

Industry Project  
CMI Internship with Alumni



# Document Comparison Chatbot for International Regulatory Compliance

Vishal Maurya  
Soumabha Roy



Shashi Satyam  
Sarvesh Bhandarkar

# Table of Content

**The Why**

**1**

**Project overview**

**2**

**3**

**The How**

**4**

**First approach: RAG**

**5**

**Other approaches**

**6**

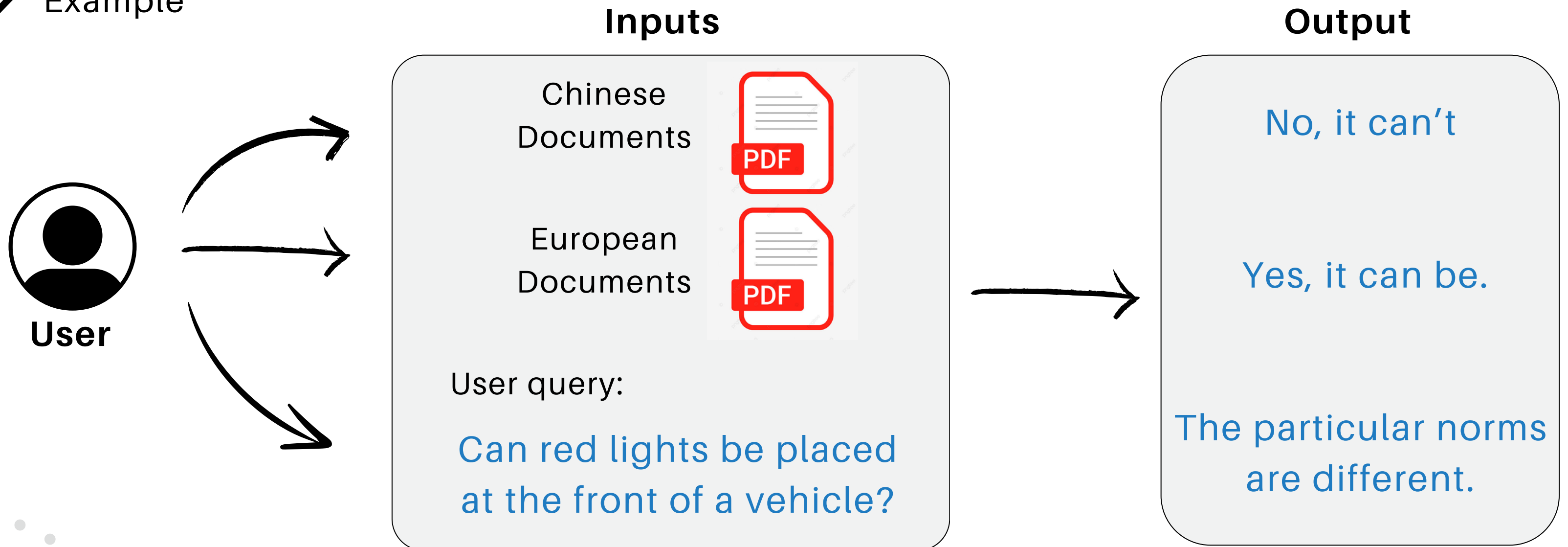
**Results & Evaluation**

**7**

**Future works**

# 1 Project Overview

- Each country have different norms/rules/regulations
- **Objective:** To compare them on basis of user query
  - ↪ **Sub-objective:** To generate an answer for user query
- Example



# 2

# The Why

➤ Vehicle manufacturing company want to sell internationally

➤ Advantages over doing it manually



Reduce cost



Reduce time - 60 hours to compare two pdfs manually



Accuracy



Language discrimination

➤ Why different countries have different norms



Material Cost



Carbon emissions



Weather conditions



Living Standards

# 3 The How



ChatGPT



Restricted norms



Prompting whole pdf to Llama2-chat  
in local computer



Limited context window



Fine-tuning a LLM



Very expensive



RAG



Reduce bias and hallucinations



Specific and Accurate  
information



Less expensive

**LLMs** - Trained for next word prediction

Once upon a



**GPT-3.5**



Once upon a time

Fine-tuned for chat

Write a story



**ChatGPT**

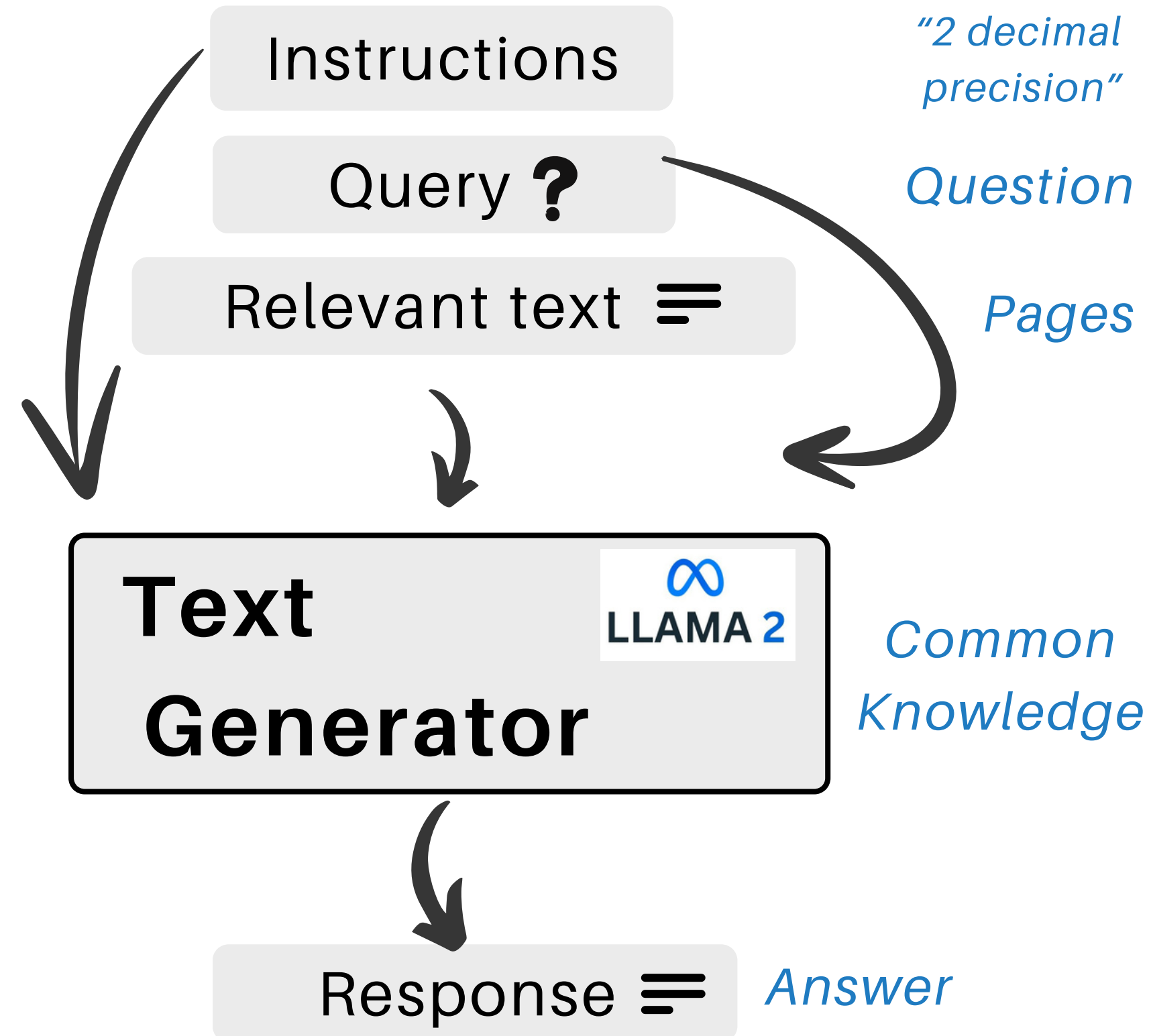
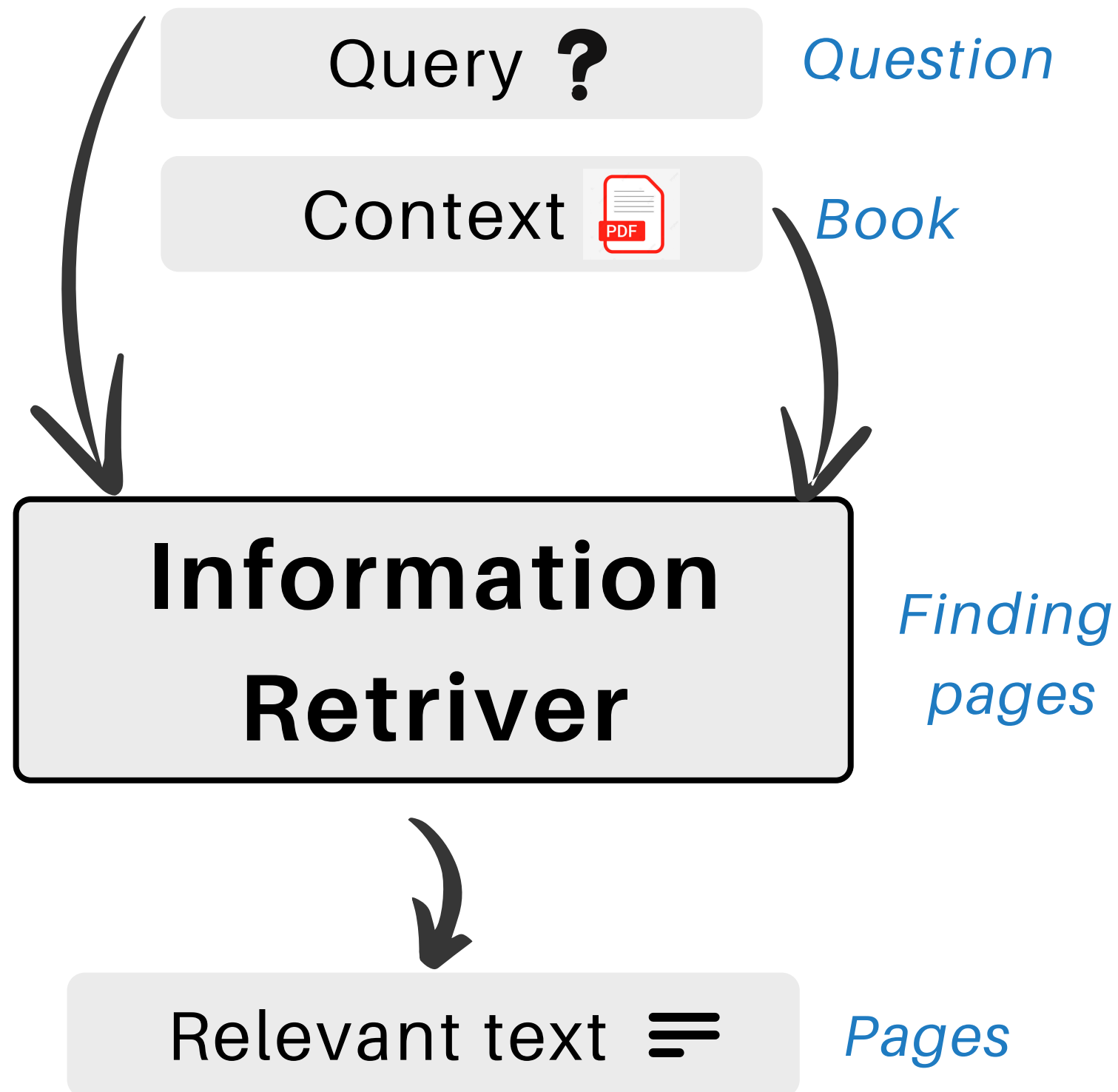


There was a king .....

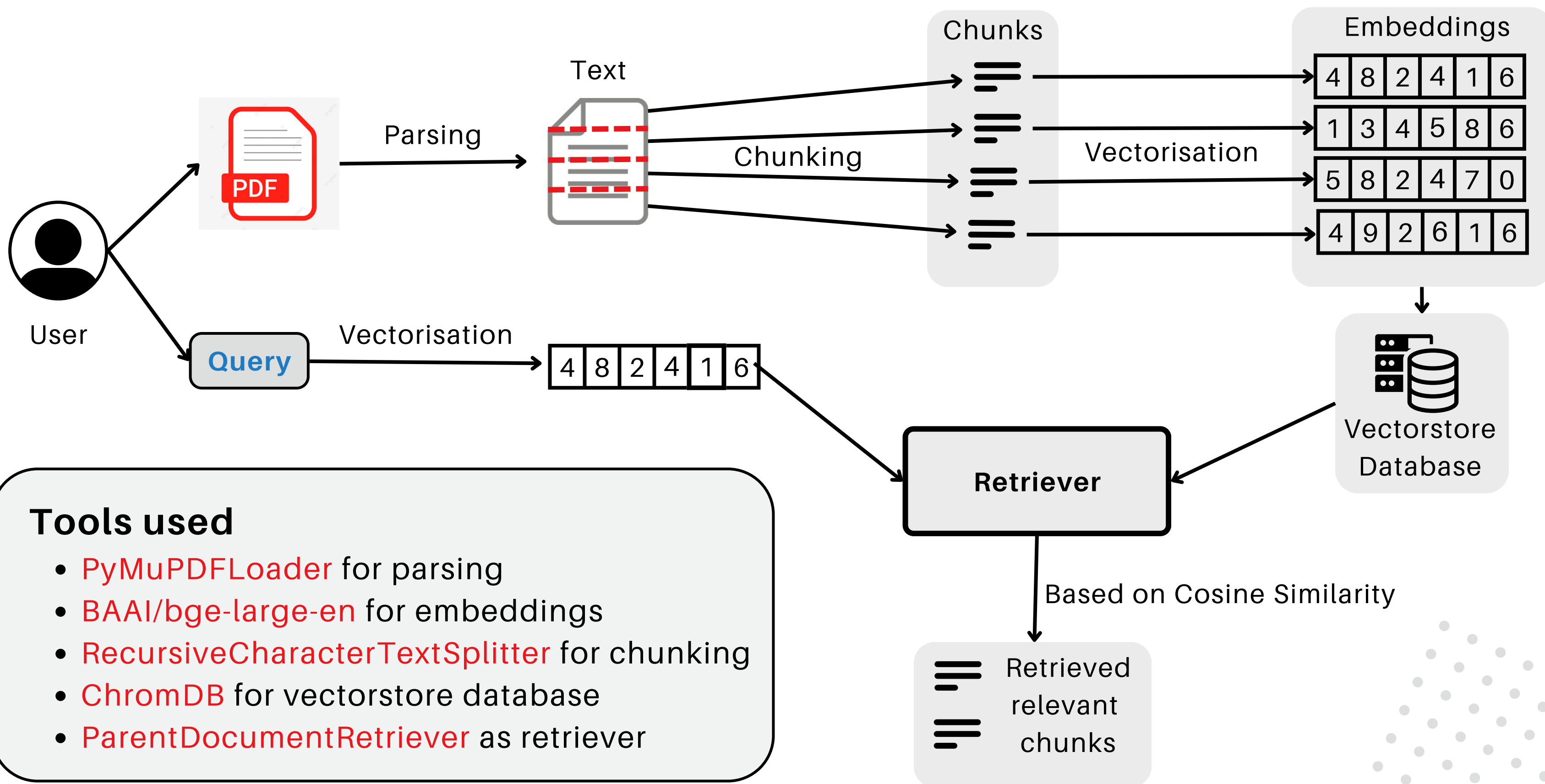
**Fine-tuned LLMs** - for specific task

# 4 First approach: RAG

*Analogy with open-book exam in which student haven't studies*



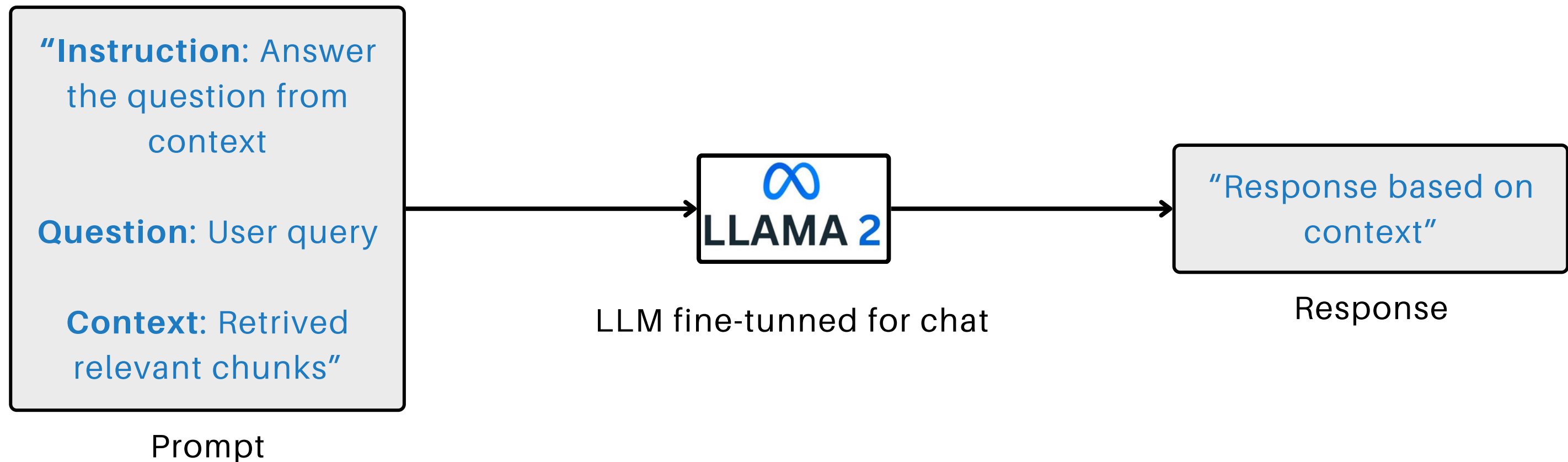
# Information Retrieval



# 4

# Text Generation

Generating response based on context with custom instructions



RAG underperformed

- Complex formatting with 4-5 levels of topic hierarchy
- Smaller chunks often struggle when detailed response is needed
- Larger chunks often struggle when concise answer is needed



# Parent document Retrieval

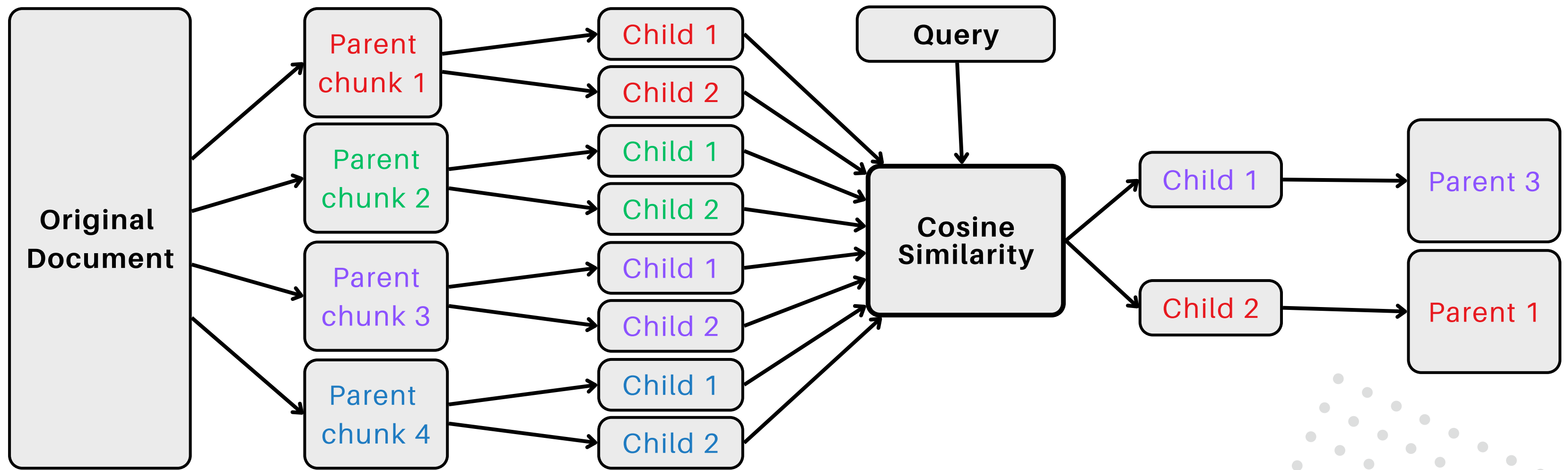
Solves conflicting advantages

Small Documents - accurate embeddings

Large Documents - retaining the context

While retrieval we need smaller chunks

While generation we need larger chunks

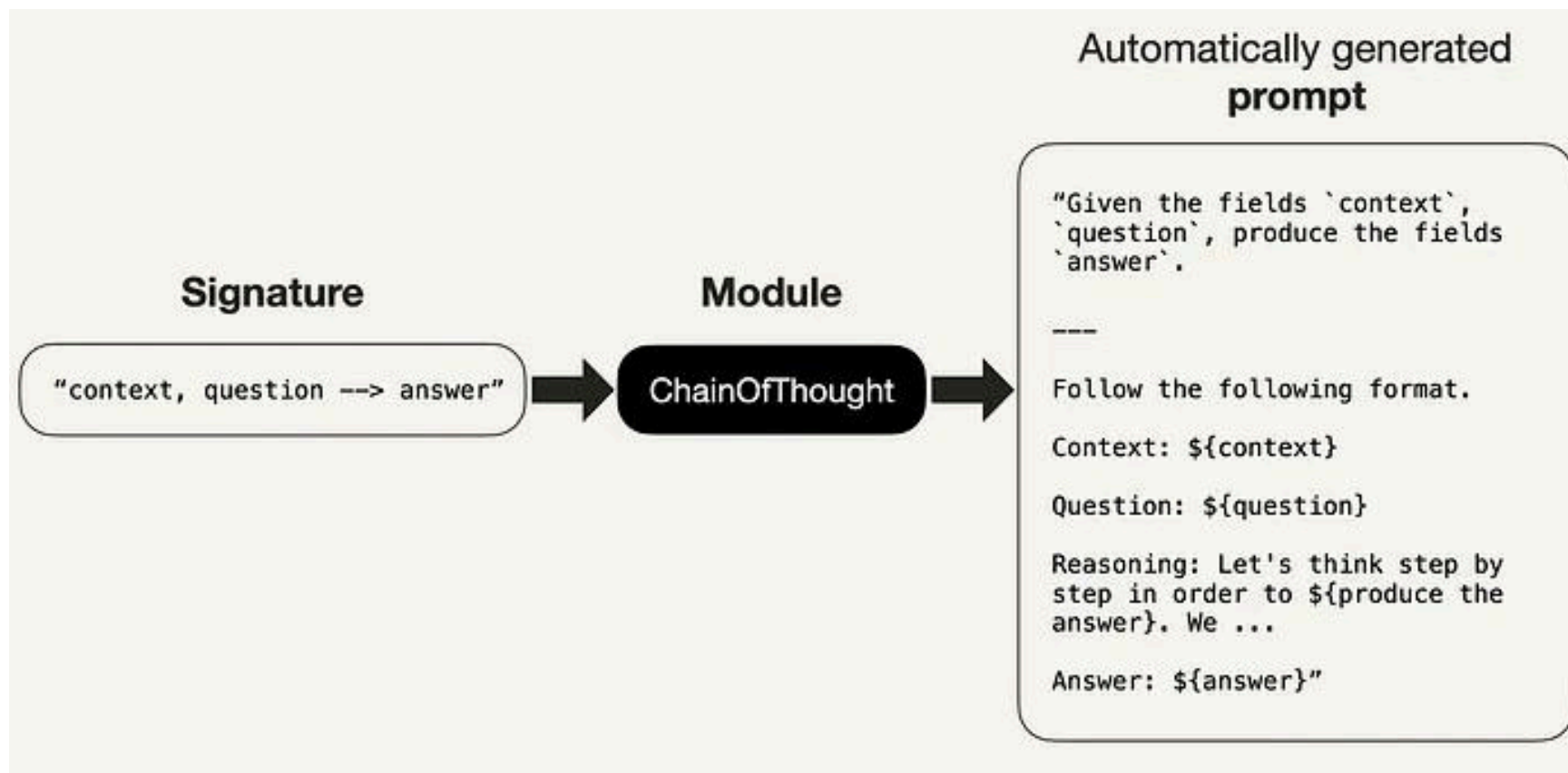


# 5 Other approaches

- Retrieval augmented generation
- Prompting whole pdf
  - ↪ Hallucinates
- Commutative page wise summarisation
- Double retriever RAG
- DSPy

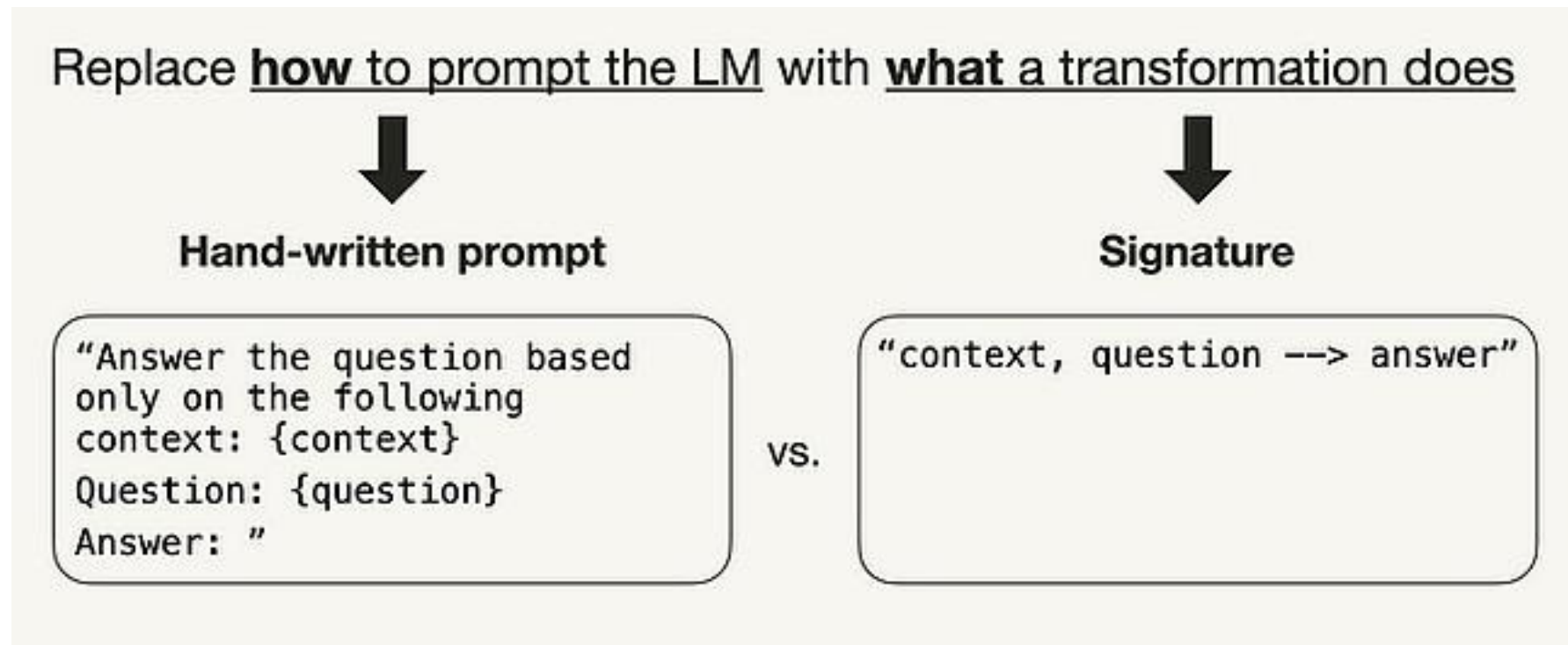
# 5 DSPy

- Declarative Self-improving Language Programs (developed by Stanford University)
- Framework for algorithmically optimizing LM prompts and weights
- Fundamental Components
  - **Signatures:** Abstracting prompting and fine-tuning
  - **Modules:** Abstracting prompting techniques
  - **Teleprompters:** Automating prompting for arbitrary pipelines



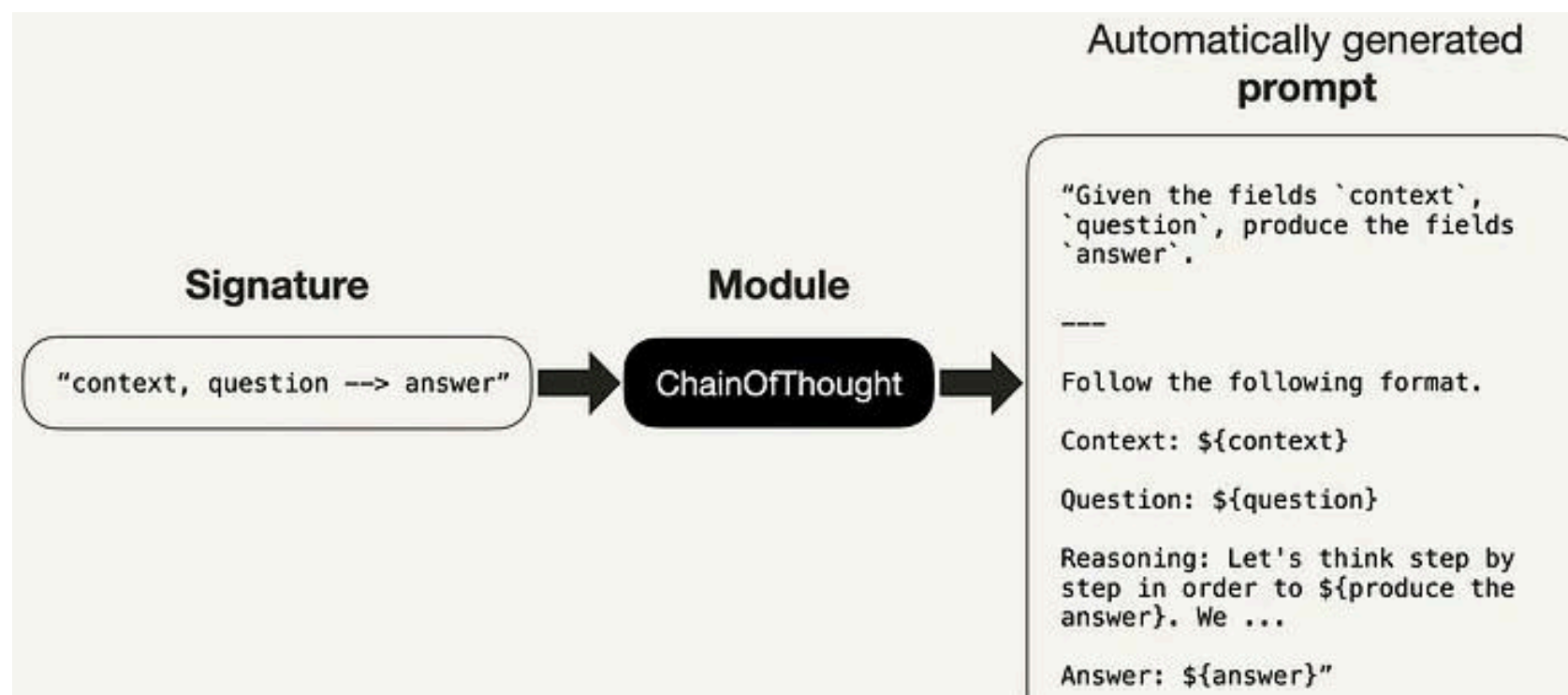
# 5 DSPy

**Signatures:** A contract that specifies the expected input/output dynamics of a module.



- Question Answering as **"question -> answer"**
- Sentiment Classification as **"sentence -> sentiment"**
- Summarization as **"document -> summary"**

**Modules:** A foundational component for constructing programs that leverage LLMs.



- **dspy.Predict**
- **dspy.ChainOfThought**
- **dspy.ProgramOfThought**
- **dspy.ReAct**
- **dspy.MultiChainComparison**

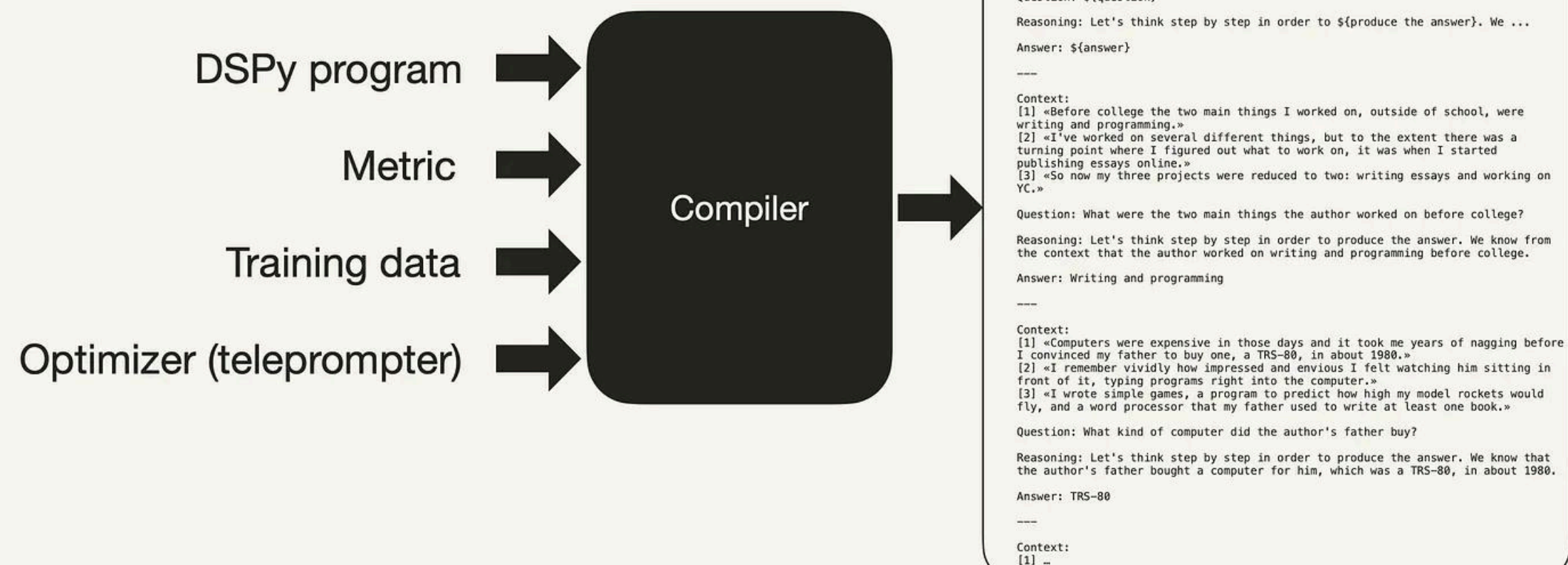
# 5 DSPy

## Optimizers (formerly Teleprompters)

a DSPy optimizer requires three key components:

- **Your DSPy Program**
- **Your Chosen Metric**
- **A Set of Training Inputs**

To adjust the settings of a DSPy program, including both the prompts and the language model (LM) weights, aiming to enhance specified metrics, such as accuracy.





Query

The total speed reduction of the subject vehicle at the time of the collision with the stationary target shall be not less than how many kilometers per hour?

When collision early warning signal shall be sent?

Response

(1) Answer from Context 1:  
The total speed reduction of the subject vehicle at the time of the collision with the stationary target shall be not less than 10km/h, as specified in Context 1.

(2) Answer from Context 2:  
The total speed reduction of the subject vehicle at the time of the collision with the stationary target shall be not less than the value specified in Table I, column D of Annex 3, as specified in Context 2.

(3) Comparison:  
Both answers are different. In Context 1, the speed reduction is specifically limited to 10km/h, while in Context 2, the speed reduction is specified to be not less than the value in Table I, column D of Annex 3. Therefore, the answer from Context 1 is different from the answer in Context 2.

Additionally, it's important to note that the speed reduction values specified in both contexts are related to the service braking system of the vehicle, and are intended to ensure the safety of the vehicle and its occupants during emergency braking situations.

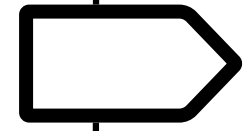
(1) Answer from Context 1:  
When the AEBS (Automatic Emergency Braking System) detects the possibility of a collision with a forward vehicle in the same lane, which is traveling at a slower speed, has slowed to a halt or is stationary, it shall send a collision early warning signal. This signal shall be sent when the AEBS has detected the risk of collision and the timing of the warning is important to enable the driver to react timely and avoid nuisance warnings (4.2.2). The warning signal can be acoustic, haptic or optical, and if an optical signal is used, it may also flash to indicate a failure (4.2.3).

(2) Answer from Context 2:  
In Context 2, the question is asked differently, focusing on the timing of the collision early warning signal. The answer is that the signal shall be sent when the AEBS detects the possibility of a collision with a forward vehicle in the same lane, which is traveling at a slower speed, has slowed to a halt or is stationary (4.2.1). The timing of the warning is important to enable the driver to react timely and avoid nuisance warnings (4.2.2).

(3) Comparison:  
Both answers agree that when the AEBS detects the possibility of a collision with a forward vehicle in the same lane, it shall send a collision early warning signal. They also agree on the importance of timing of the warning to enable the driver to react timely and avoid nuisance warnings. However, Context 1 provides more details on the types of signals that can be used for the collision early warning, while Context 2 focuses more on the timing of the warning signal.

# 6

# Evaluation



Manual



Our target



Correct information



Relevant information



Complete information



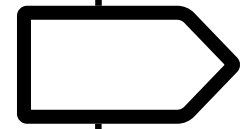
No additional related information



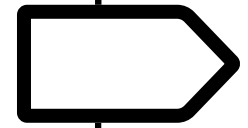
May look for evaluation , after achieving the target.

# 7

# Future Works



Handling multiple languages



Knowledge Database





# Thank you!

**Created and presented by:**

Vishal Maurya (vishal.mds2023@cmi.ac.in)  
Soumabha Roy (soumabha.ug2023@cmi.ac.in)