

# RAG Pipeline in Non-Life Insurance Domain



Intelligent Systems - COS30018

Option C - Mitigating AI Hallucination

# Members

**Vu Minh Dung**

Data Engineer,  
Leader

**Le Hoang Long**

Data Engineer

**Vu The Thanh  
Viet**

Document  
Writer

**Nguyen Ngoc Anh**

AI Engineer

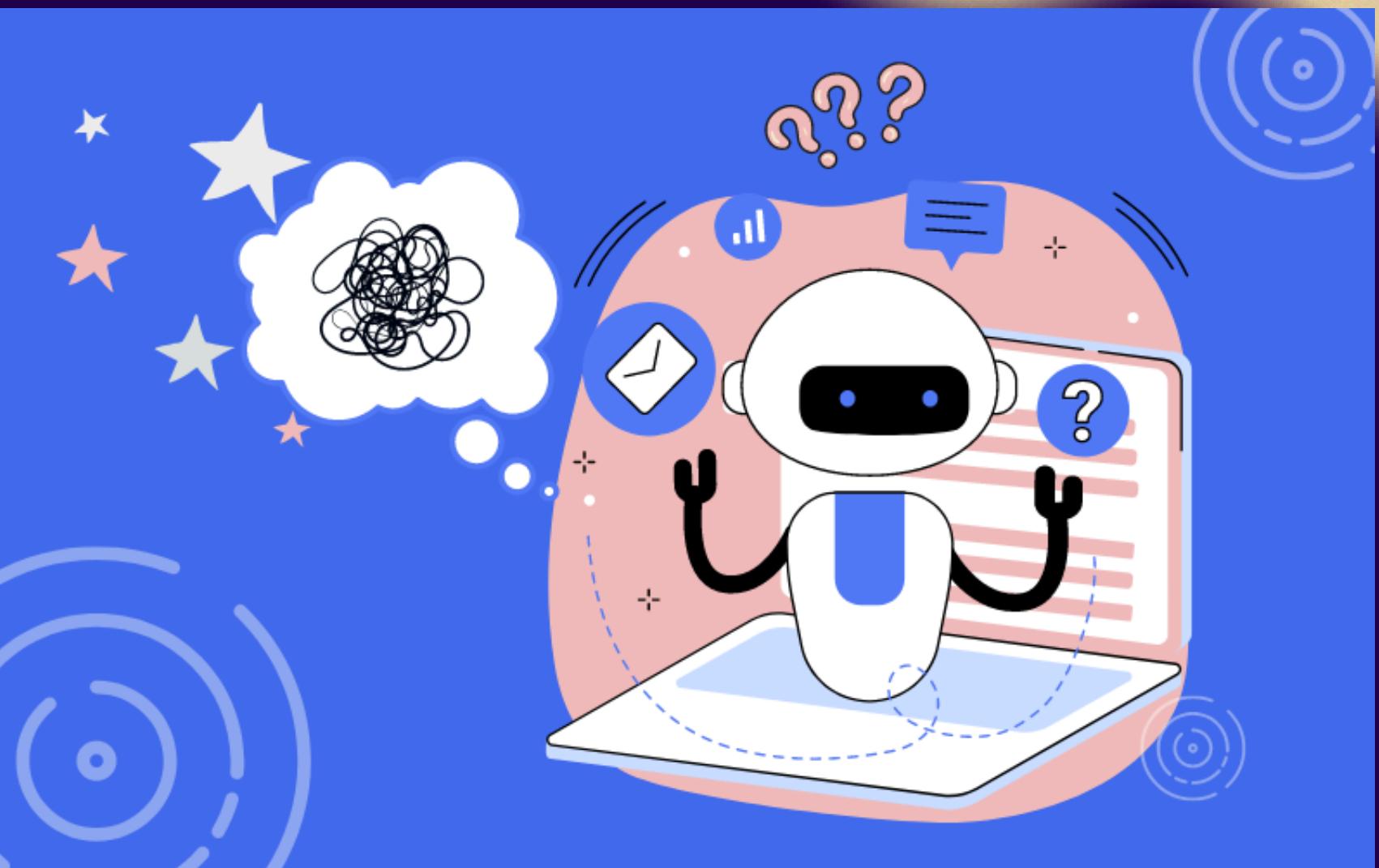
**Arlene Phuong  
Brown**

AI Engineer

# Problem Statement

Despite advancements, hallucinations undermine AI reliability. Existing solutions (RAG, fine-tuning, self-refinement) need better optimization for domain-specific use. Risks include:

- Misinformation in decision-making.
- Loss of trust in AI-generated outputs.
- Legal and financial risks from incorrect information.



# Objective

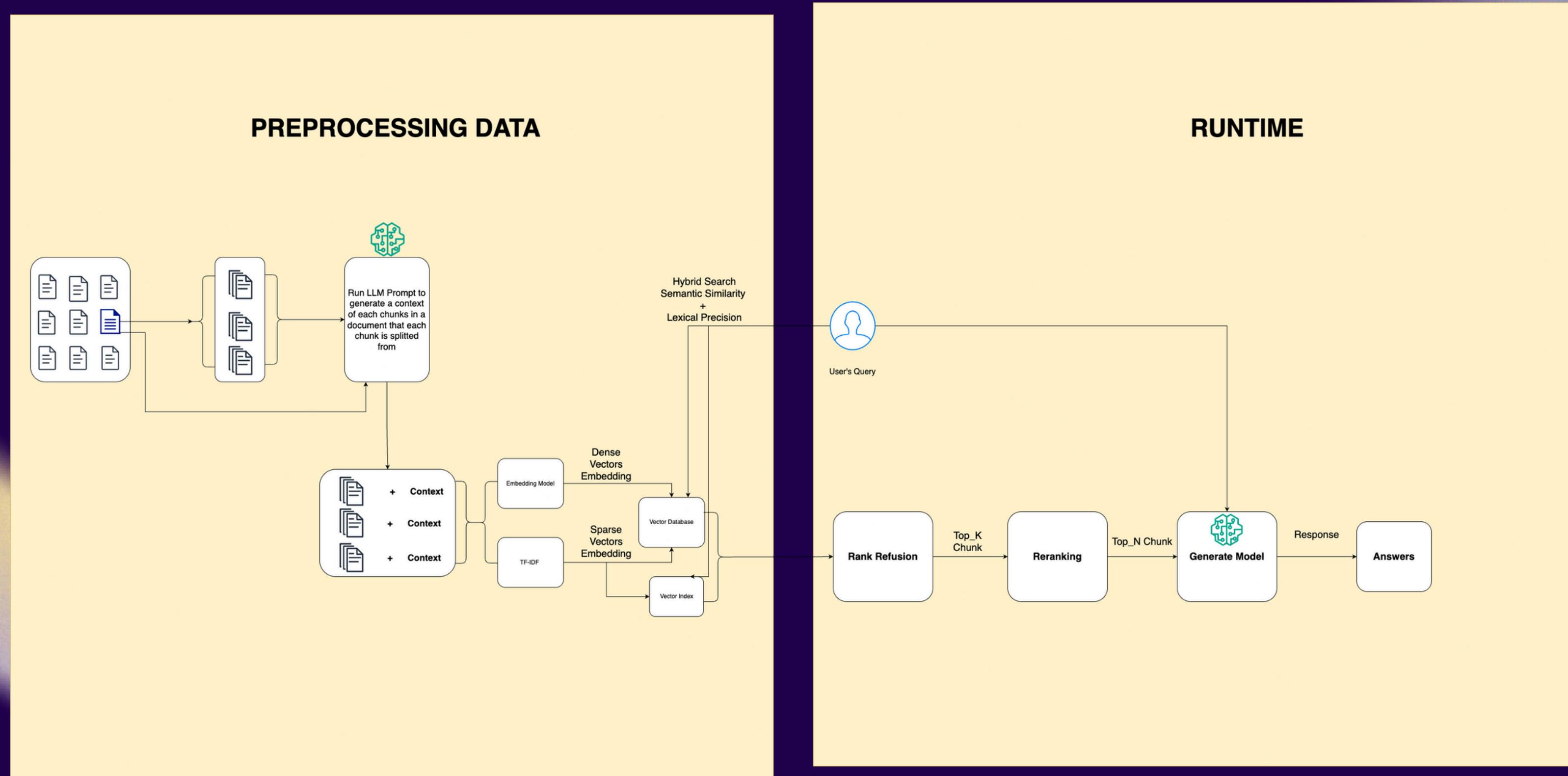
This project aims to implement and evaluate RAG to reduce hallucinations in the non-life insurance domain, ensuring accurate and context-aware responses.

Key challenges include:

- Adapting RAG for legal and regulatory documents.
- Balancing accuracy with efficiency.
- Aligning retrieved context with user queries to prevent fabrications.



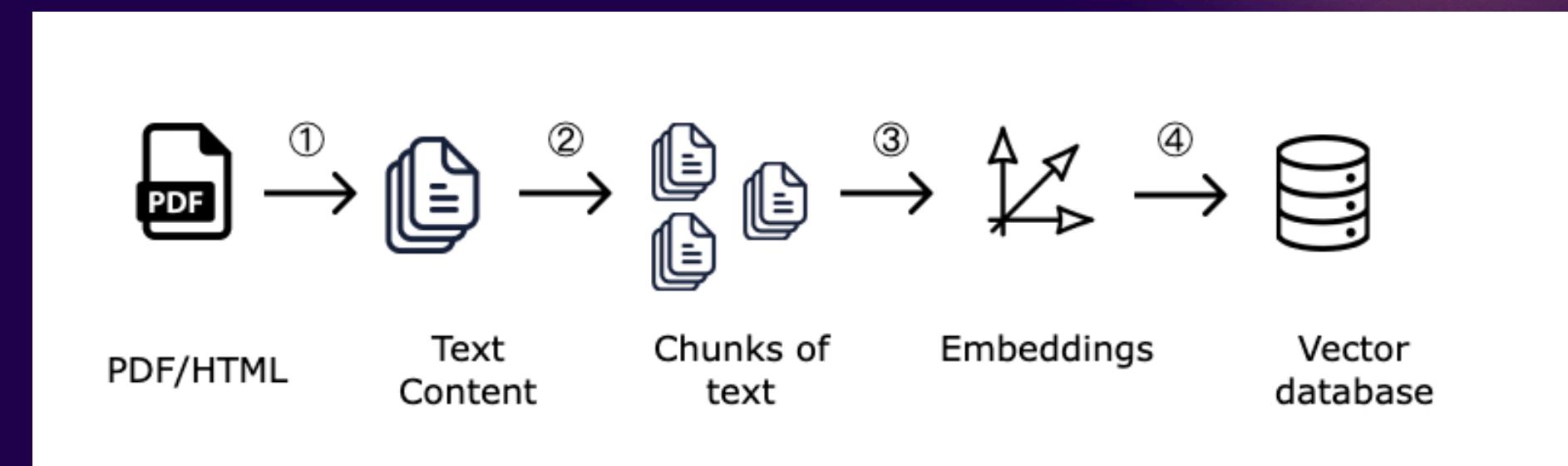
# System Architecture



# Implementation

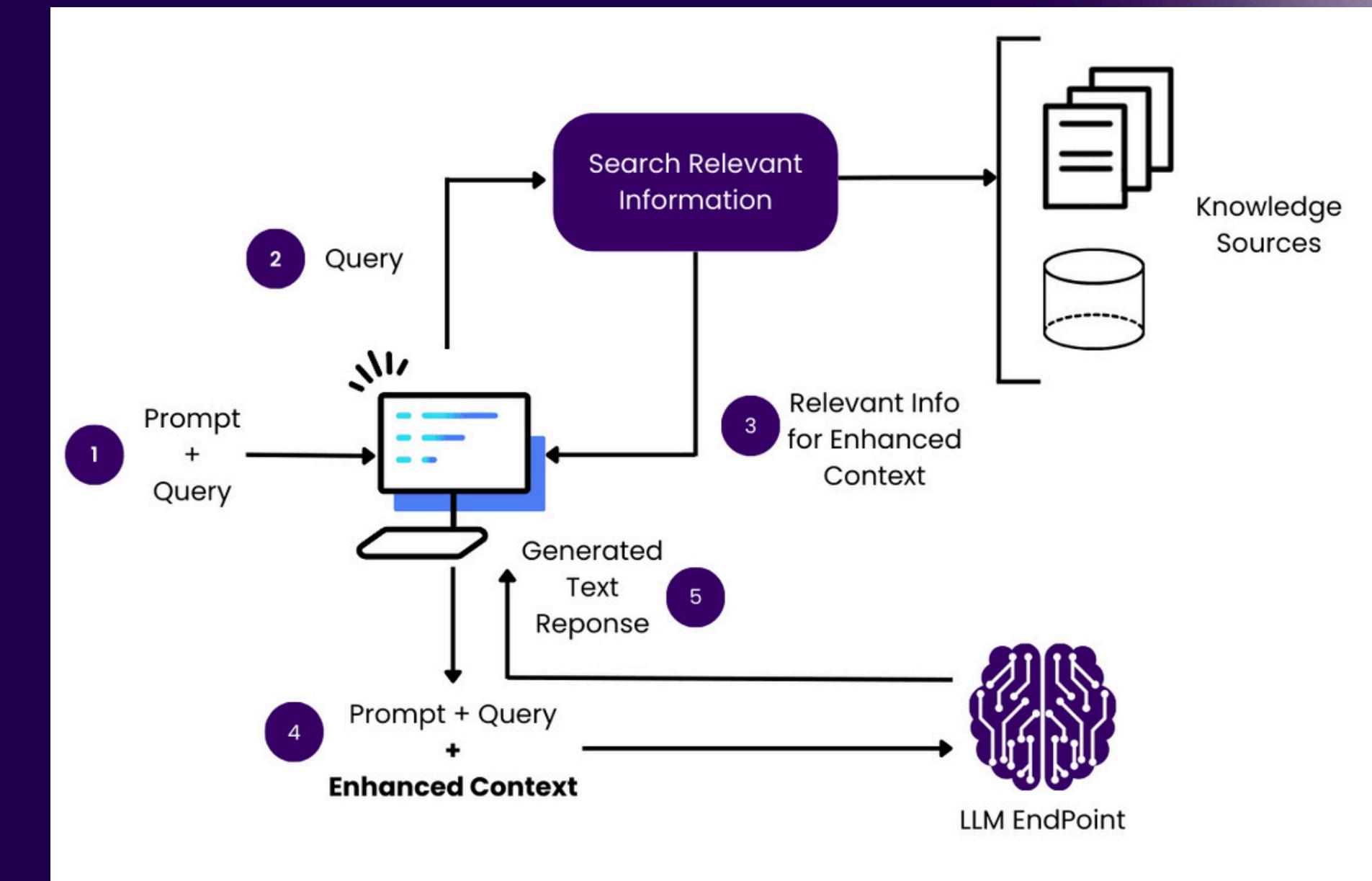
## 1. Data chunking & ingestion

1. Load PDF Document
2. Split Document into smaller Chunks
3. Enhance Chunks with Generated Context from LLM
4. Embedding each chunks with Dense Vector and Sparse Vector Embedding
5. Store Vector and metadata of each chunks into a Vector Store Database



# Implementation

## 2. Retrieval & Inference



# Evaluation

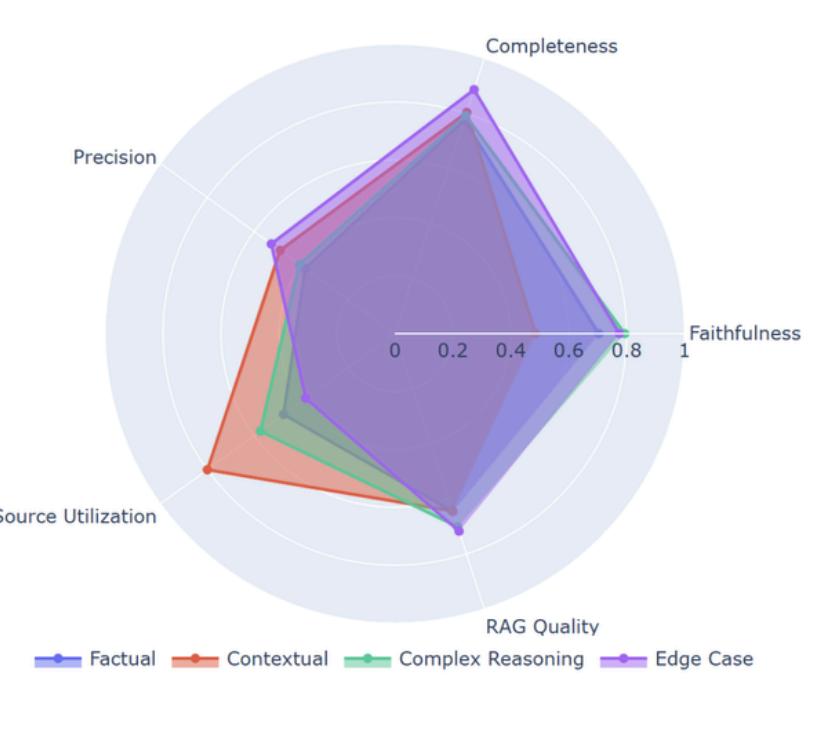
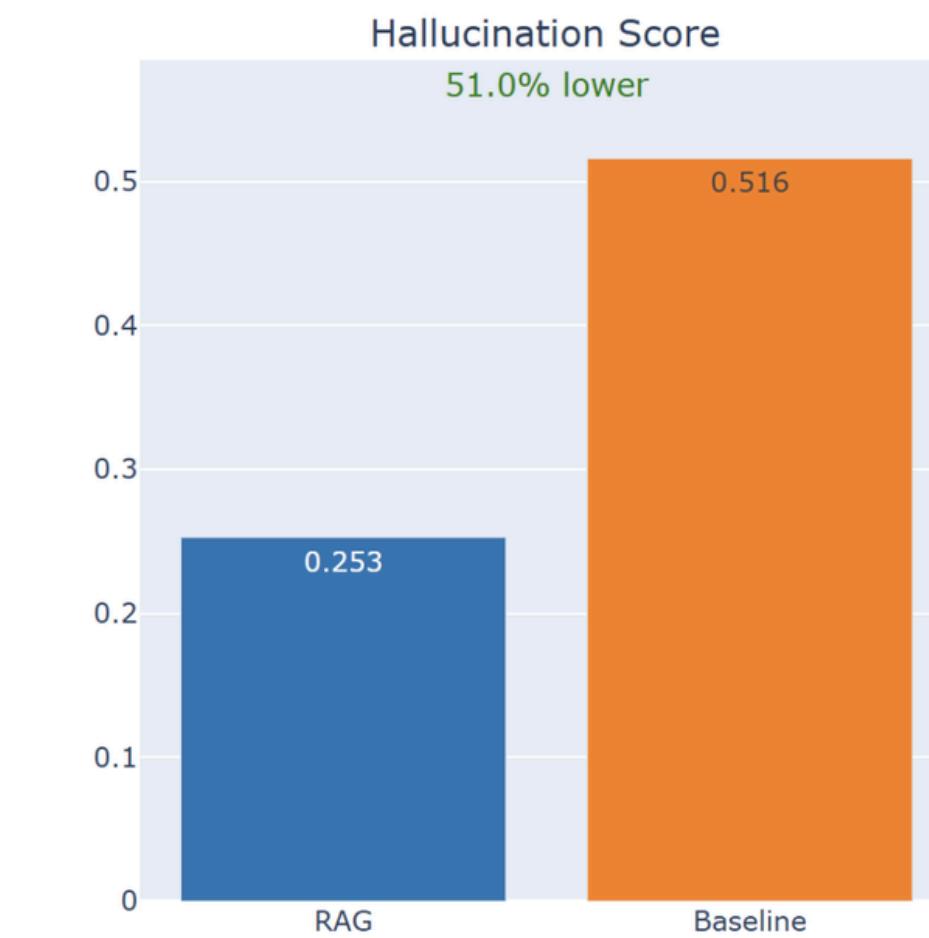
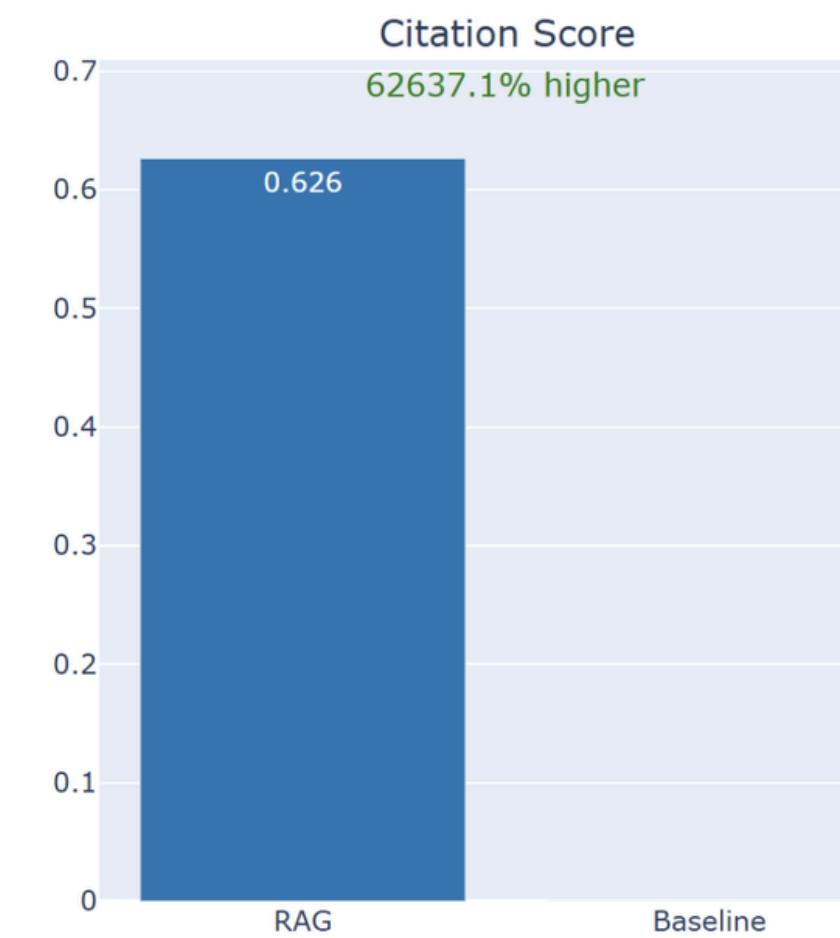
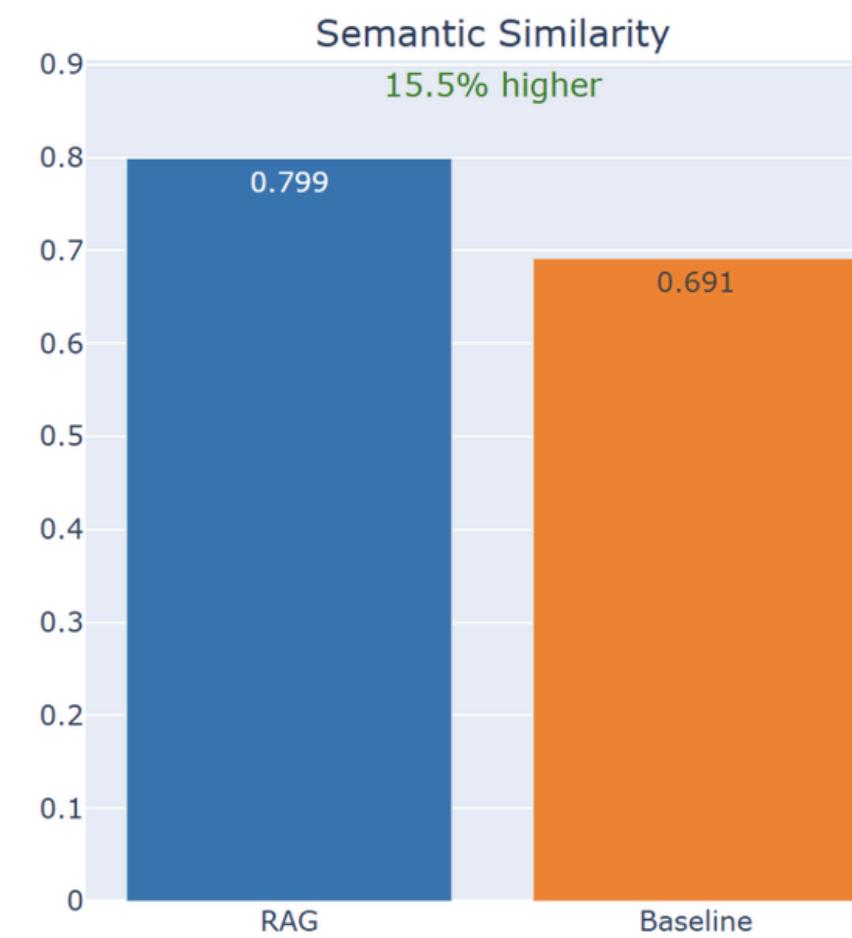
## Comparison with Baseline

Metric	RAG	Baseline	Improvement	% Change	p-value
Semantic Similarity	0.7990	0.6915	<b>0.1075</b>	<b>15.55%</b> better	<b>0.0000</b> (significant)
Citation Score	0.6264	0.0000	<b>0.6264</b>	<b>626371.20%</b> better	<b>0.0000</b> (significant)
Hallucination Score	0.2528	0.5160	<b>0.2632</b>	<b>51.00%</b> better	<b>0.0000</b> (significant)

## RAG Performance Summary

Metric	Value
Total Questions	74
Errors	0
Avg Semantic Similarity	0.8029
Avg Citation Score	0.6060
Avg Hallucination Score	0.2561
Avg Faithfulness	0.6580
Avg Completeness	0.8029
Avg Precision	0.4430
Avg Source Utilization	0.5919
Avg RAG Quality	0.6666

## RAG vs Baseline Comparison



## Performance by Question Type

Question Type	Count	Semantic Similarity	Citation Score	Hallucination Score	RAG Quality
Factual	24	0.7746	0.8795	0.1939	0.6462
Contextual	25	0.8034	0.2908	0.3503	0.6445
Complex Reasoning	15	0.7920	0.5695	0.2748	0.7020
Edge Case	10	0.8860	0.7925	0.1421	0.7174

# Thank you for listening



**Github Repo**

<https://github.com/Arlphuongy/COS30018>

