

# RAG Evaluation



[gollnickdata.de](http://gollnickdata.de)

# RAG Evaluation: Possible Issues

# RAG Evaluation

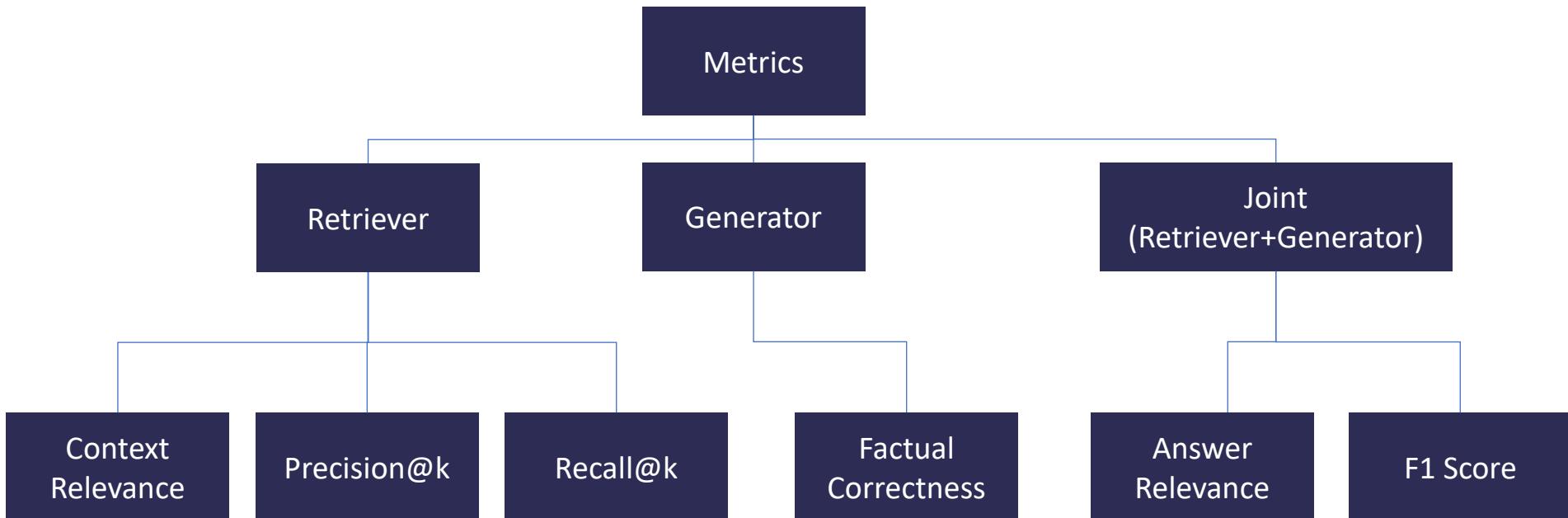
## Issues

- hallucinations due to
  - Retriever...fails to retrieve relevant context
  - Generation...LLM does not consider context
  - Generation...LLM picks irrelevant information from context



# RAG Evaluation

## Metrics



# RAG Evaluation

## Required Abilities

### Noise Robustness

- possible that document is related to user query, but does not have relevant info

### Counter Fact Robustness

- info might be wrong
- RAG should reject known inaccuracies

### Negative Rejection

- LLMs always generate text
- ability to NOT give answer when no info is available

### Latency

### Information Integration

- ability to assimilate information from multiple docs

### Bias

### Toxicity



# RAG Evaluation

## Evaluation Metrics

Accuracy

$$= \frac{\text{count relevant} + \text{irrelevant documents retrieved}}{\text{Total number docs}}$$

Precision

$$= \frac{\text{count relevant documents retrieved}}{\text{Total count docs}}$$

Precision@k

$$= \frac{\text{count relevant documents retrieved in Top} - k \text{ docs}}{\text{Top} - k \text{ documents count}}$$

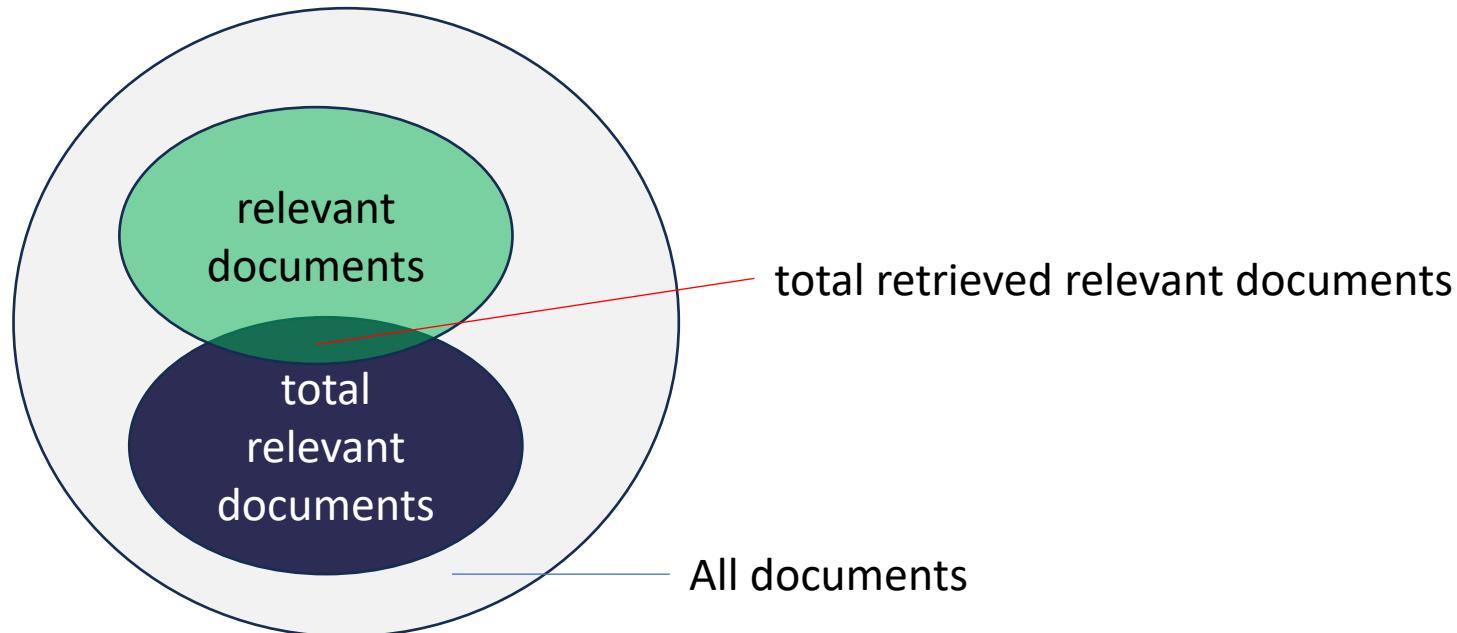
Recall

$$= \frac{\text{count relevant documents retrieved}}{\text{Count relevant documents count}}$$



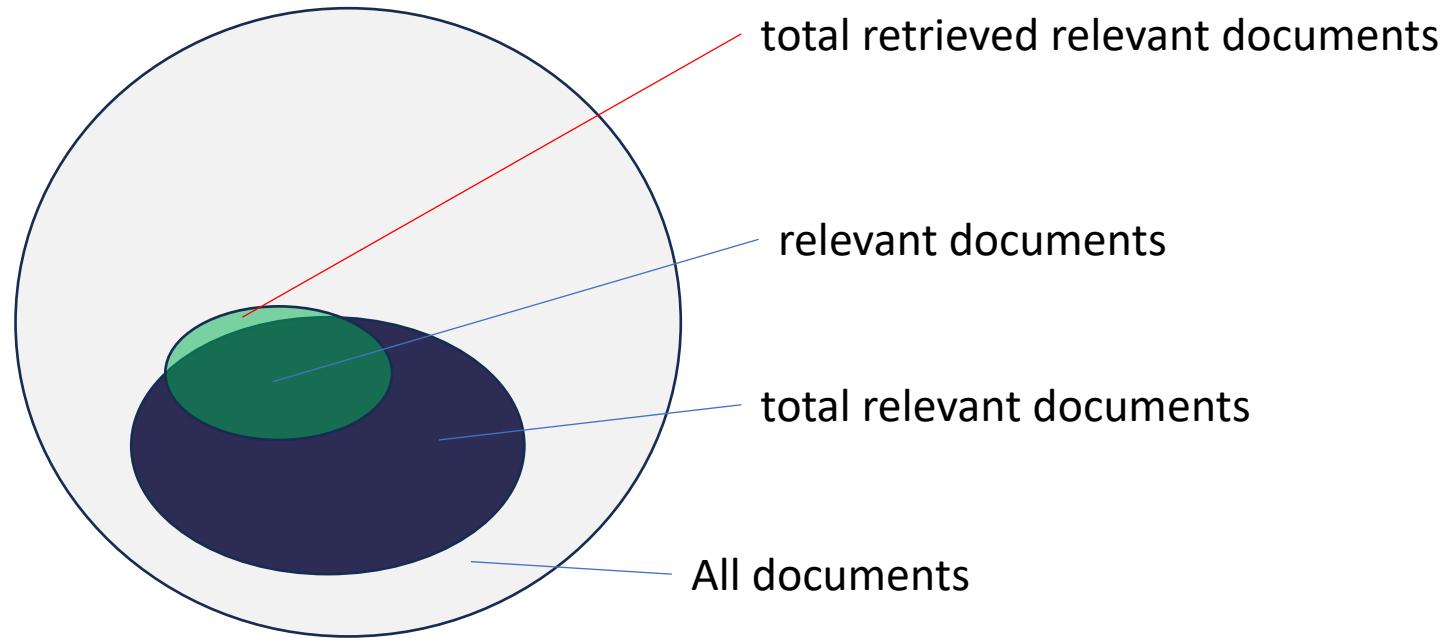
# RAG Evaluation

## Evaluation Metrics



# RAG Evaluation

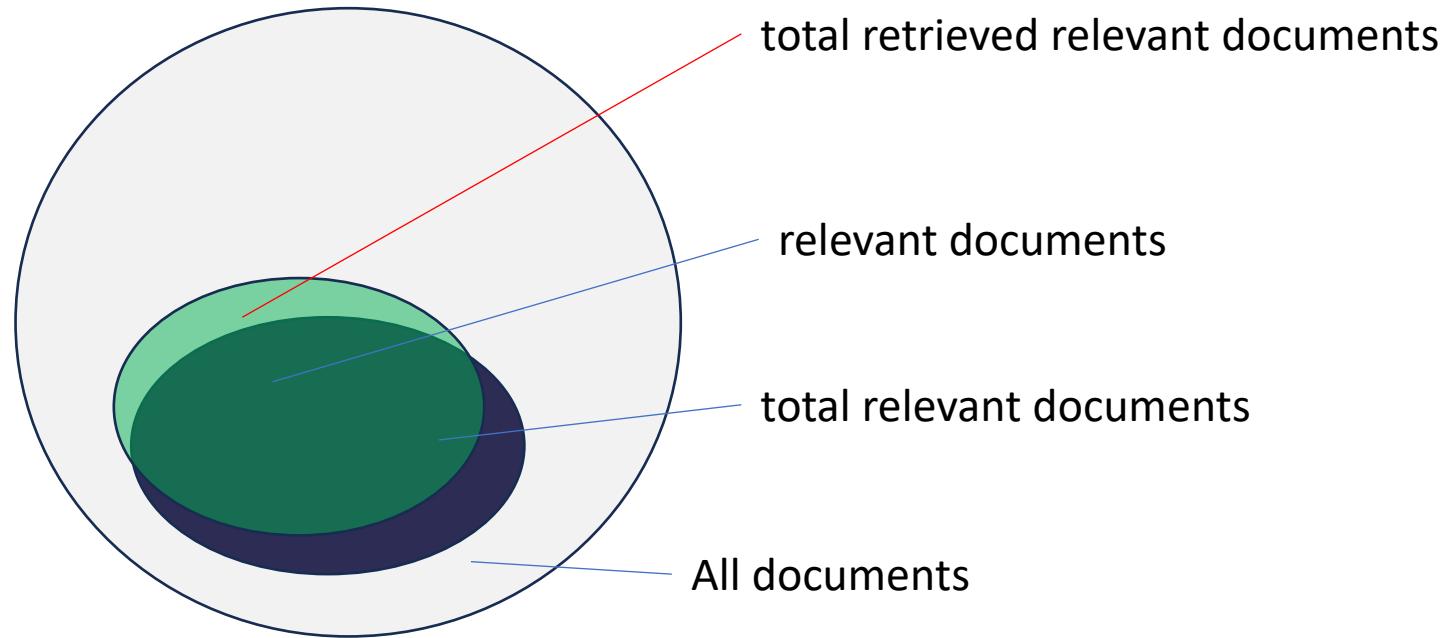
High Precision, Low Recall



gollnickdata.de

# RAG Evaluation

High Precision, High Recall



# RAG Evaluation

RAG Questions and KPIs: Context Relevance

Questions

KPIs

Definitions

## Retrieval

- Is retrieval needed to answer user query?

Context  
Relevance

$$= \frac{\text{Count relevant sentences in retrieval}}{\text{Total count sentences in retrieval}}$$



gollnickdata.de

# RAG Evaluation

RAG Questions and KPIs: Context Relevance

## Context

Apple Inc. is a multinational technology company headquartered in Cupertino, California.

The company designs, manufactures, and sells consumer electronics, software, and online services.

Its flagship products include the iPhone, iPad, and MacBook.

Recently, Apple announced the launch of its latest AI-powered wearable device.

In unrelated news, a global survey on renewable energy adoption showed significant growth in solar energy usage in Europe.

Total count sentences = 5

Relevant sentences = 4

context relevance = 80%

## User Query

What are the latest AI-powered products launched by Apple?



gollnickdata.de

# RAG Evaluation

RAG Questions and KPIs: Answer Faithfulness

Questions

KPIs

Definitions

## Generation

- Is answer based on retrieval?

Answer  
Faithfulness

$$= \frac{\text{Count generated claims in context}}{\text{Total count claims in response}}$$



gollnickdata.de

# RAG Evaluation

RAG Questions and KPIs: Answer Faithfulness

Context

The 2020 United States presidential election was held on November 3, 2020. Joe Biden, the Democratic candidate, defeated the incumbent president, Donald Trump, to become the 46th president of the United States. Biden's victory was confirmed by the Electoral College on December 14, 2020.

User Query

Who won the last U.S. presidential election in 2020, and when exactly was it held?

**Joe Biden won the U.S. presidential election** which was held on **November 8, 2020**.

partially faithful answer



gollnickdata.de

# RAG Evaluation

RAG Questions and KPIs: Answer Relevance

Questions

KPIs

Definitions

## Generation

- Is answer relevant to the user query?

Answer  
Relevance

$$= \sum_{i=0}^N \text{Similarity}(\text{user query}, \text{LLM generated synthetic query}[i])$$



gollnickdata.de

# RAG Evaluation

RAG Questions and KPIs: Answer Relevance

