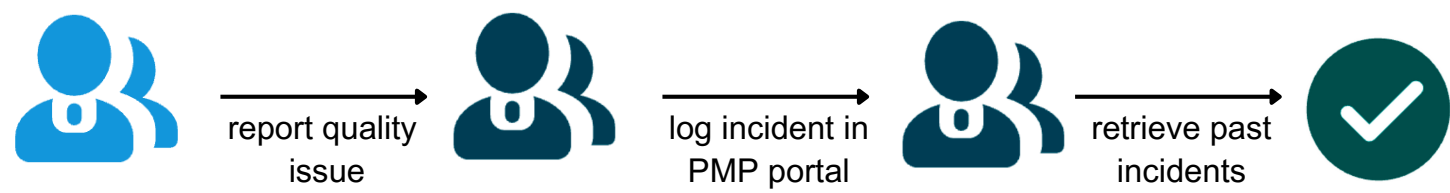




Motivation

Traditional workflow of Reporting Line Alignment (RLA) for Problem Management Process (PMP) is based on manual filters and column name searching. It's time consuming and error-prone.



Scope

Task 1

Enhanced Incident Retrieval

Existing RLA process

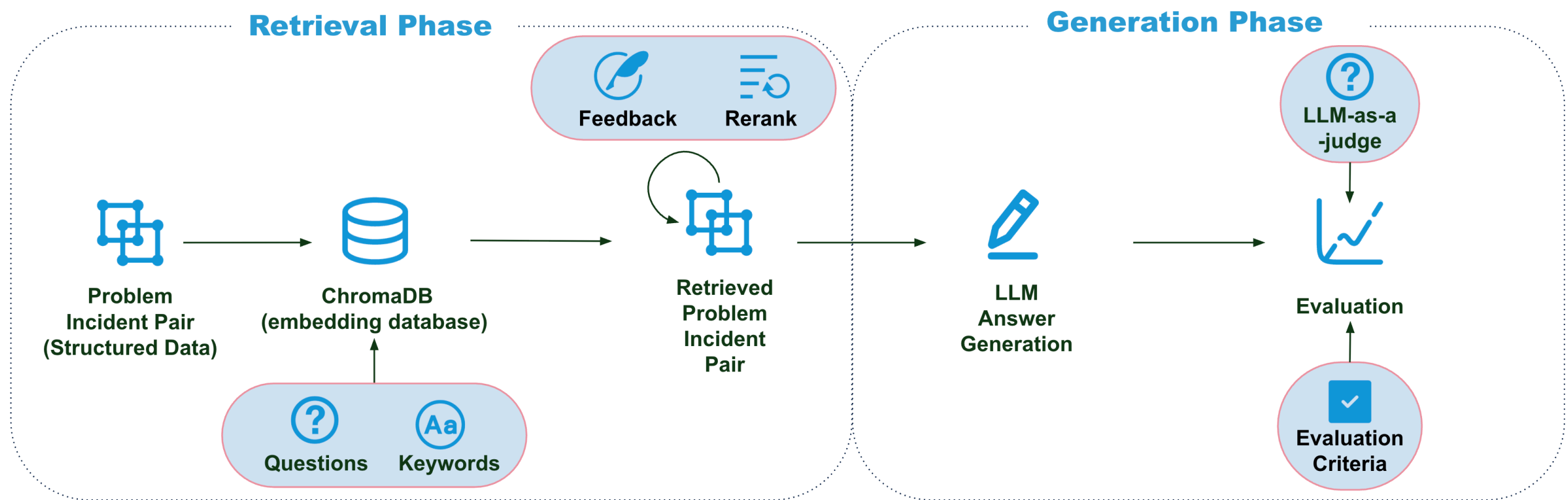
Task 2

Q&A based on Incident-Problem Pairs

New feature

Project Overview

- 1 Intelligent retrieval of past incidents, iterated with **human feedback**
- 2 Semantic **reranking** and keyword-guided generation
- 3 Human-in-the-Loop & LLM-based **evaluation**
- 4 Seamless **UI** integration with existing search tool



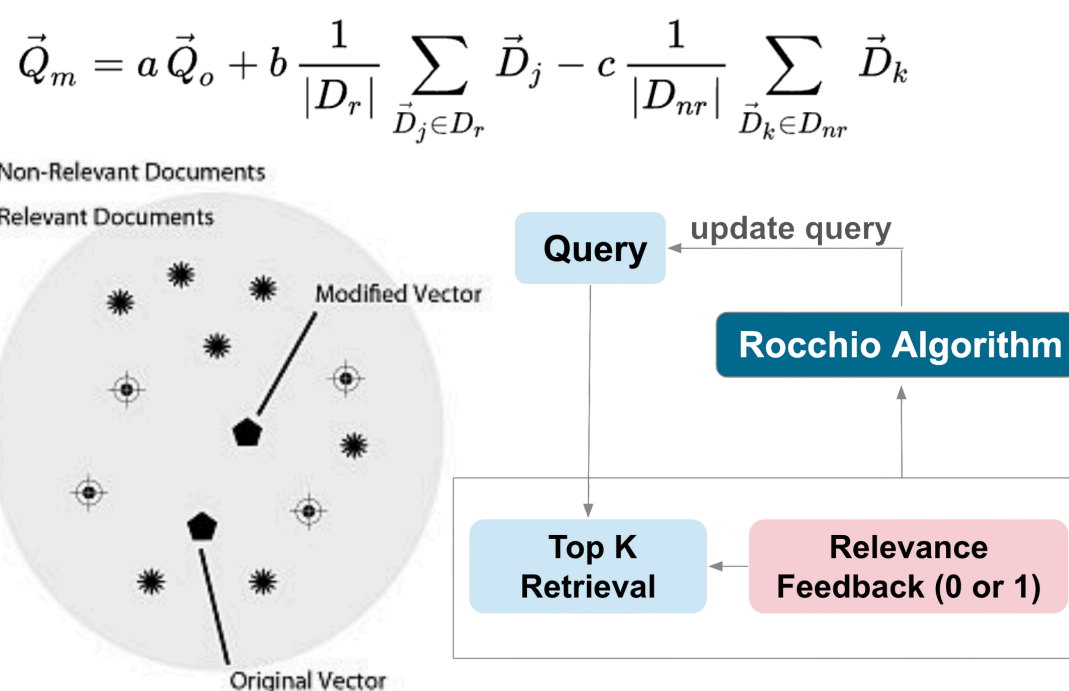
Methodology

Expert Keyword Extraction

1. **Preprocess** historical expert query → look for **common search pattern**
2. Build **keyword catalogue**
3. **Cross-reference** with incident-problem text
4. **N-gram** keyword extraction

The majority of incident-problem pairs have **5–7 matched expert keywords**

Retrieval based on Relevance Feedback: Rocchio Algorithm

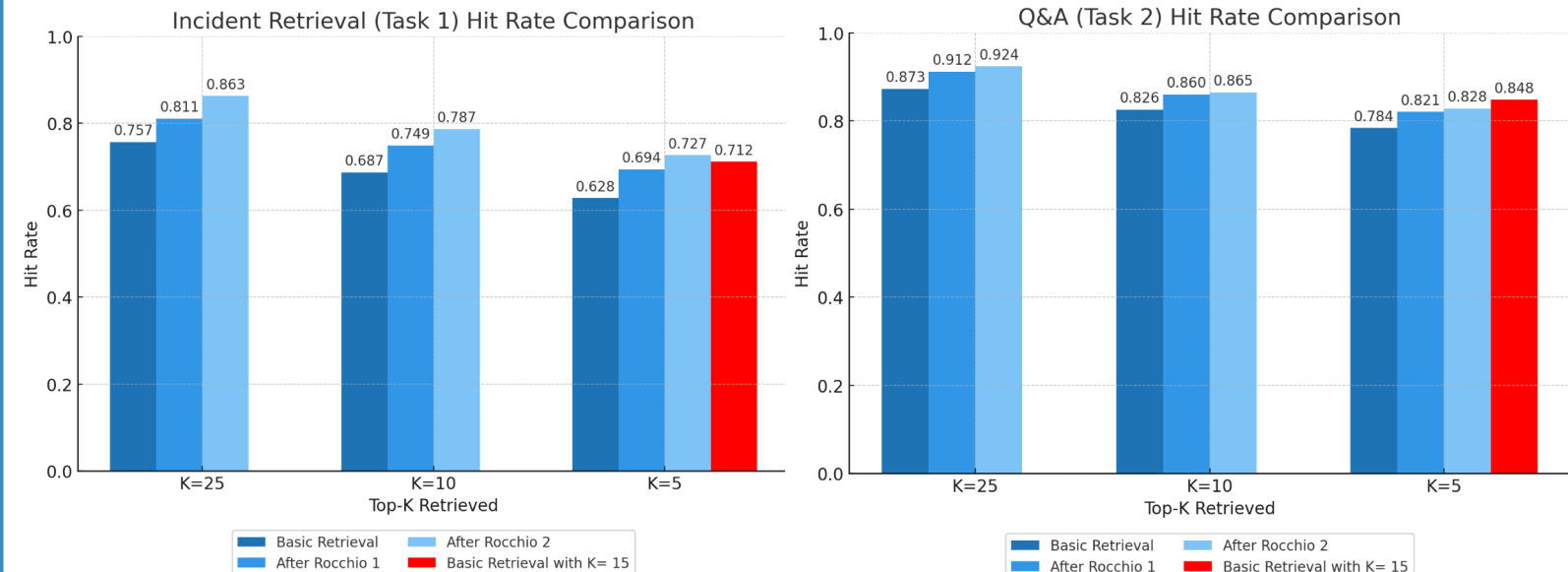


Retrieval Reranking

- 1 **Keyword Re-ranking**
 - Uses expert query keywords extracted in Methodology 1
 - Reranks based on the count of keyword in retrieved documents
- 2 **Query-aware Re-ranking (Langchain)**
 - Compresses and re-orders the candidate results by reconsidering the query's intent more deeply – is the retrieved item really relevant to the question asked by user?

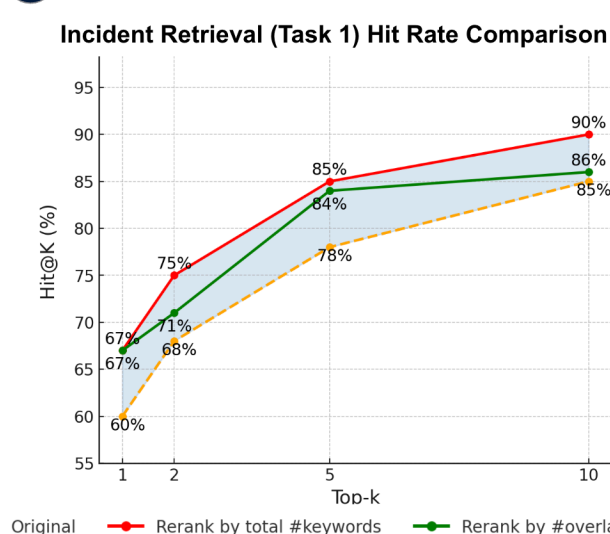
Result

Conclusion 1: Rocchio consistently improves the HitRate across K, but the significance is scenario-dependent



Conclusion 2: Both reranking methods are effective; Chainfilter balances the time and performance the best

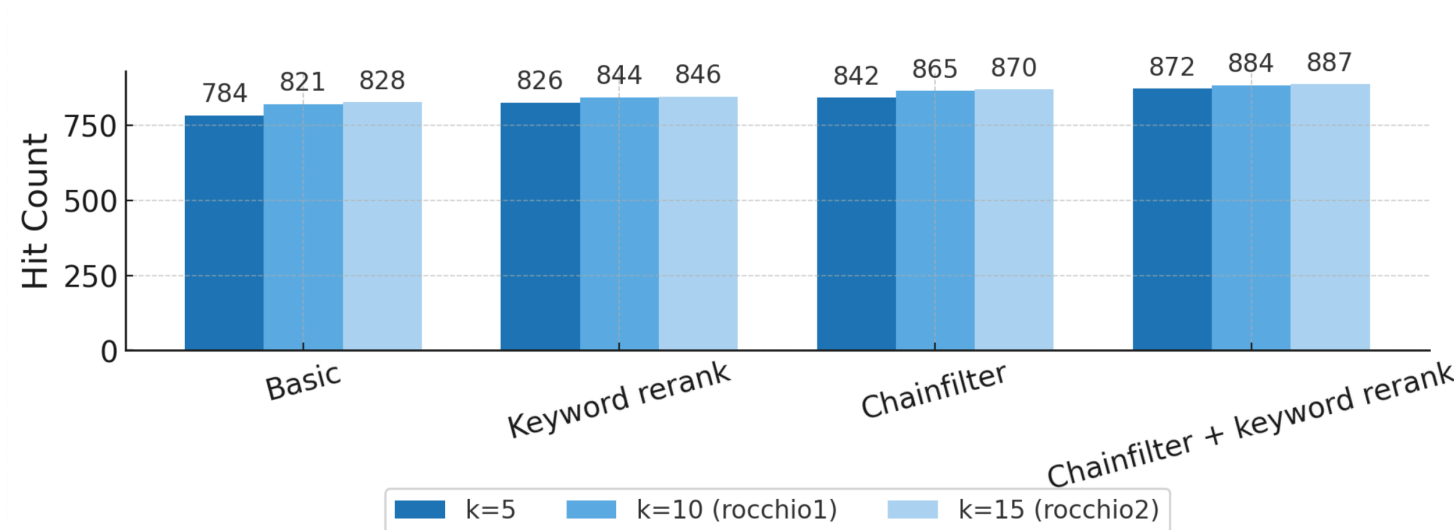
1 Keyword Re-ranking



2 Query-aware Re-ranking (Langchain)

Method	Time (sec/query)	Hit Rate
Basic Retrieval	0.003	0.784
Cross encoder	3.206	0.819
Embedding filter	0.325	0.826
Chain filter	1.422	0.842

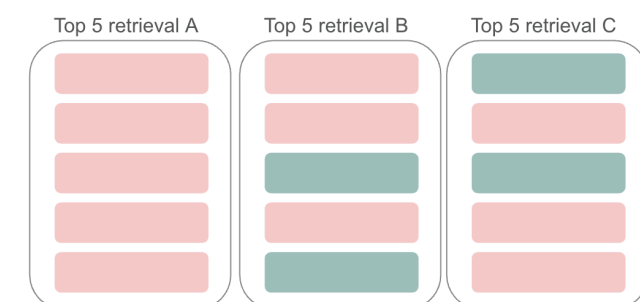
Conclusion 3: Collectively, Rocchio + keyword + Chainfilter reranking boosted the HitRate@5 by 11.2%



Evaluation

Metrics:

- **HitRate @K**: % of queries retrieving at least one relevant result within K
- **Mean Reciprocal Rank @K**: average reciprocals of ranks of first relevant result within K → both higher the better



In this example of Top3 retrieval:

- HitRate@3 is 2/3 = 66.7%
- MRR is (0 + 0.333 + 1.0) / 3 = 44.4%

LLM-as-judge:

- designed a 3-point scale (1 = poor, 2 = partial, 3 = correct) criteria
- human-annotated 100 Q&A pairs for comparison with LLM evaluation

Impact

Based on preliminary expert feedback:

Watch our demo!



Time Spent

Manual:

~30 minutes per incident

Spent on searching, categorizing, prioritizing, routing

Our tool:

Instantaneous retrieval

2-3s generation

Total 1-2 mins

⚡ ~93% faster per case

Scalability

Manual:

25 core employees

40-50 additional users

Our tool:

Handled by AI

Depends on compute resource

⚡ Growing incident process volume