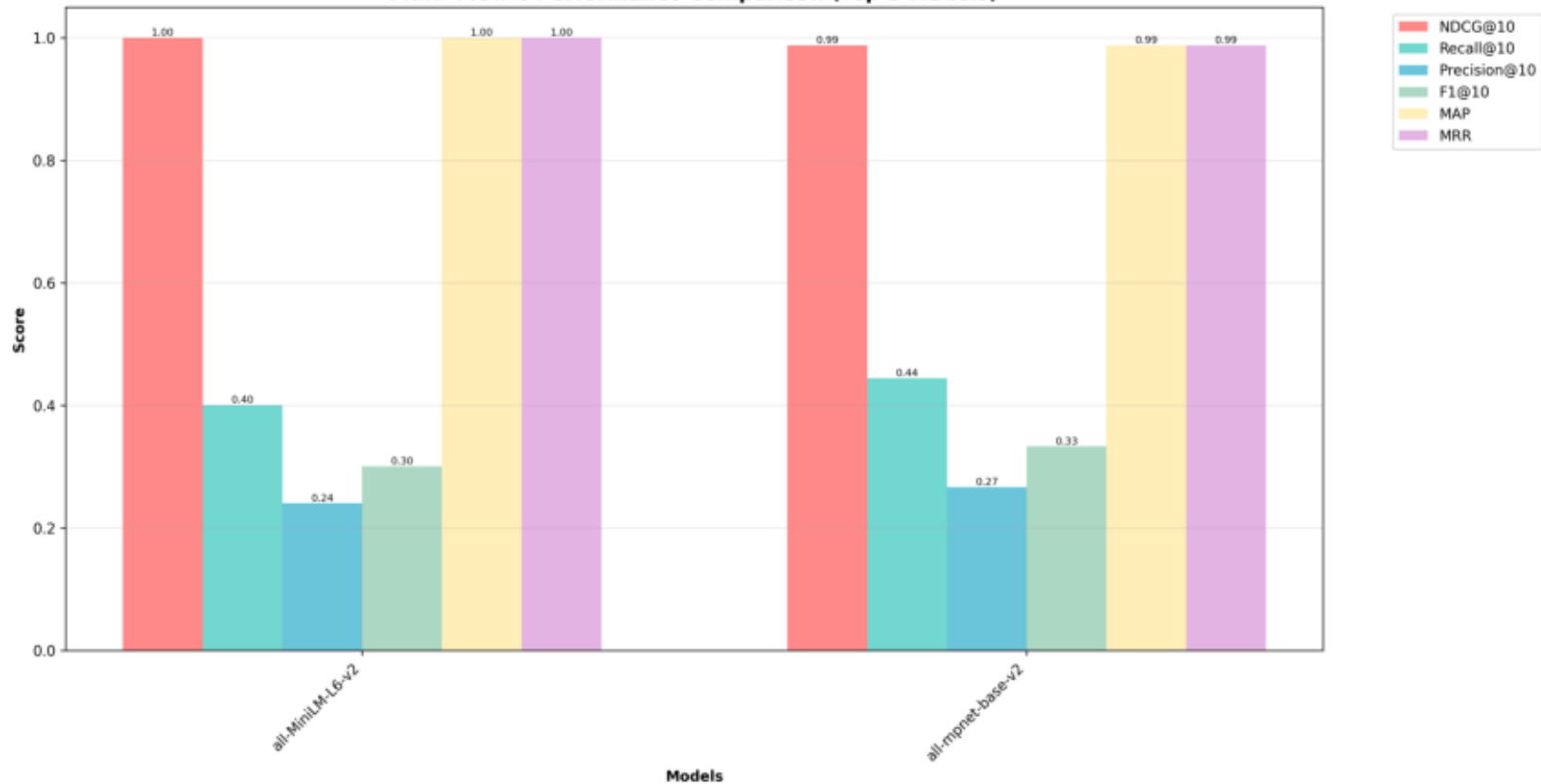
Top Performers by Metric

Bar chart showing Best Score for each metric:
- NDCG@10: 1.000 (all-MiniLM-L6-v...)
- Recall@10: 0.444 (all-mpnet-base-...)
- Precision@10: 0.267 (all-mpnet-base-...)
- F1@10: 0.333 (all-mpnet-base-...)
- MAP: 1.000 (all-MiniLM-L6-v...)
- MRR: 1.000 (all-MiniLM-L6-v...)
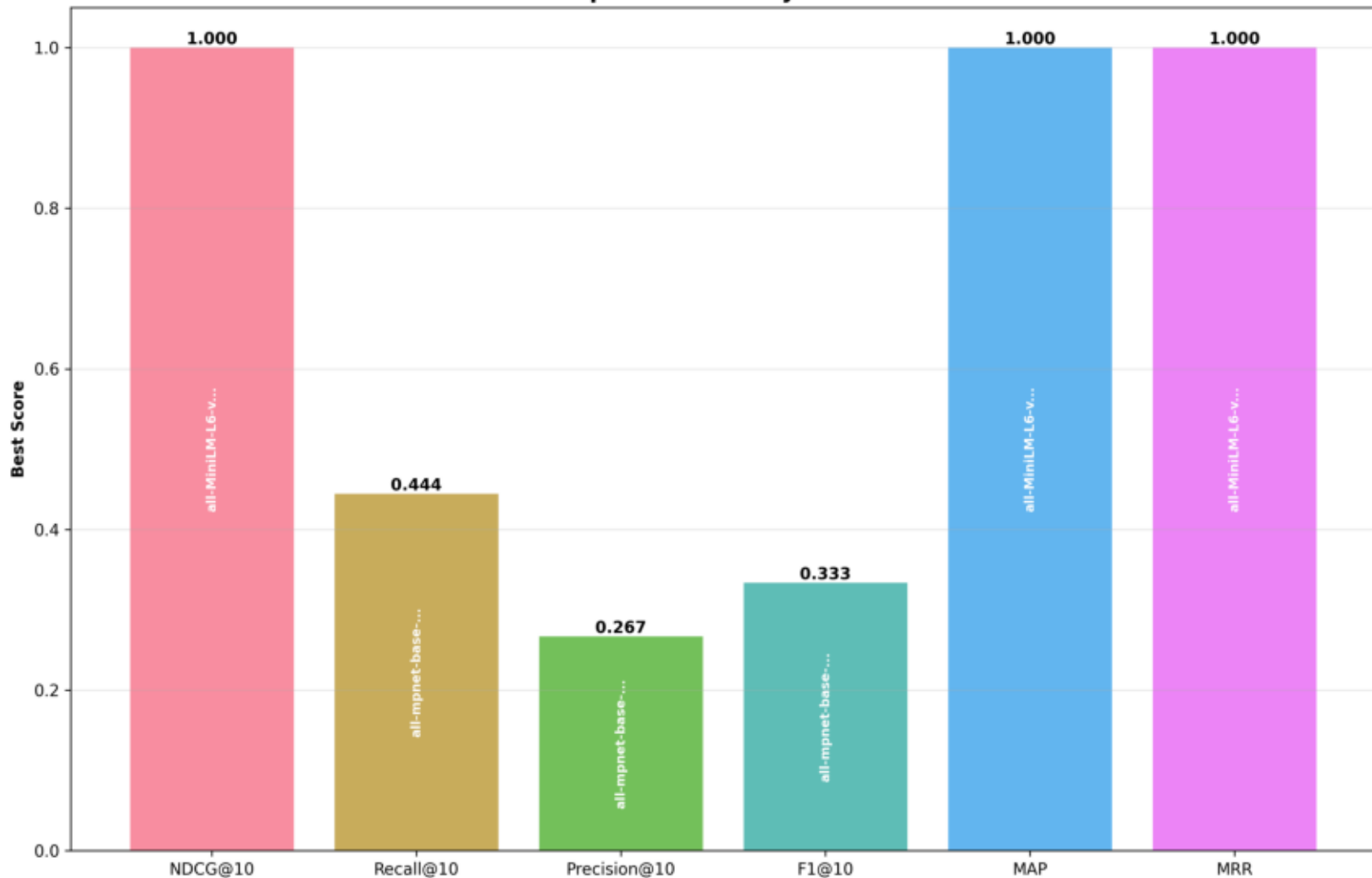
QUAC Dataset - Model Comparison
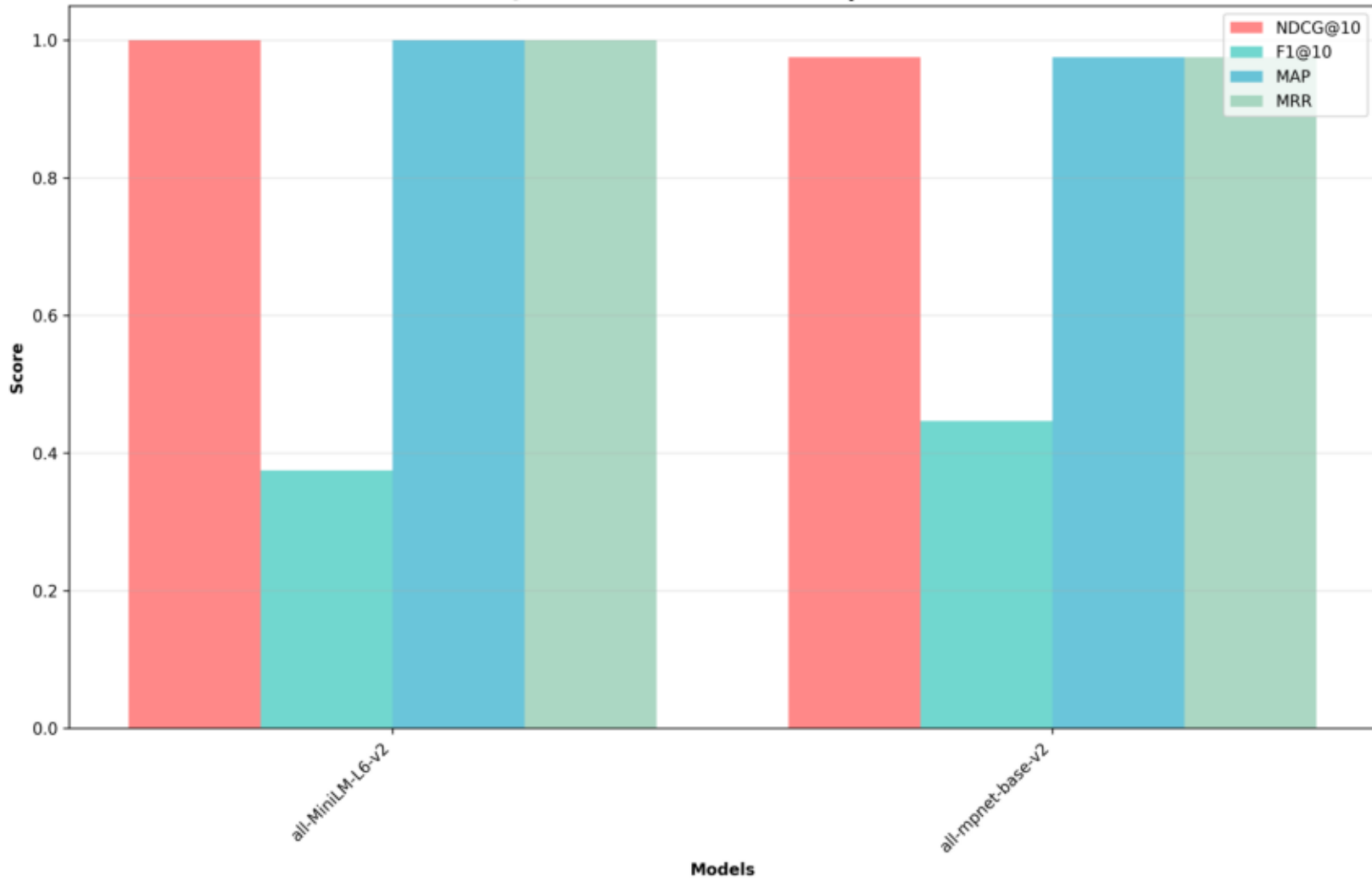
TOPICQA Dataset - Model Comparison