

Evaluation Metrics

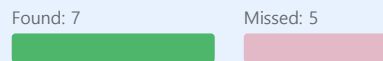
Retrieval Metrics

Precision@10



Formula: $\text{relevant}/K = 7/10 = 70\%$

Recall@10



Formula: $\text{found}/\text{total} = 7/12 = 58\%$

NDCG Score



Score: 0.85 (considers rank order)

Generation Metrics

ROUGE-L Score

Generated:
Diabetes requires insulin therapy
Reference:
Diabetes needs insulin treatment

LCS match: 75%

BERTScore



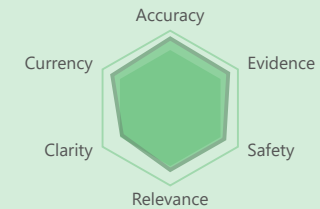
Semantic similarity: 92%

Hallucination Rate

2% (Target: <5%)

Medical Metrics

Clinical Relevance



Evidence Quality

RCT 40% Cohort 30% Case 20%

Weighted score: 8.2/10

Citation Accuracy

98.5%

Target Thresholds for Production

Precision@10: >90%

NDCG: >0.8

ROUGE-L: >0.7

Hallucination: <5%