







Evaluation Metrics

 **Retrieval Metrics**


Precision@10

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


Recall@10
Found: 7  Missed: 5 
Formula: $\text{found}/\text{total} = 7/12 = 58\%$


NDCG Score
 Position weight decay
Score: 0.85 (considers rank order)

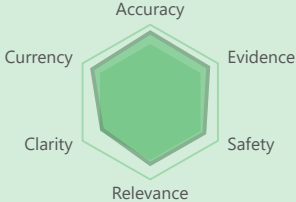
 **Generation Metrics**


ROUGE-L Score
Generated: Diabetes requires insulin therapy
Reference: Diabetes needs insulin treatment
LCS match: 75%

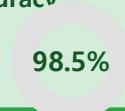
BERTScore

Semantic similarity: 92%


Hallucination Rate


 **Medical Metrics**

Clinical Relevance


Evidence Quality

Weighted score: 8.2/10

Citation Accuracy


 **Target Thresholds for Production**

Precision@10: >90%

NDCG: >0.8

ROUGE-L: >0.7

Hallucination: <5%