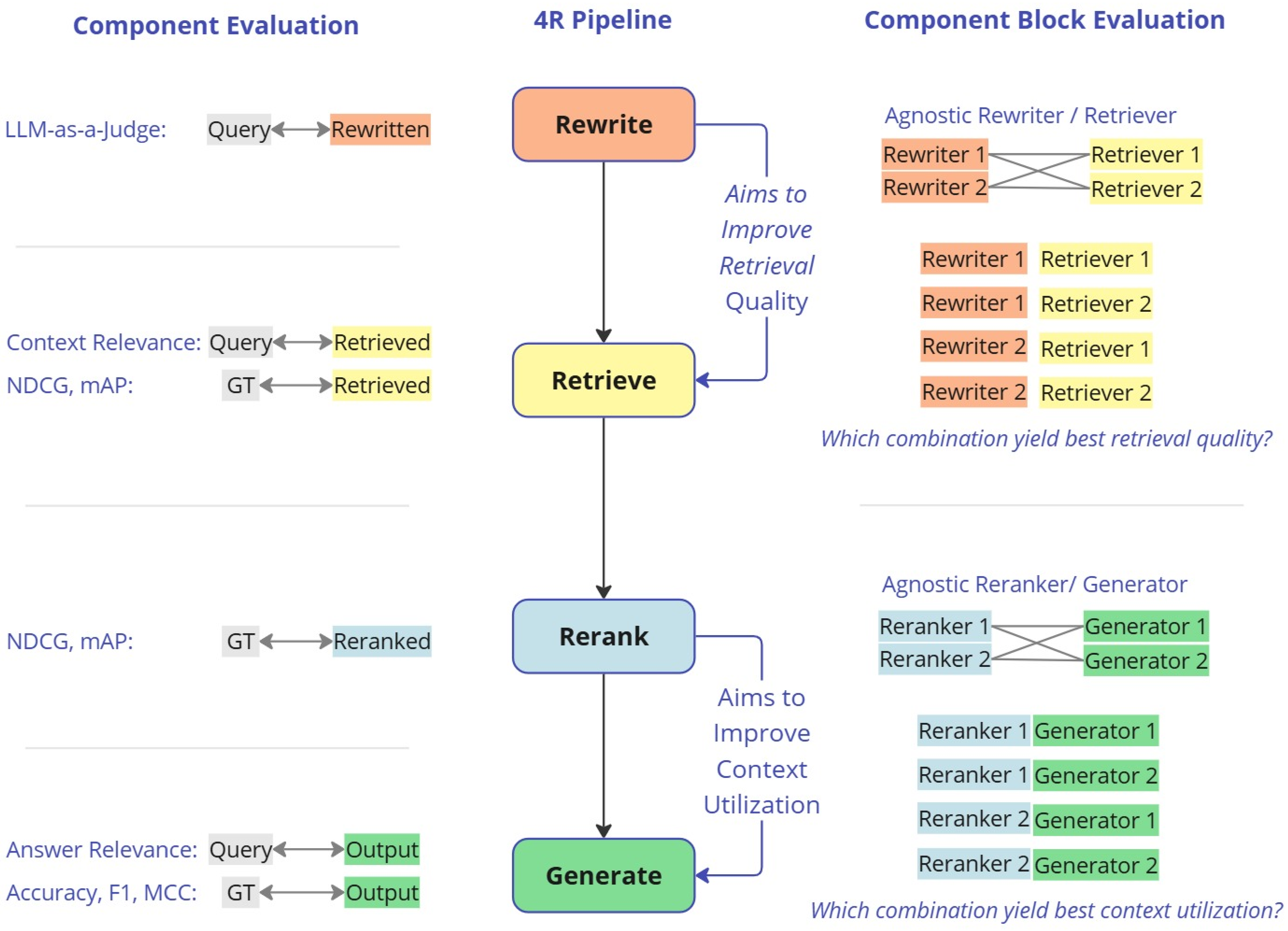
Component Evaluation
4R Pipeline
Component Block Evaluation
LLM-as-a-Judge:
Query
Rewritten
Context Relevance:
Query
Retrieved
NDCG, mAP:
GT
Retrieved
NDCG, mAP:
GT
Reranked
Answer Relevance:
Query
Output
Accuracy, F1, MCC:
GT
Output
Rewrite
Retrieve
Rerank
Generate
Aims to Improve Retrieval Quality
Aims to Improve Context Utilization
Agnostic Rewriter / Retriever
Rewriter 1
Rewriter 2
Retriever 1
Retriever 2
Rewriter 1 Retriever 1
Rewriter 1 Retriever 2
Rewriter 2 Retriever 1
Rewriter 2 Retriever 2
Which combination yield best retrieval quality?
Agnostic Reranker/ Generator
Reranker 1
Reranker 2
Generator 1
Generator 2
Reranker 1 Generator 1
Reranker 1 Generator 2
Reranker 2 Generator 1
Reranker 2 Generator 2
Which combination yield best context utilization?