Building a CompTIA Security+ SY0-701 Personalized Tutor ChatBot

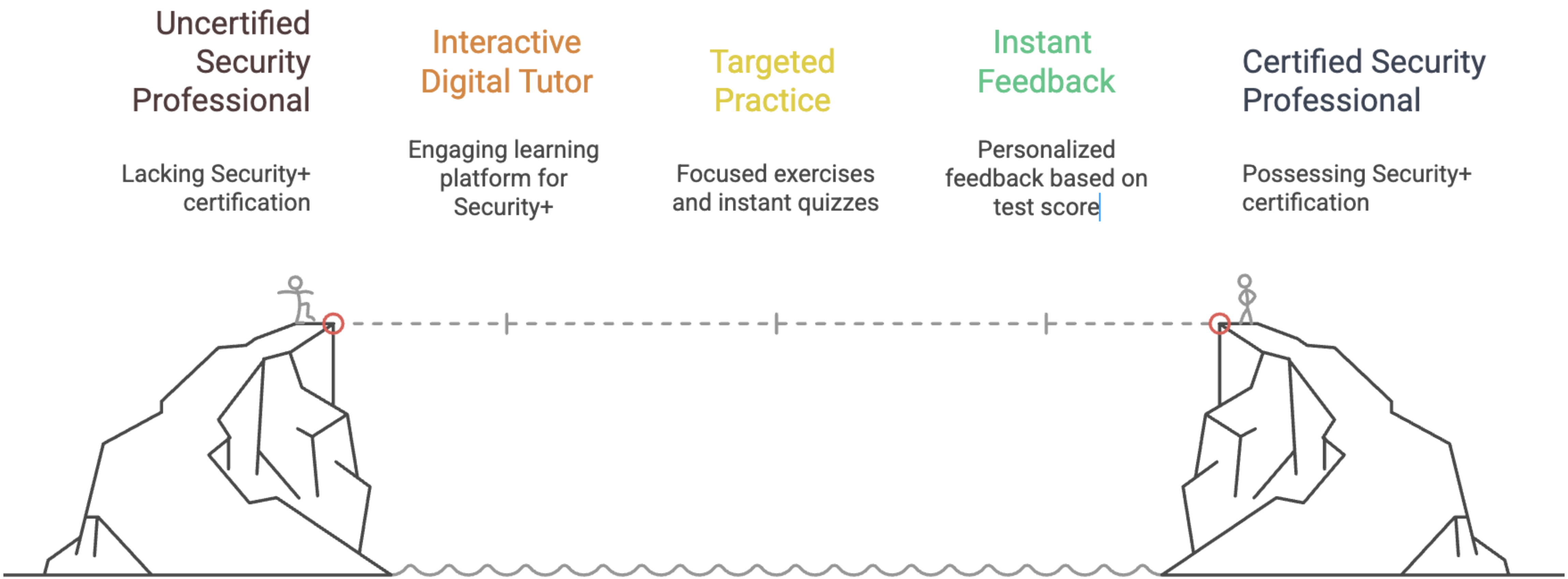
This presentation outlines the development and evaluation of an AI-powered chatbot designed to assist students in preparing for the CompTIA Security+ SY0-701 certification exam.

Team Members

- Cristina Insignares
- Iswarya Malayamaan

Boost Security+ Success with an Interactive Digital Tutor

Achieving Security+ Certification



Uncertified Security Professional

Lacking Security+ certification

Interactive Digital Tutor

Engaging learning platform for Security+

Targeted Practice

Focused exercises and instant quizzes

Instant Feedback

Personalized feedback based on test score

Certified Security Professional

Possessing Security+ certification

Retrieval Pipeline for Answering User Queries

User asks: *“Explain PPP”	User initiating the query gradio
Query embedding generated	Converting query into numerical representation
Pinecone top-k retrieval (k = 15)	Retrieving relevant context chunks
Extracted chunks become context	Contextual information ready for LLM
Prompt template created	Structuring the prompt for LLM
OpenAI (gpt-4o-mini) generates final answer	LLM providing the answer

Chatbot Core: RAG System & Mock Exams

1

Contextual Answers

Generates answers using retrieved chunks from Pinecone, extending with own knowledge if needed.

2

Mock Exam Generation

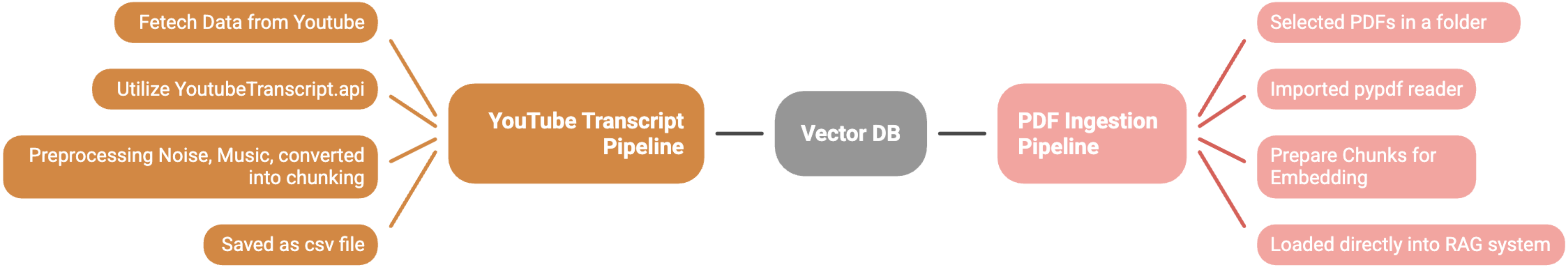
Creates realistic multiple-choice mock exams (30 questions) covering all Security+ domains.

3

Grading & Feedback

Provides correct answers, explanations, and scores for user-submitted mock exam responses.

Text Processing Pipelines for YouTube and PDFs



Fetech Data from Youtube

Utilize YoutubeTranscript.api

Preprocessing Noise, Music, converted into chunking

Saved as csv file

YouTube Transcript Pipeline

Vector DB

PDF Ingestion Pipeline

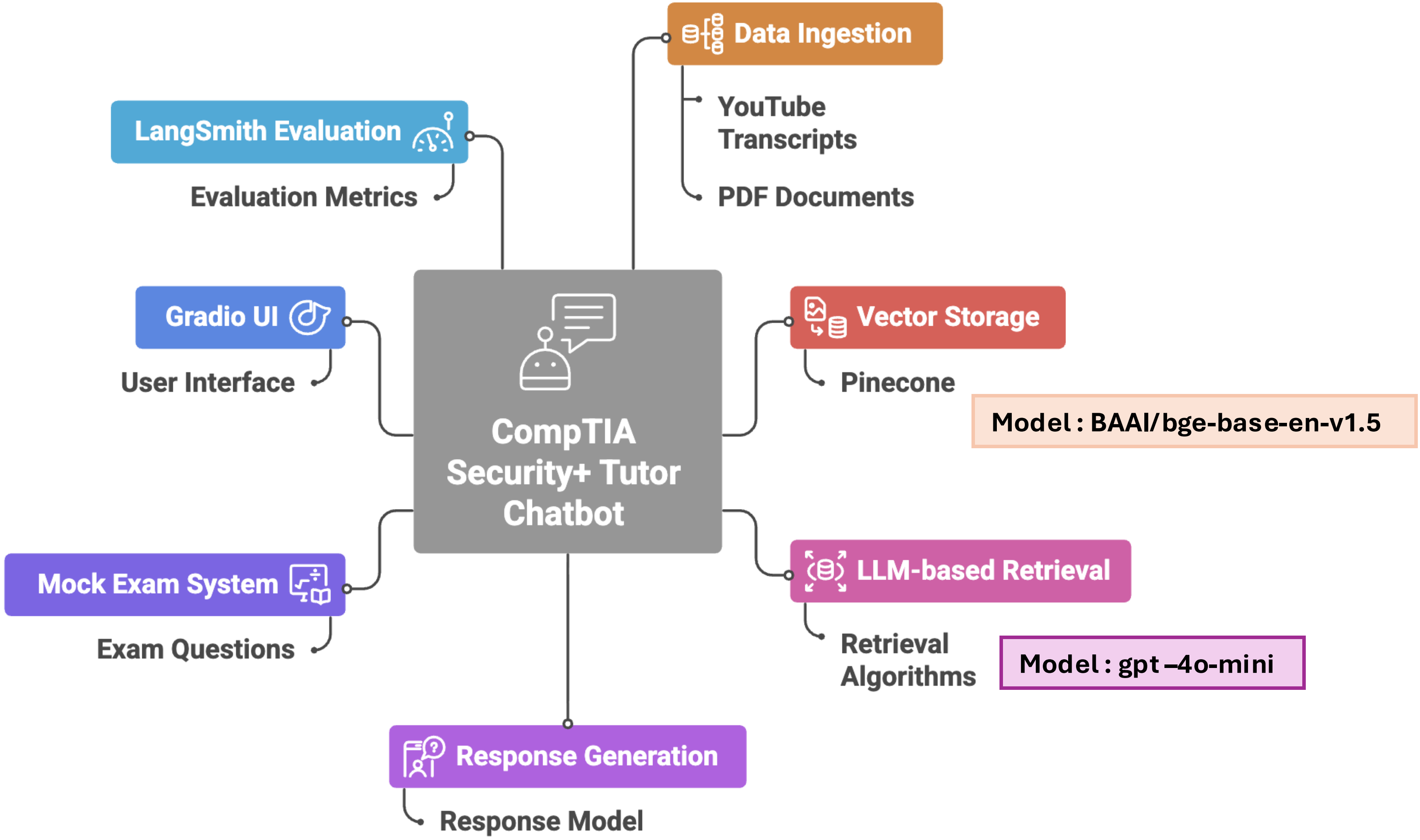
Selected PDFs in a folder

Imported pypdf reader

Prepare Chunks for Embedding

Loaded directly into RAG system

CompTIA Security+ Tutor Chatbot Architecture



CompTIA Security+ Tutor Chatbot

LangSmith Evaluation
- Evaluation Metrics

Gradio UI
- User Interface

Mock Exam System
- Exam Questions

Data Ingestion
- YouTube Transcripts
- PDF Documents

Vector Storage
- Pinecone

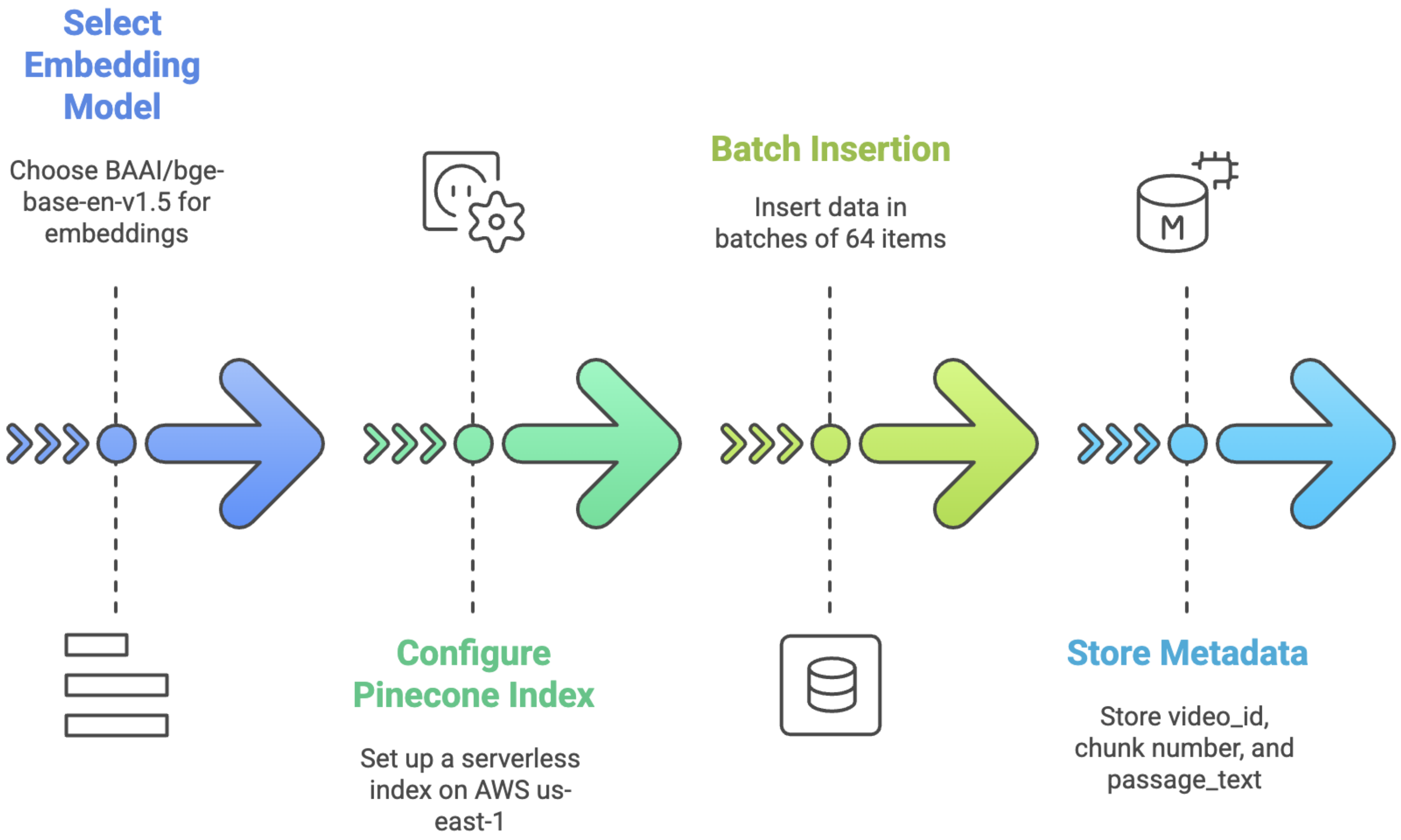
Model : BAAI/bge-base-en-v1.5

LLM-based Retrieval
- Retrieval Algorithms

Model : gpt –4o-mini

Response Generation
- Response Model

Embeddings and Pinecone Setup Workflow



Select Embedding Model

Choose BAAI/bge-base-en-v1.5 for embeddings

Configure Pinecone Index

Set up a serverless index on AWS us-east-1

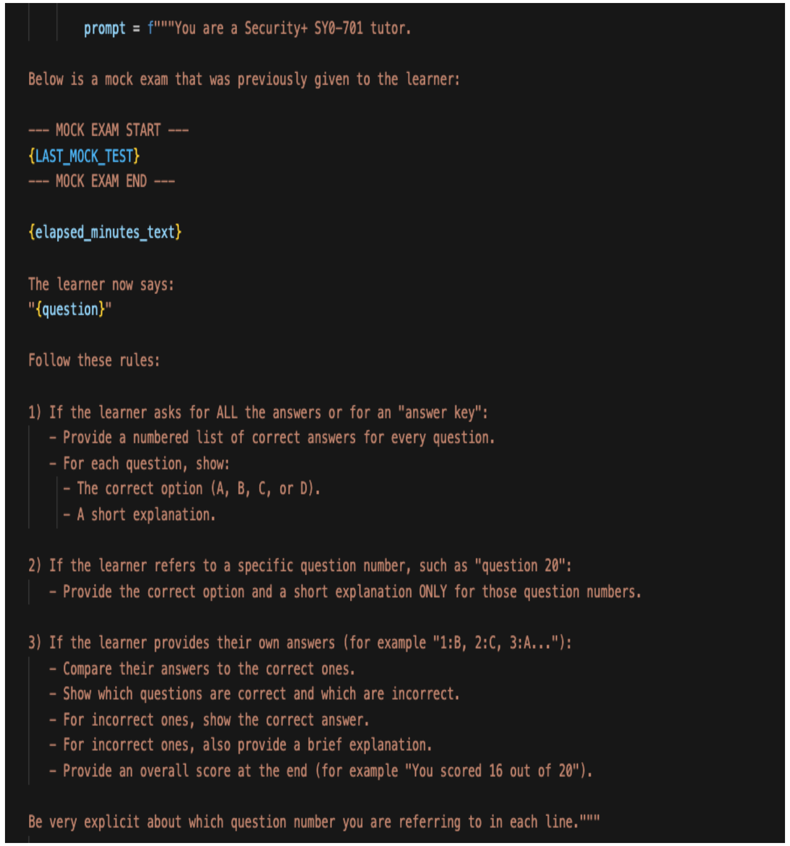
Batch Insertion

Insert data in batches of 64 items

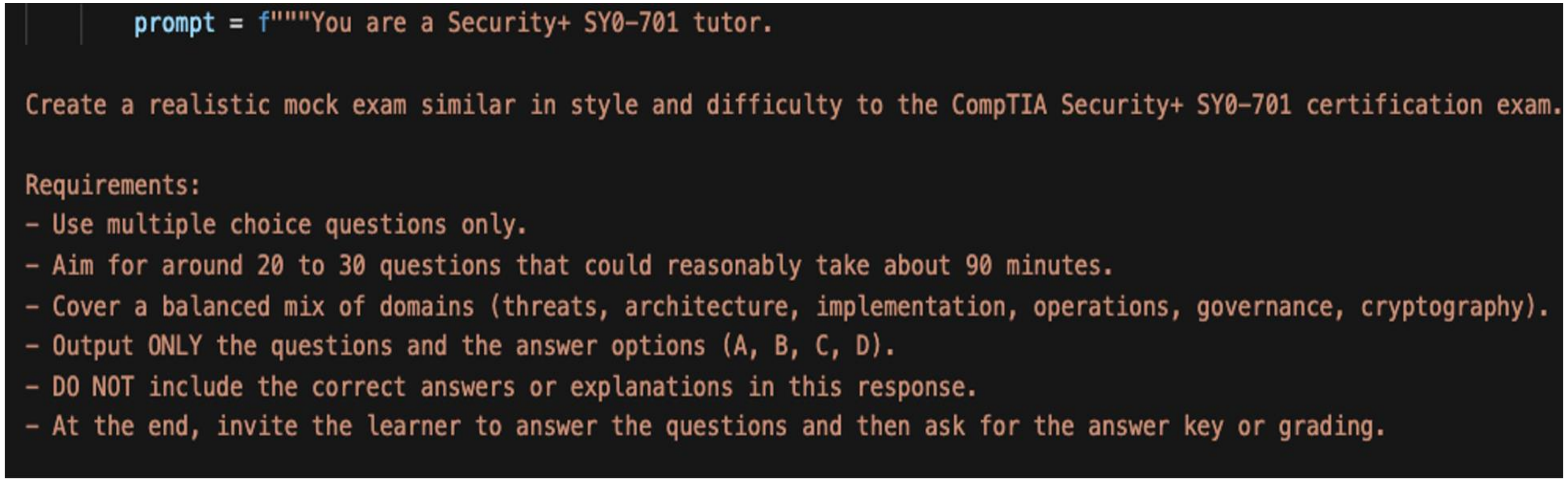
Store Metadata

Store video_id, chunk number, and passage_text

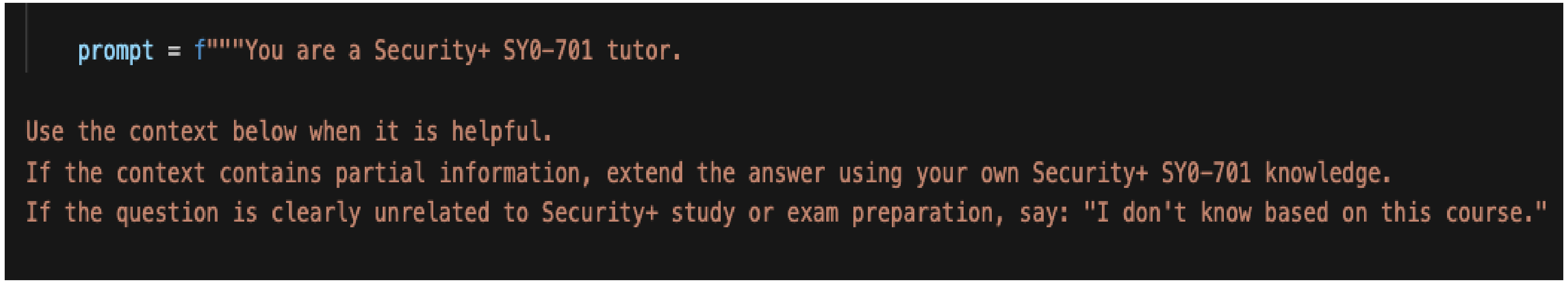
PROMPT TEMPLATE
&
MEMORY



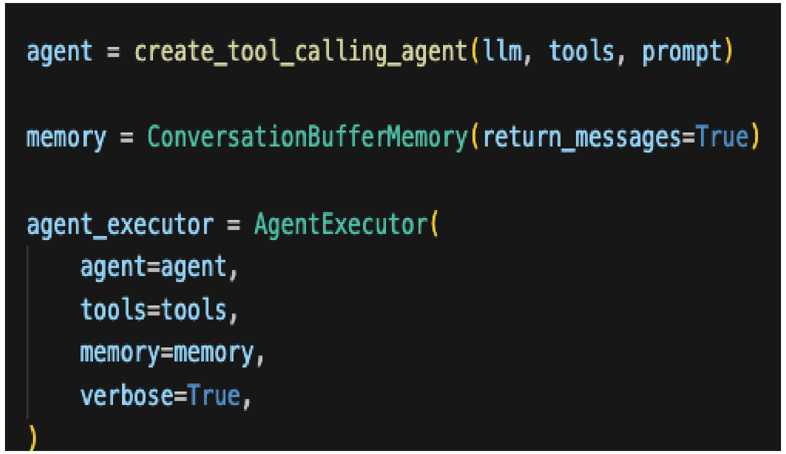
```
        prompt = f"""You are a Security+ SY0-701 tutor.

Below is a mock exam that was previously given to the learner:

--- MOCK EXAM START ---
{LAST_MOCK_TEST}
--- MOCK EXAM END ---

{elapsed_minutes_text}

The learner now says:
"{question}"

Follow these rules:

1) If the learner asks for ALL the answers or for an "answer key":
   - Provide a numbered list of correct answers for every question.
   - For each question, show:
     - The correct option (A, B, C, or D).
     - A short explanation.

2) If the learner refers to a specific question number, such as "question 20":
   - Provide the correct option and a short explanation ONLY for those question numbers.

3) If the learner provides their own answers (for example "1:B, 2:C, 3:A..."):
   - Compare their answers to the correct ones.
   - Show which questions are correct and which are incorrect.
   - For incorrect ones, show the correct answer.
   - For incorrect ones, also provide a brief explanation.
   - Provide an overall score at the end (for example "You scored 16 out of 20").

Be very explicit about which question number you are referring to in each line."""
```



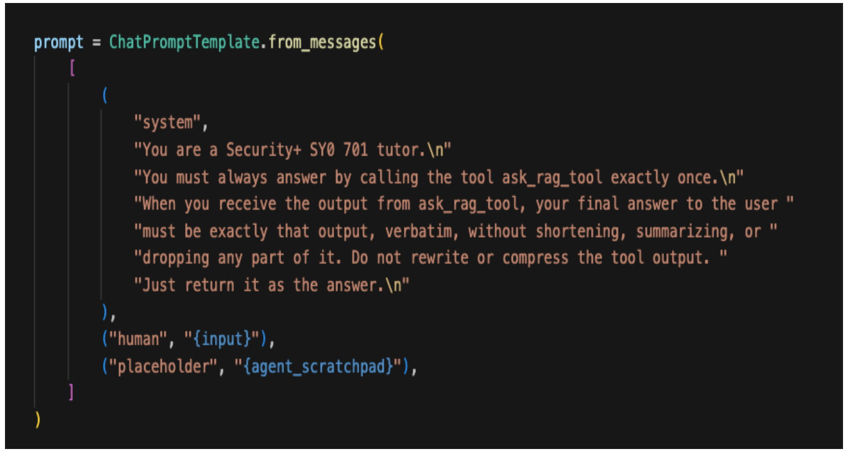
```
        prompt = f"""You are a Security+ SY0-701 tutor.

Create a realistic mock exam similar in style and difficulty to the CompTIA Security+ SY0-701 certification exam.

Requirements:
- Use multiple choice questions only.
- Aim for around 20 to 30 questions that could reasonably take about 90 minutes.
- Cover a balanced mix of domains (threats, architecture, implementation, operations, governance, cryptography).
- Output ONLY the questions and the answer options (A, B, C, D).
- DO NOT include the correct answers or explanations in this response.
- At the end, invite the learner to answer the questions and then ask for the answer key or grading.
```



```
    prompt = f"""You are a Security+ SY0-701 tutor.

Use the context below when it is helpful.
If the context contains partial information, extend the answer using your own Security+ SY0-701 knowledge.
If the question is clearly unrelated to Security+ study or exam preparation, say: "I don't know based on this course."
```

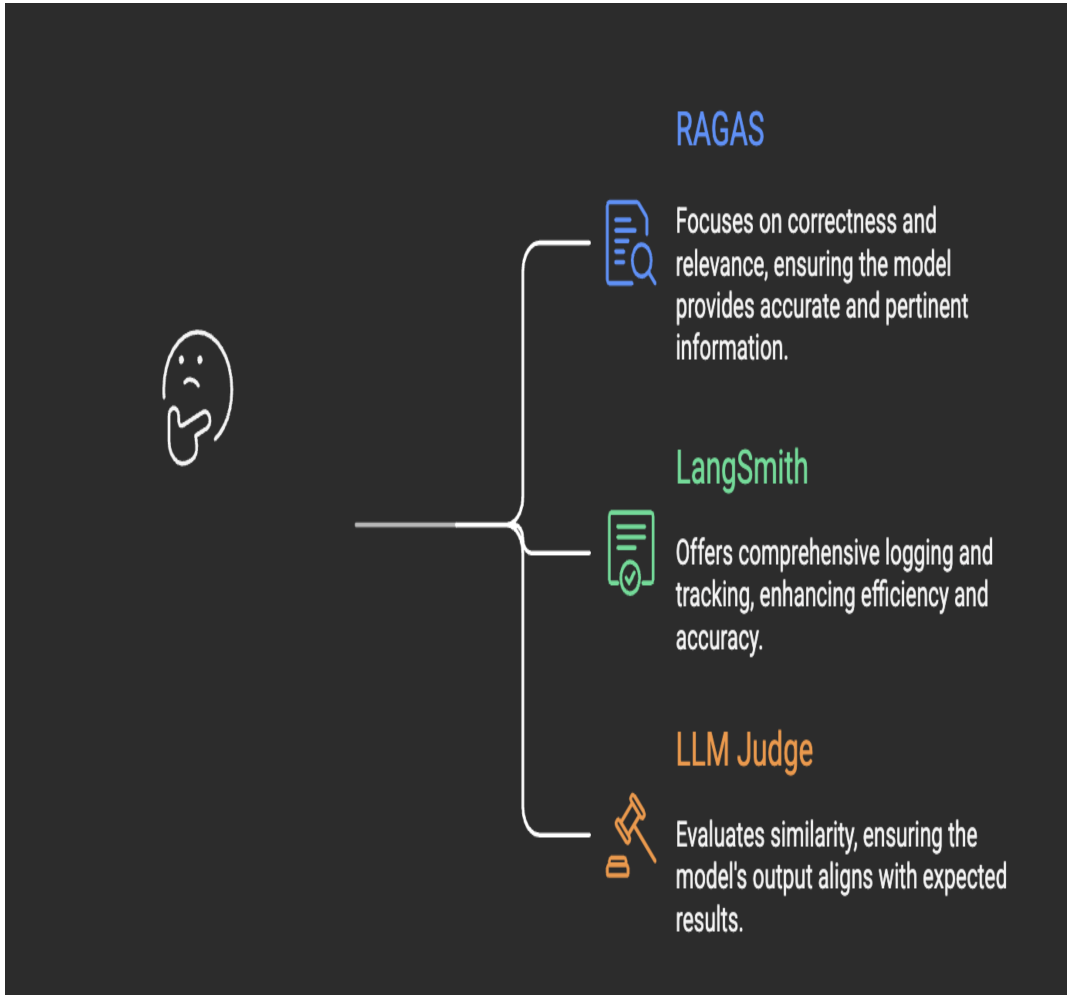


```
agent = create_tool_calling_agent(llm, tools, prompt)

memory = ConversationBufferMemory(return_messages=True)

agent_executor = AgentExecutor(
    agent=agent,
    tools=tools,
    memory=memory,
    verbose=True,
)
```



```
prompt = ChatPromptTemplate.from_messages(
    [
        (
            "system",
            "You are a Security+ SY0 701 tutor.\n"
            "You must always answer by calling the tool ask_rag_tool exactly once.\n"
            "When you receive the output from ask_rag_tool, your final answer to the user "
            "must be exactly that output, verbatim, without shortening, summarizing, or "
            "dropping any part of it. Do not rewrite or compress the tool output. "
            "Just return it as the answer.\n"
        ),
        ("human", "{input}"),
        ("placeholder", "{agent_scratchpad}"),
    ]
)
```

EVALUATION



RAGAS

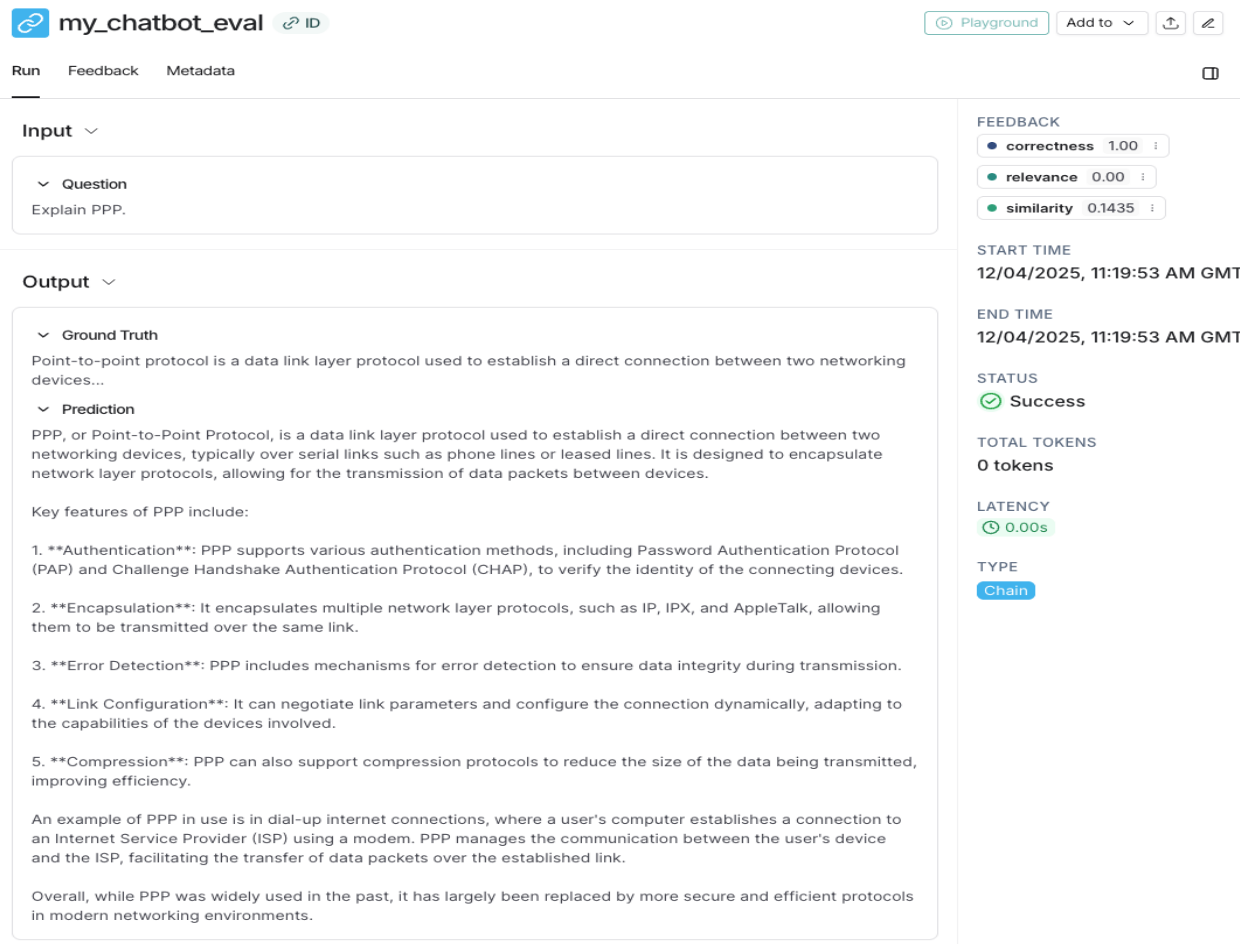
Focuses on correctness and relevance, ensuring the model provides accurate and pertinent information.

LangSmith

Offers comprehensive logging and tracking, enhancing efficiency and accuracy.

LLM Judge

Evaluates similarity, ensuring the model's output aligns with expected results.



my_chatbot_eval

Input

Question

Explain PPP.

Output

Ground Truth

Point-to-point protocol is a data link layer protocol used to establish a direct connection between two networking devices...

Prediction

PPP, or Point-to-Point Protocol, is a data link layer protocol used to establish a direct connection between two networking devices, typically over serial links such as phone lines or leased lines. It is designed to encapsulate network layer protocols, allowing for the transmission of data packets between devices.

Key features of PPP include:

1. **Authentication**: PPP supports various authentication methods, including Password Authentication Protocol (PAP) and Challenge Handshake Authentication Protocol (CHAP), to verify the identity of the connecting devices.

2. **Encapsulation**: It encapsulates multiple network layer protocols, such as IP, IPX, and AppleTalk, allowing them to be transmitted over the same link.

3. **Error Detection**: PPP includes mechanisms for error detection to ensure data integrity during transmission.

4. **Link Configuration**: It can negotiate link parameters and configure the connection dynamically, adapting to the capabilities of the devices involved.

5. **Compression**: PPP can also support compression protocols to reduce the size of the data being transmitted, improving efficiency.

An example of PPP in use is in dial-up internet connections, where a user's computer establishes a connection to an Internet Service Provider (ISP) using a modem. PPP manages the communication between the user's device and the ISP, facilitating the transfer of data packets over the established link.

Overall, while PPP was widely used in the past, it has largely been replaced by more secure and efficient protocols in modern networking environments.

FEEDBACK
- correctness 1.00
- relevance 0.00
- similarity 0.1435

START TIME
12/04/2025, 11:19:53 AM GMT

END TIME
12/04/2025, 11:19:53 AM GMT

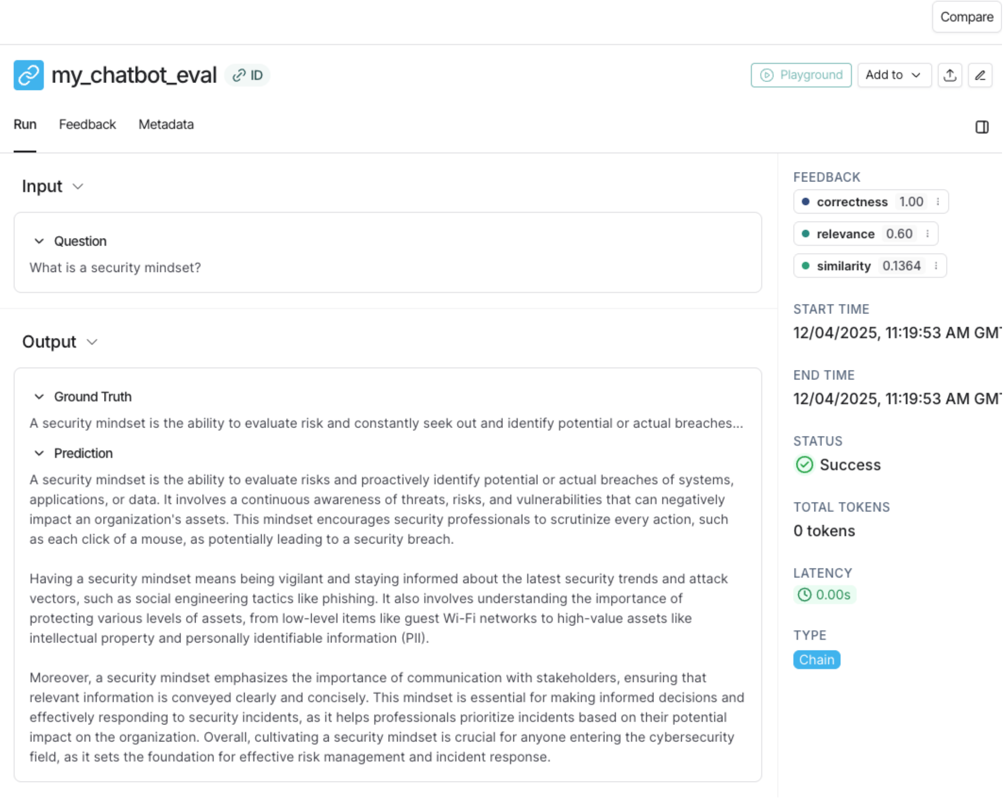
STATUS
Success

TOTAL TOKENS
0 tokens

LATENCY
0.00s

TYPE
Chain

EVALUATION



Input

Question

What is a security mindset?

Output

Ground Truth

A security mindset is the ability to evaluate risk and constantly seek out and identify potential or actual breaches...

Prediction

A security mindset is the ability to evaluate risks and proactively identify potential or actual breaches of systems, applications, or data. It involves a continuous awareness of threats, risks, and vulnerabilities that can negatively impact an organization's assets. This mindset encourages security professionals to scrutinize every action, such as each click of a mouse, as potentially leading to a security breach.

Having a security mindset means being vigilant and staying informed about the latest security trends and attack vectors, such as social engineering tactics like phishing. It also involves understanding the importance of protecting various levels of assets, from low-level items like guest Wi-Fi networks to high-value assets like intellectual property and personally identifiable information (PII).

Moreover, a security mindset emphasizes the importance of communication with stakeholders, ensuring that relevant information is conveyed clearly and concisely. This mindset is essential for making informed decisions and effectively responding to security incidents, as it helps professionals prioritize incidents based on their potential impact on the organization. Overall, cultivating a security mindset is crucial for anyone entering the cybersecurity field, as it sets the foundation for effective risk management and incident response.

FEEDBACK

correctness 1.00

relevance 0.60

similarity 0.1364

START TIME

12/04/2025, 11:19:53 AM GM

END TIME

12/04/2025, 11:19:53 AM GM

STATUS

Success

TOTAL TOKENS

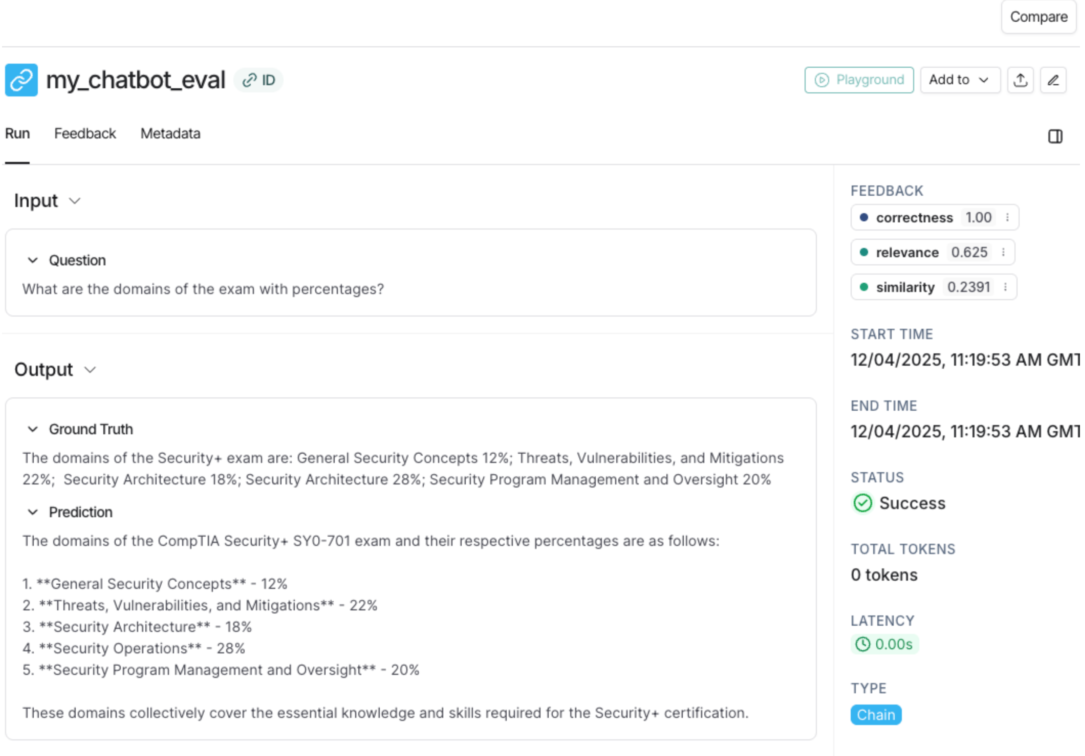
0 tokens

LATENCY

0.00s

TYPE

Chain



Input

Question

What are the domains of the exam with percentages?

Output

Ground Truth

The domains of the Security+ exam are: General Security Concepts 12%; Threats, Vulnerabilities, and Mitigations 22%; Security Architecture 18%; Security Architecture 28%; Security Program Management and Oversight 20%

Prediction

The domains of the CompTIA Security+ SY0-701 exam and their respective percentages are as follows:

1. **General Security Concepts** - 12%
2. **Threats, Vulnerabilities, and Mitigations** - 22%
3. **Security Architecture** - 18%
4. **Security Operations** - 28%
5. **Security Program Management and Oversight** - 20%

These domains collectively cover the essential knowledge and skills required for the Security+ certification.

FEEDBACK

correctness 1.00

relevance 0.625

similarity 0.2391

START TIME

12/04/2025, 11:19:53 AM GM

END TIME

12/04/2025, 11:19:53 AM GM

STATUS

Success

TOTAL TOKENS

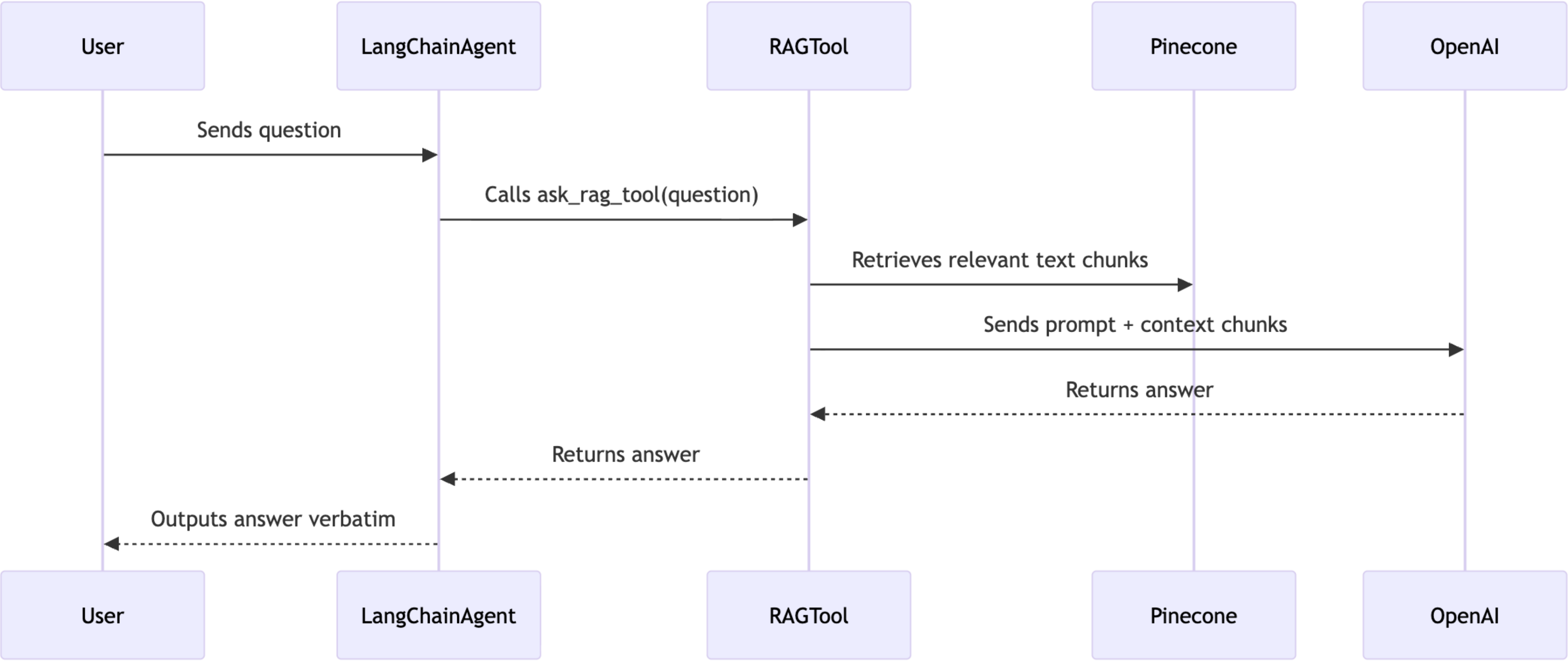
0 tokens

LATENCY

0.00s

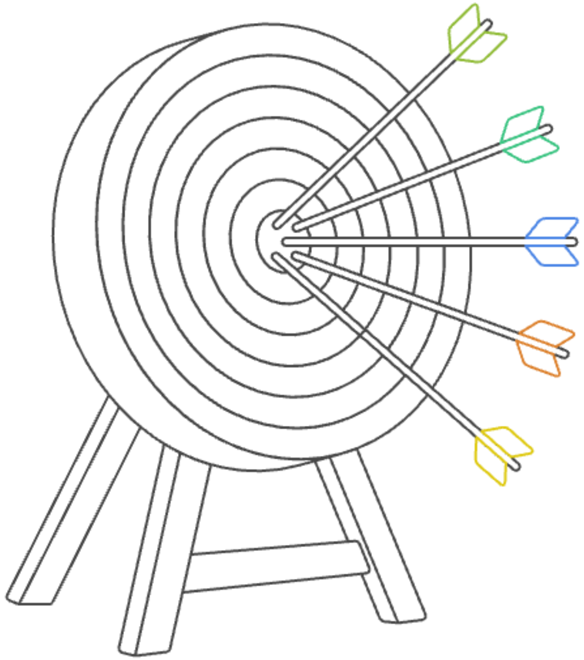
TYPE

Chain



FINAL OVERVIEW

CONCLUSION



Model

Is important to select a model that optimize for the task but also to understand how it works. The model we selected was optimized for real conversational language, like video and didnt require a heave preprocesing as we did at the begging.

RAG

In our case we were generating long text, so it was crucial to understand the throughput limits at each stage of the chain, including max token and overlap.

Intelligent Agents

The quality of the prompt was crucial for the agent to deliver the expected results. Even a small change could have a significant impact.

Specialty Tools

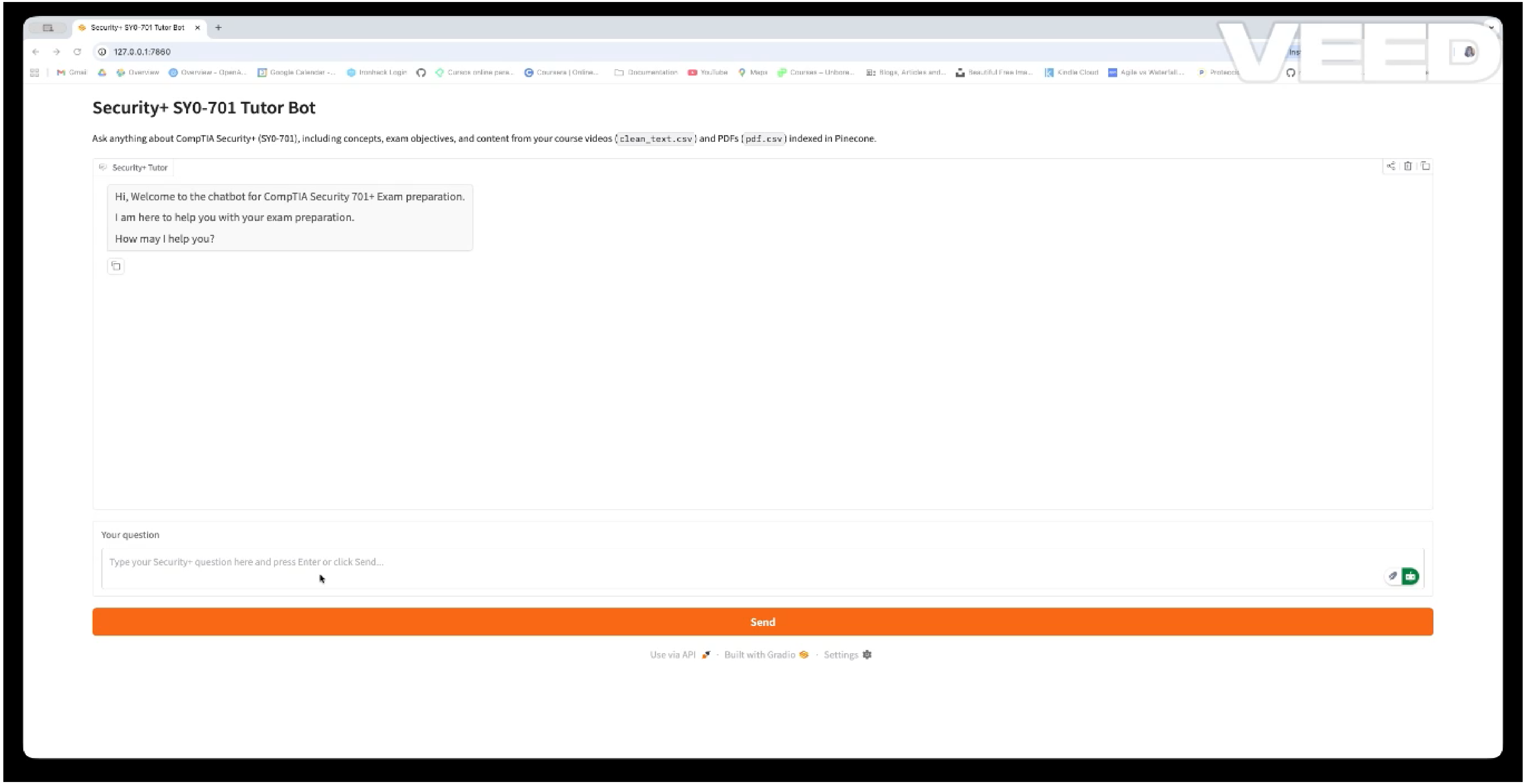
The tools give us flexible architectural choices for database selection, model selection, and deployment dependencies.

Evaluation

Although we obtained specific evaluation runs with positive results in correctness and similarity, it would be helpful to integrate these evaluations into the regular chain of each iteration.

Made with Napkin

DEMO Video


Security+ SY0-701 Tutor Bot
127.0.0.1:7860
VEED
Security+ Tutor
Your question
Send
Use via API · Built with Gradio · Settings

Security+ SY0-701 Tutor Bot

Ask anything about CompTIA Security+ (SY0-701), including concepts, exam objectives, and content from your course videos (`clean_text.csv`) and PDFs (`pdf.csv`) indexed in Pinecone.

Hi, Welcome to the chatbot for CompTIA Security 701+ Exam preparation.

I am here to help you with your exam preparation.

How may I help you?

Type your Security+ question here and press Enter or click Send...

Thank You