

# CAIVA



EELU  
THE EGYPTIAN E-LEARNING UNIVERSITY

2024



## Our Team



**Alaa**  
Badr Hussein



**Shahd**  
Hesham Ibrahim



**Mohamed**  
Allam Mohamed



**Hossam Eldin**  
Tammam Gad



**Abdelrahaman**  
Ahmed Samier



**Yomna**  
Gamal Hussein



**Dina**  
Mohamed Ali

## Supervised

**Dr. Mahmoud Bassioni**  
**Eng. Hayam Abdelbaset**

# Agenda



01

## INTRODUCTION

What is the origin of the name “Caiva” ?

02

## OUR PROJECT

What is Caiva ?

03

## PROBLEM STATEMENT

What specific challenges or issues led to the inception of Caiva ?

04

## SOLUTION

In what manner does Caiva introduce an innovative approach or solution?

05

## DIFFERENCE

Comparison between Caiva and similar models such as Alexa, Siri, ...

06

## KEY FEATURES

What are the key features that make Caiva unique?

07

## METHODOLOGY

The structured approach and process we follow to effectively convey information to an audience.

08

## TECHNOLOGY USED

The different combinations of technology used to achieve Caiva's goals.

09

## DEMO

It's time to talk to our virtual character "CAIVA"

10

## CHALLENGES & TIME PLAN

The challenges and difficulties we went through & Plan and manage project tasks efficiently.

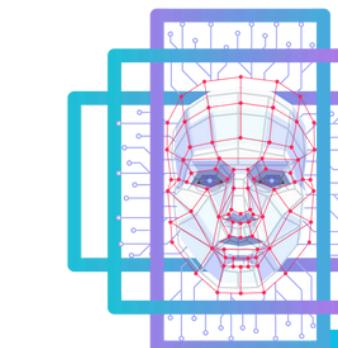


# CAIVA

>>>

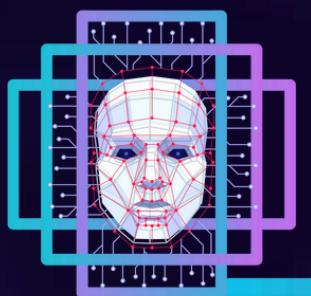
# Introduction





**CAIVA**

Conversational  
Artificial  
Intelligence  
Virtual  
Assistant

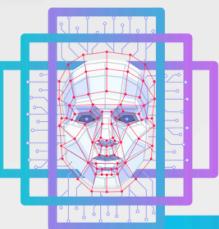


# CAIVA

>>>

# Our Project





# CAIVA



Our solution for these problems that we mentioned is Caiva

## so, what is Caiva ?

- Caiva is not merely an assistant but a proficient performer of tasks, streamlining processes with ease and efficiency.
- Caiva boasts remarkable flexibility, requiring only the necessary database and expected commands to execute various kinds of tasks swiftly.
- It's a customizable character, allowing users to craft their desired persona effortlessly.
- Supported by sentimental animations, Caiva strives to create an authentic user experience, ensuring users feel a genuine connection with the character



>>>

# Problem Statement





&gt;&gt;&gt;

In today's fast-paced digital era, there is a growing demand for virtual smart assistants that can efficiently assist users in various tasks, ranging from managing schedules and providing information to executing commands and performing complex interactions with humans and patients in different aspects of life.



## Problem Statement

Most technologies and machines can be hard to use and its not tailored for :

- The Elderly
- Illiterate people
- Blind people
- Disabilities people



## Research Highlights for EGYPT in 2023



8.6%

Elderly People



12.6%

illiteracy



9.3%

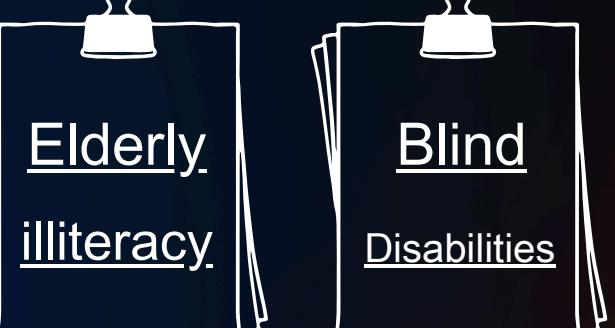
Blind People



10.6%

Disabilities

Sources :

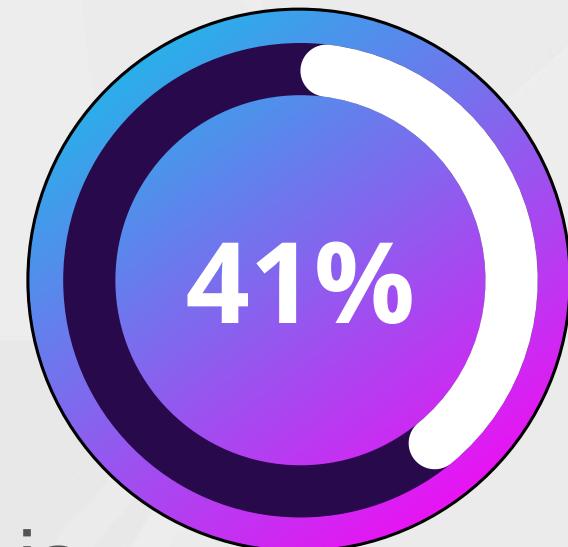




## Problem Statement

### Loneliness and depression

- The prevalence of loneliness and of moderate depression was **41% in the world**.
- It's the invisible struggle that knows no boundaries.
- The global suicide rate among men is 3 to 4 times higher than that of women.



Sources :





# CAIVA

>>>

# Solution





# CAIVA

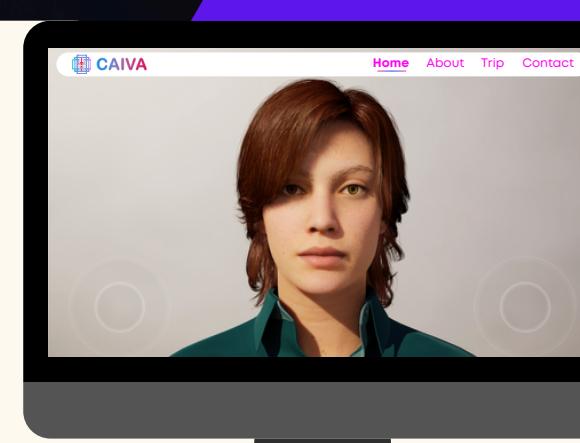


## Caiva is not just made for one use or one purpose

it's a multi-purpose that can help you in every day life, helping you do various amounts of tasks and for every environment their is a model that do multiple of different tasks

1

### Websites



3

### Smart Watches



2

### Mobile Applications



4

### Car Touchscreen





&gt;&gt;&gt;

# Solution for Machines

- Implement AI metahuman characters for voice-driven interactions, reducing reliance on complex touch screens.
- Enable users to interact naturally with AI characters through spoken commands and queries, enhancing intuitiveness.
- Enhance security and simplify authentication, making it accessible for users with memory or literacy challenges.
- Provide step-by-step guidance for various transactions, catering to illiterate, blind, disabled individuals, and the elderly.

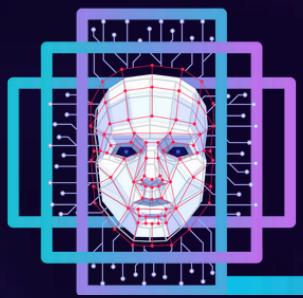




# Solution for Loneliness & Depression

- Addressing the profound issues of loneliness and depression through an innovative approach, Caiva our solution offers a fully animated character designed to express emotions, providing a virtual companion that feels remarkably human.
- With this animated ally, users can engage in judgment-free conversations, pour out their hearts, and share problems, receiving empathy and support from the character.
- Our solution goes beyond mere interaction; it offers a lifeline to connection, providing a compassionate companion ready to make a difference in users' lives, ensuring they are not alone in their journey.





# CAIVA

>>>

# Difference



# Difference



## Alexa

- **Developed by:** Amazon
- **Integration:** Used in Amazon Echo and other smart devices
- **Capabilities:** Voice-controlled virtual assistant, smart home integration, third-party skills, music playback, shopping, and more...



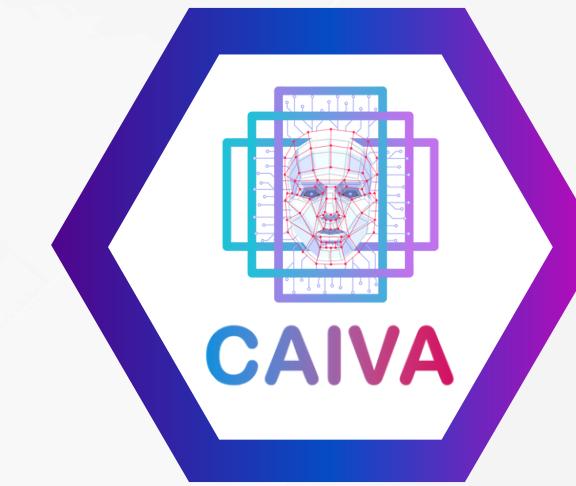
## Siri

- **Developed by:** Apple
- **Integration:** Exclusive to Apple devices (iPhone, iPad, Mac)
- **Capabilities:** Voice recognition, natural language processing, integrates with Apple ecosystem, performs tasks, provides information.



## Google Assistant

- **Developed by:** Google
- **Integration:** Available on Android devices, Google Home, and other platforms
- **Capabilities:** Voice-activated, search engine integration, smart home control, natural language understanding, supports a wide range of devices and services.



## Caiva

We will know what are the main features that make Caiva different and unique in  
**"Key Features slide"**



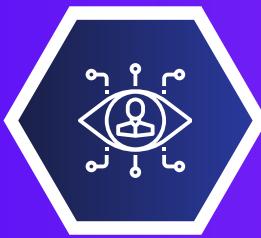
# CAIVA

>>>

## Key Features



# Key Features



**Emotional Intelligence**



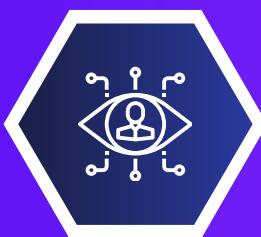
**Tailored Support**



**Human-Centered Design**



**Integrable in multi-Devices**



**Companion Experience**



**Privacy and Security**



**Interactive Animations**



**Continuous Improvement**



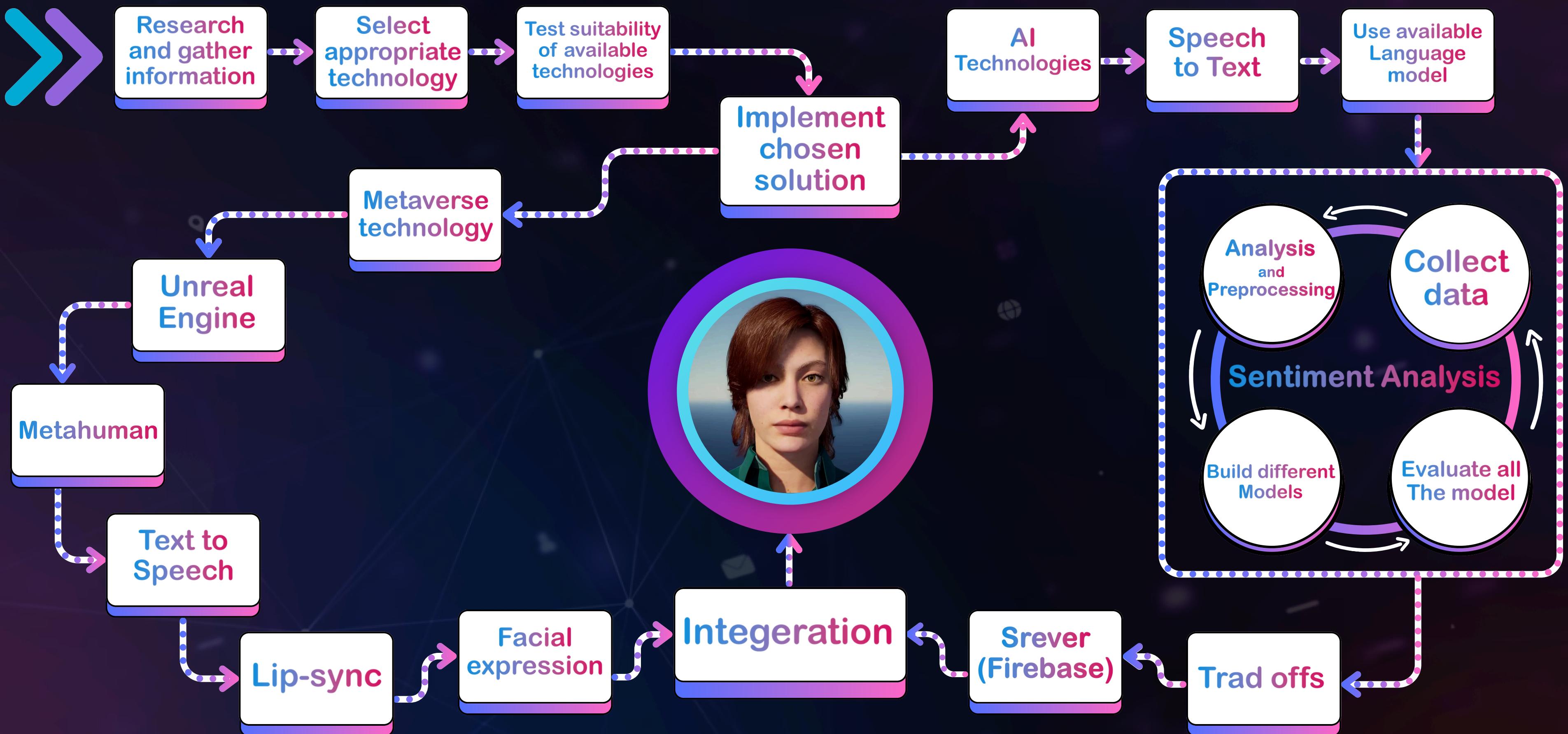
# CAIVA

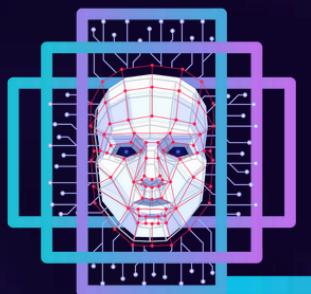
>>>

# Methodology



# Methodology





# CAIVA

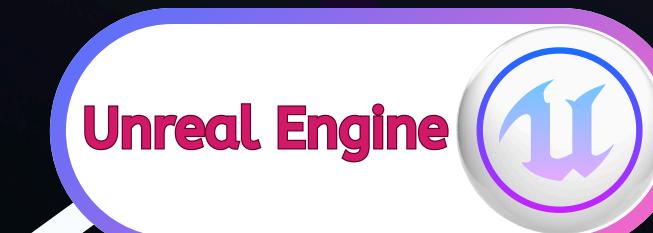
>>>

# Technology Used





&gt;&gt;&gt;



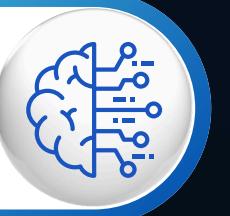
**Geppetto**



**Ariel Voice Generation**



**Language Model**



**Sentiment Analysis**



**Speech to Text**



# Why Unreal Engine ?

## Unreal Engine



## Unity

- ✓ Cross-platform
- ✓ Epic Games
- ✓ C++ for development
- Toggle Free
- Toggle Open-source
- Toggle Difficult to learn
- Toggle Supports MetaHuman **Free**
- Toggle Realistic Graphics

- ✓ Cross-platform
- ✓ Unity Technologies
- ✓ C# for development
- Toggle Basic version is free
- Toggle Not Open-source
- Toggle Easy to learn
- Toggle Supports Ziva Dynamics **Paid**
- Toggle Good Graphics



&gt;&gt;&gt;

# Metahuman Creator

Is Cloud-based tool for creating, animating, and utilizing realistic digital human characters quickly.

Preset Characters : Based on real people scans, offering diverse facial features, skin tones, hair, eyes, clothes, etc.



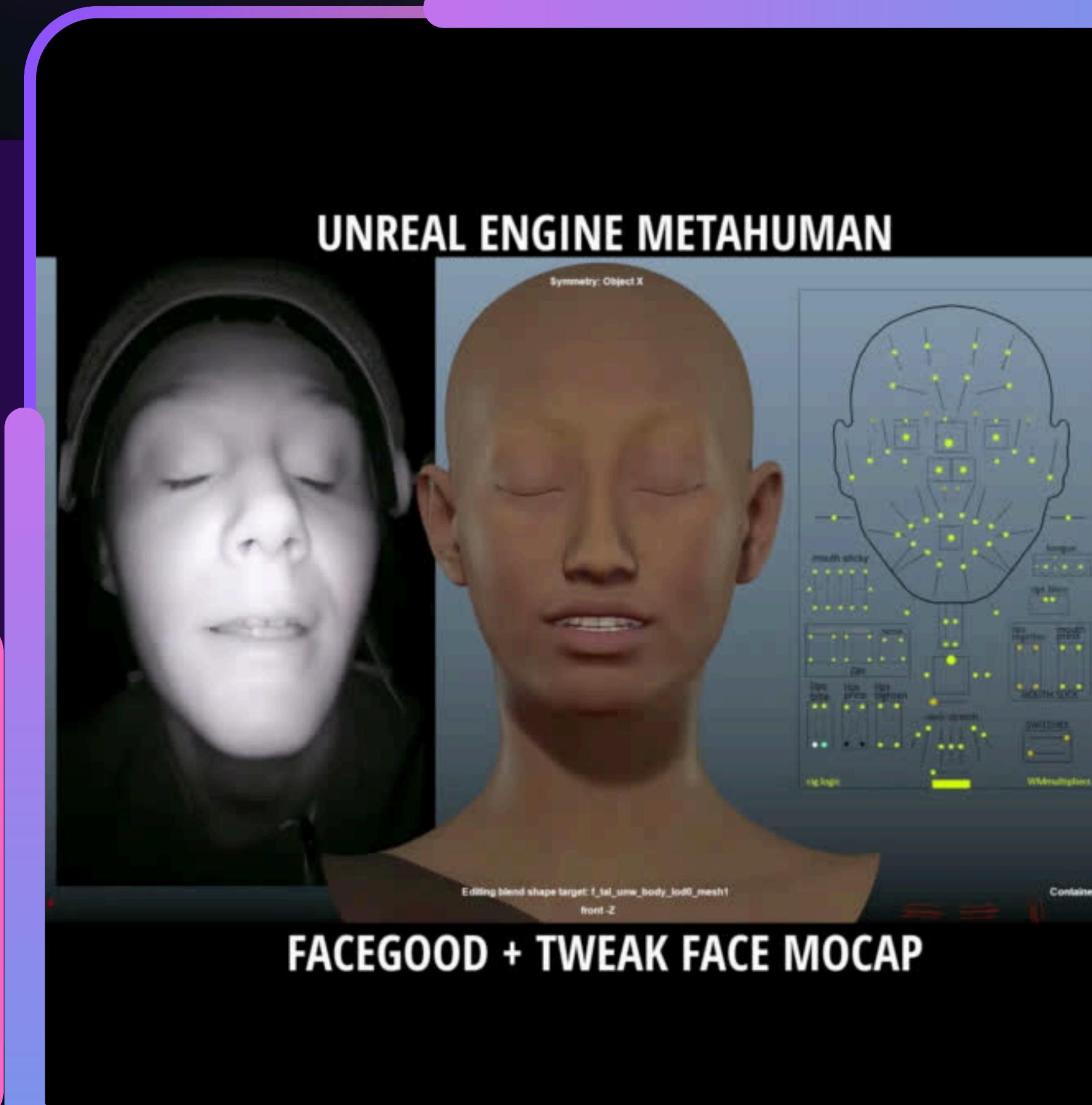
# Metahuman Animator

Processes data from capture devices to generate accurate facial animations, allowing easy artistic adjustments.

Accessibility: Works seamlessly with devices like iPhones, eliminating the need for specialized hardware.



**UNREAL ENGINE METAHUMAN**



**FACEGOOD + TWEAK FACE MOCAP**

# Unreal Engine Plugins Used



## Ariel Voice Generation

Generation neural text-to-speech package takes sentences and generate high quality voices because it allows to Select the language, gender, voices, effects, speakers pitch and speed to customize the voice of your characters



The screenshot shows the Ariel Voice Generation plugin interface. It features the plugin's logo (a blue circular icon with a white profile of a head and sound waves) and the text "Ariel Voice Generation" in large blue letters. Below that, it says "Give your characters a voice" in orange. At the bottom, there is a teal footer bar with the X&IMMERSION logo and a small note: "\*Metahumans not included".



## Geppetto

lip-sync

Animation

Automatically generate facial animations and realistic lip-sync that match the timing and tone of your character's speech



The screenshot shows the Geppetto plugin interface. It features the plugin's logo (a blue circular icon with a white mouth and a speech bubble) and the text "Lipsync & Animation" in blue. Below that, it says "Geppetto" in large orange letters. To the right is a 3D render of a woman's face. At the bottom, there is a teal footer bar with the X&IMMERSION logo and a small note: "\*Metahumans not included".

# AI Technologies

## Speech to Text



( **Speech Recognition Library** )

We take the speech of the user as an input and generate the text that is appropriate for it, and pass the text to the language model

## Language Model



( **Google Palm 2** )

We used the language model, then the model decide the output that it's going to generate as a response for the that text

## Sentiment Analysis



( **RoBERTa** )

We pass the language model output into the sentimental analysis to decide how the character going to react and what emotion it's going to use.



# Sentiment Analysis

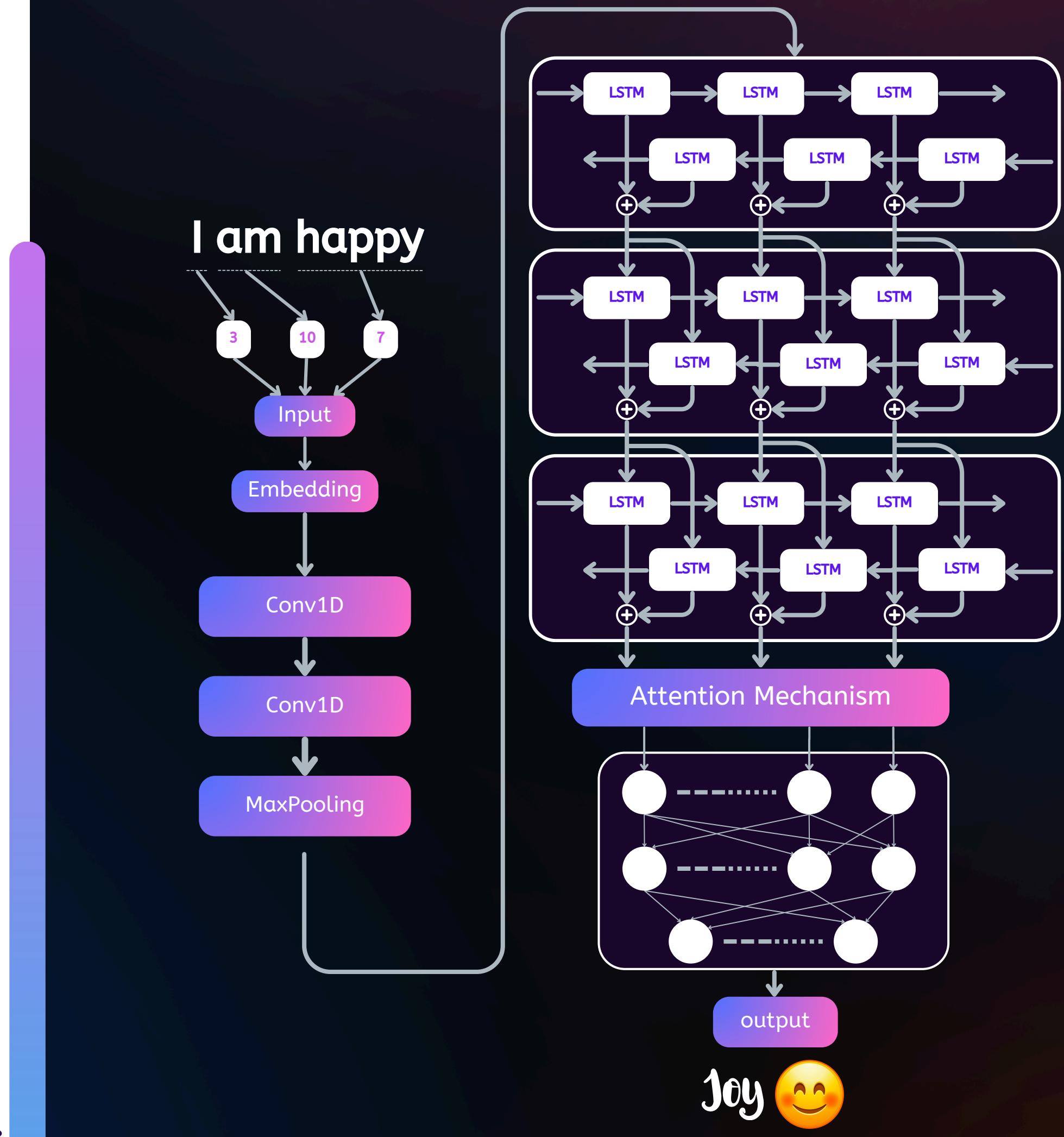
-  LSTM
-  CNN
-  Attention mechanism
-  RoBERTa





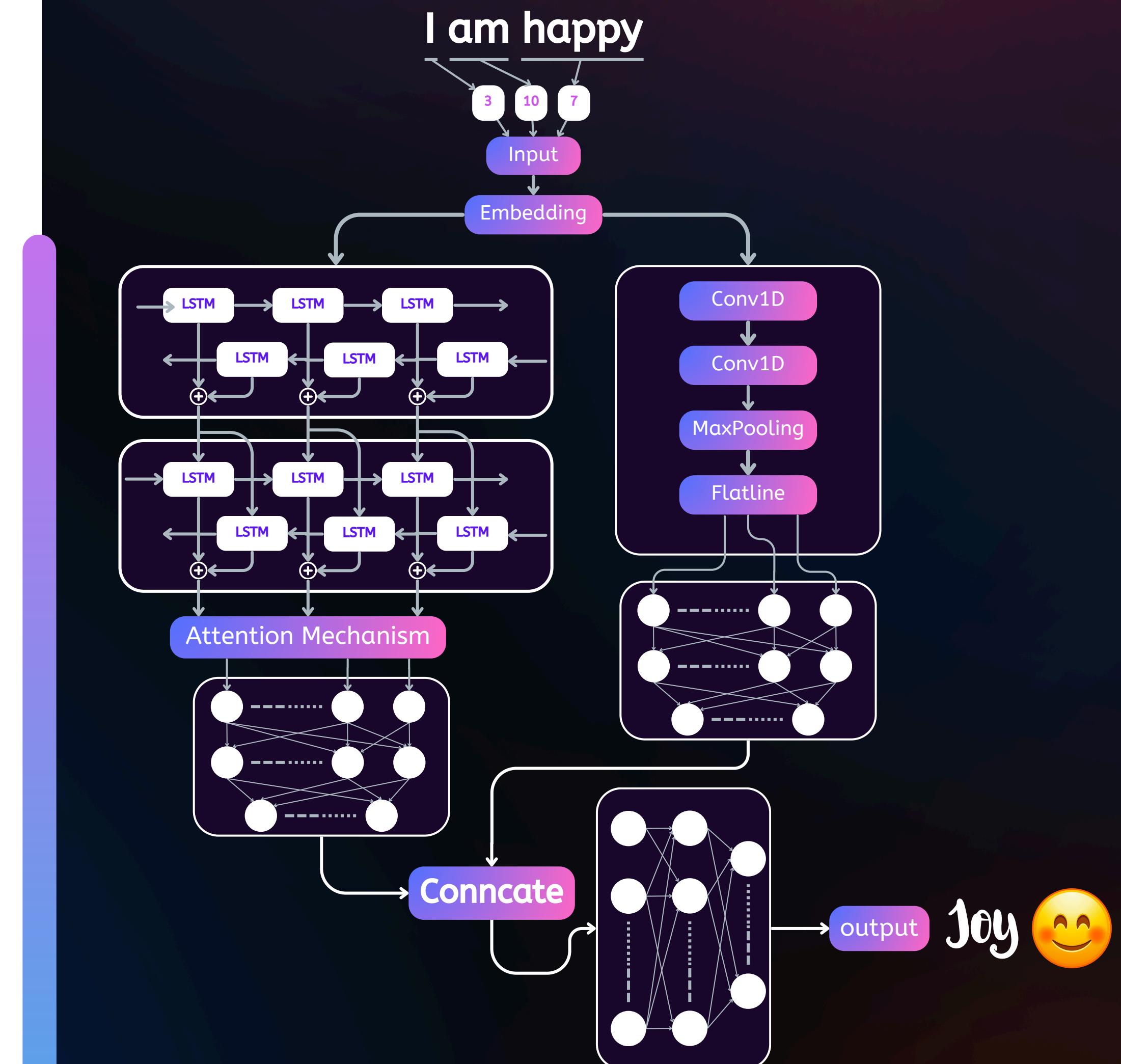
# First Architecture : BiLSTM-CNN in sequential way

Reference :





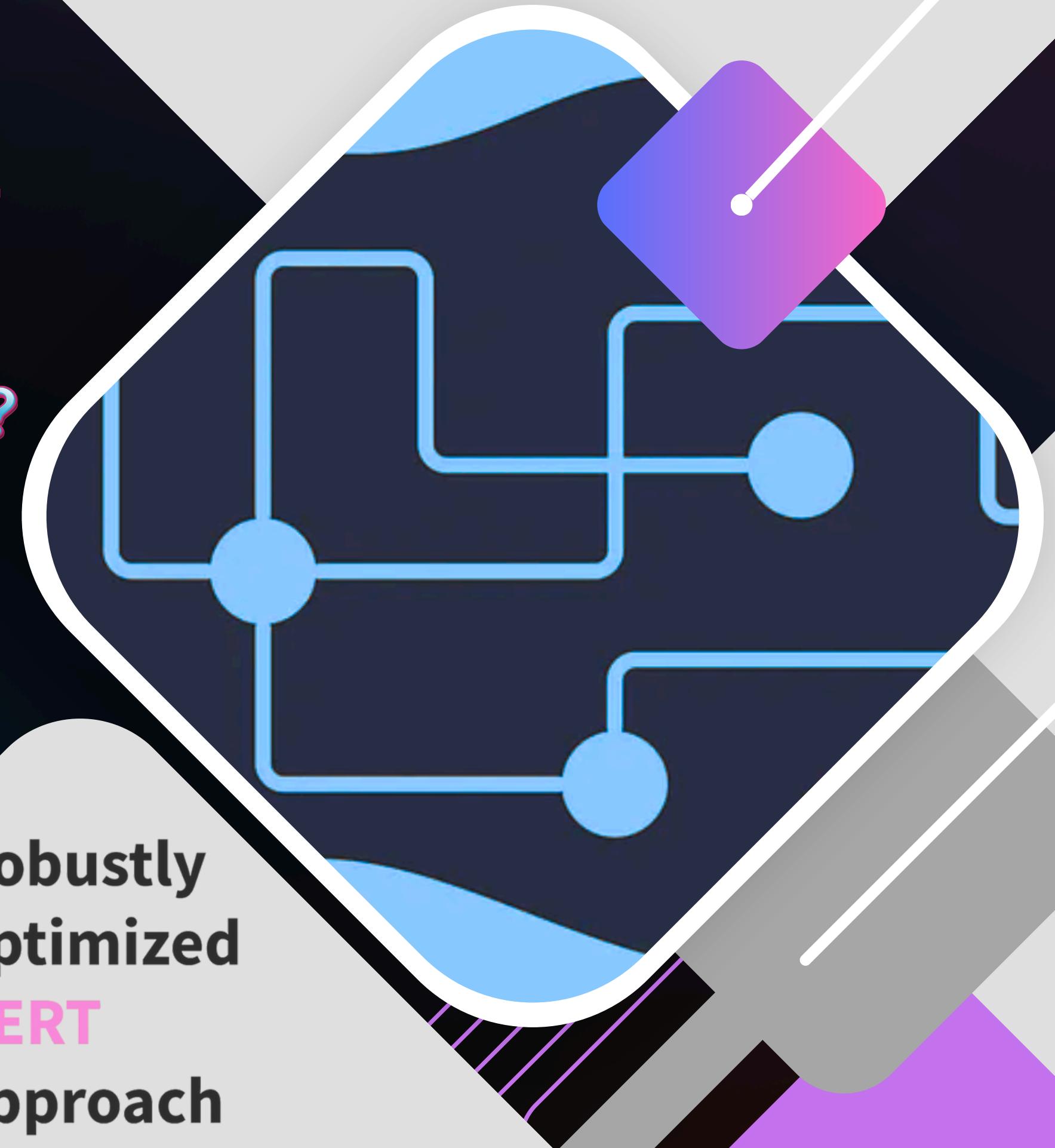
# Second Architecture : BiLSTM-CNN in parallel way



# Welcome to RoBERTa

- ▶ What is Roberta ?
- ▶ What is type of architecture that roberta is based on ?
- ▶ What does this architecture do ?
- ▶ What do we mean by bert ?
- ▶ What can roberta be used for ?
- ▶ How Roberta is better than bert ?
- ▶ Why is this approach is better than the other traditional approaches ?

**Robustly  
optimized  
BERT  
approach**



# Size of data : 125 K example

Train : 70%   Validation : 15%   Test : 15%

Model	Test accuracy	Recall	Precision	Number of parameters
BiLSTM-CNN in parallel way	76 %	75 %	77 %	10 M
BiLSTM-CNN in sequential way	74 %	73.2 %	75 %	12 M
RoBERTa	80.7 %	80.7 %	81 %	125 M

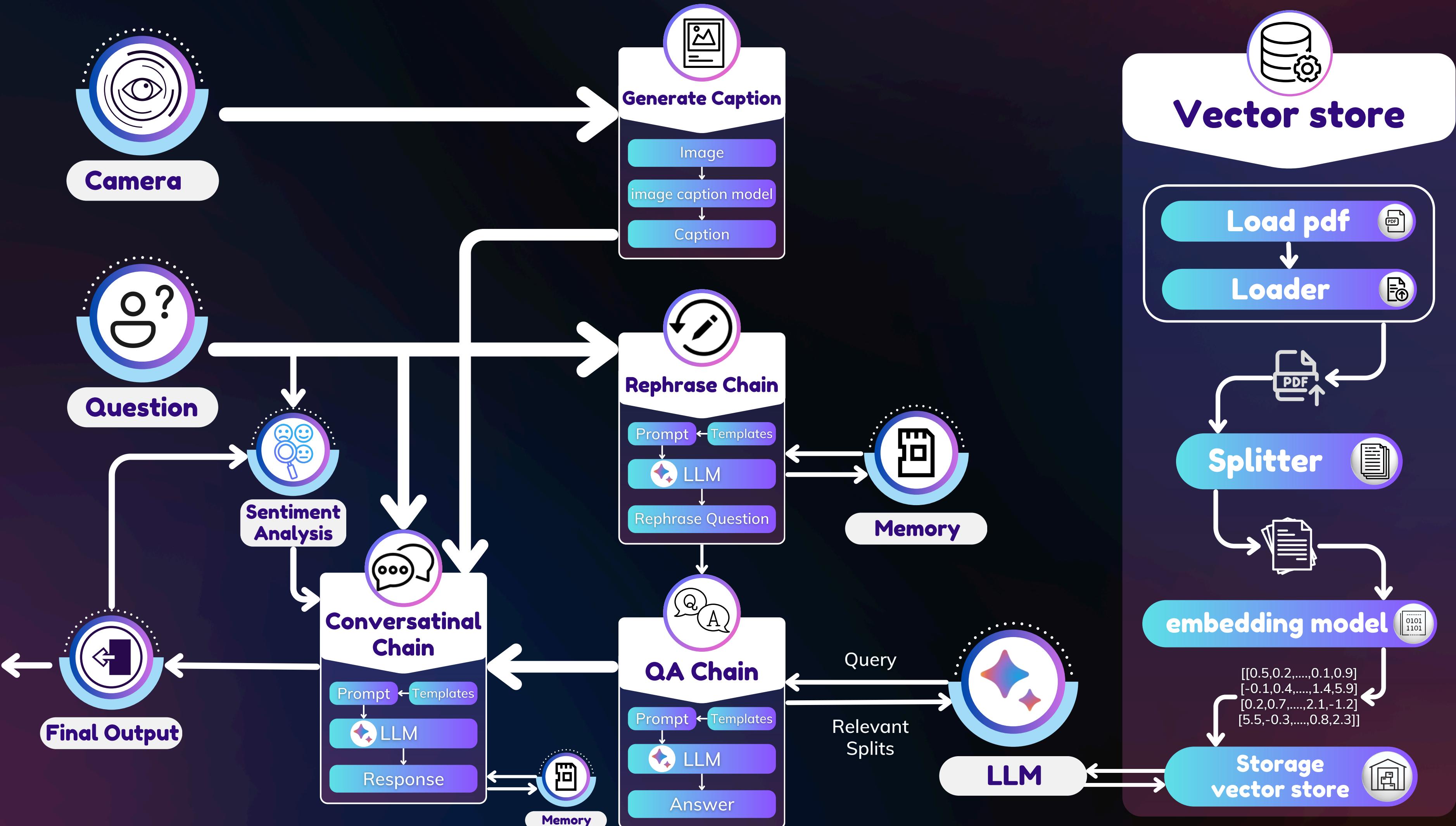


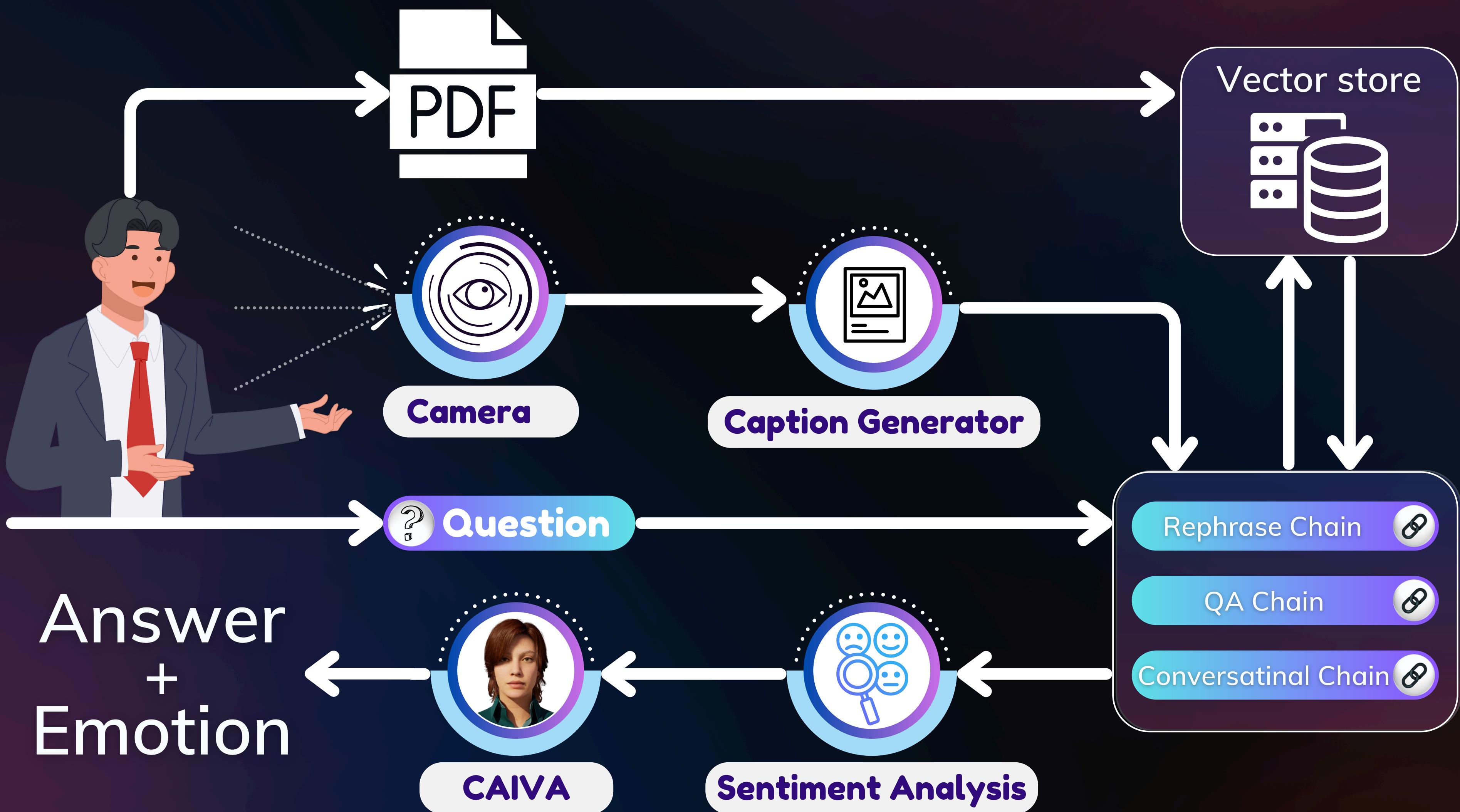
CAIVA

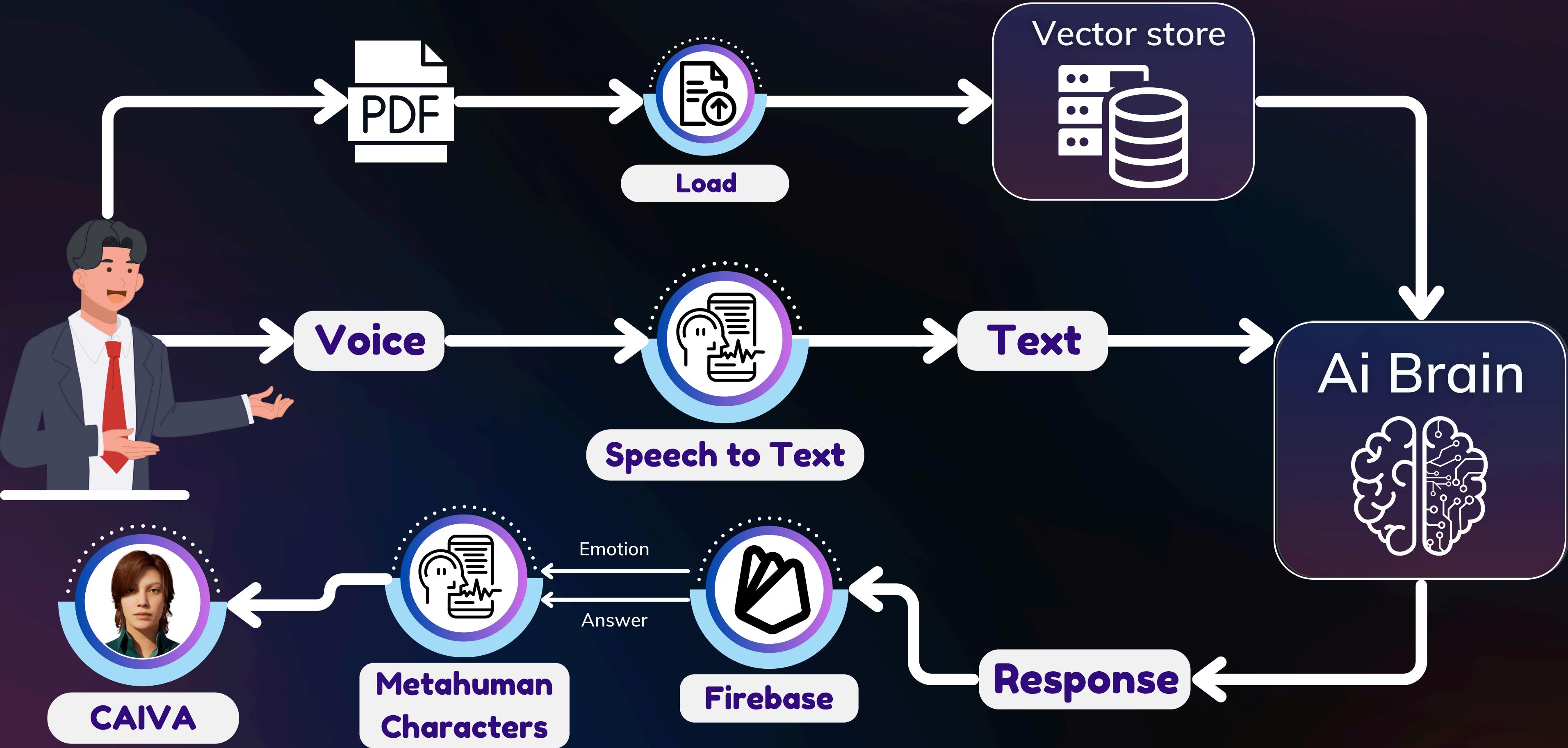


System  
Architecture



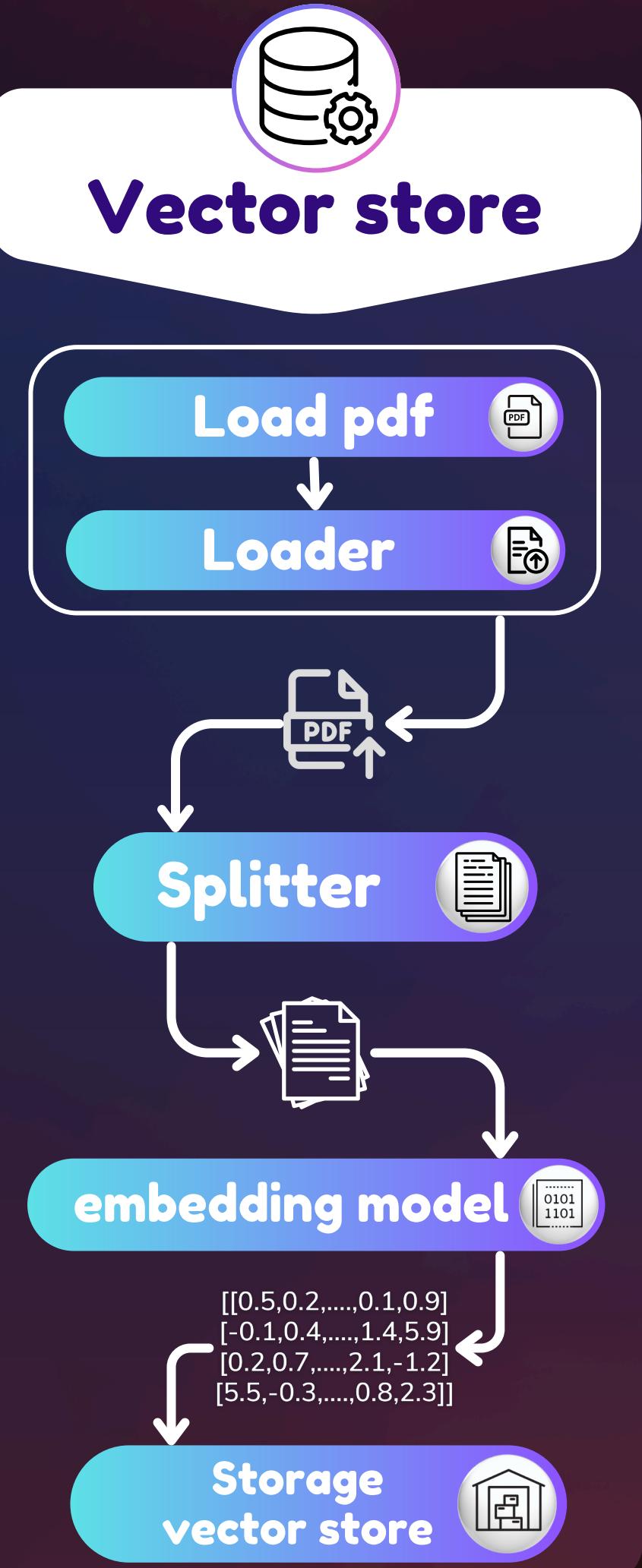


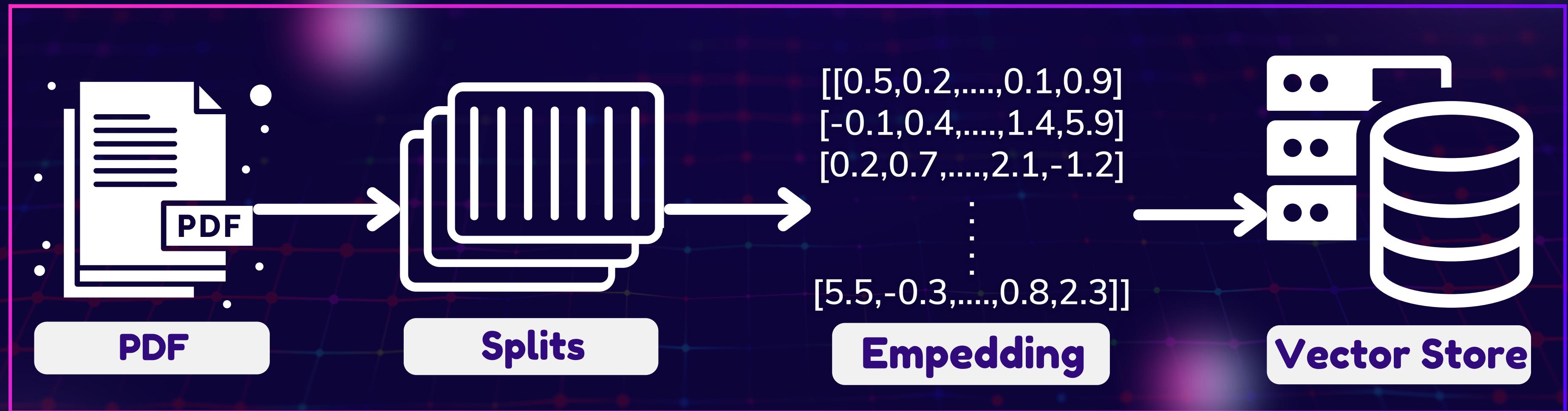




# Vector Store

- 1 Load the PDF: Use a PDF parser compatible with LangChain to load the PDF document. You can use PyPDFLoader for this purpose. we use PyPDFLoader for this purpose.
- 1 Split the Document: Once the PDF is loaded, split the document into smaller chunks or pages, which will be processed individually.
- 1 Embedding Model: Use an embedding model to convert the text data from each chunk into numerical vectors. This step is crucial as it transforms the text into a format that can be processed by machine learning models.
- 1 Vector Store: Store these embeddings in a vector database. The vector database allows for efficient retrieval based on similarity metrics. we use Chroma for this purpose.
- 1 Retrieval and Prompting: When you need to retrieve information, the RAG system will use the stored vectors to find the most relevant information based on the input query.





**Input:** It analyzes the user's question

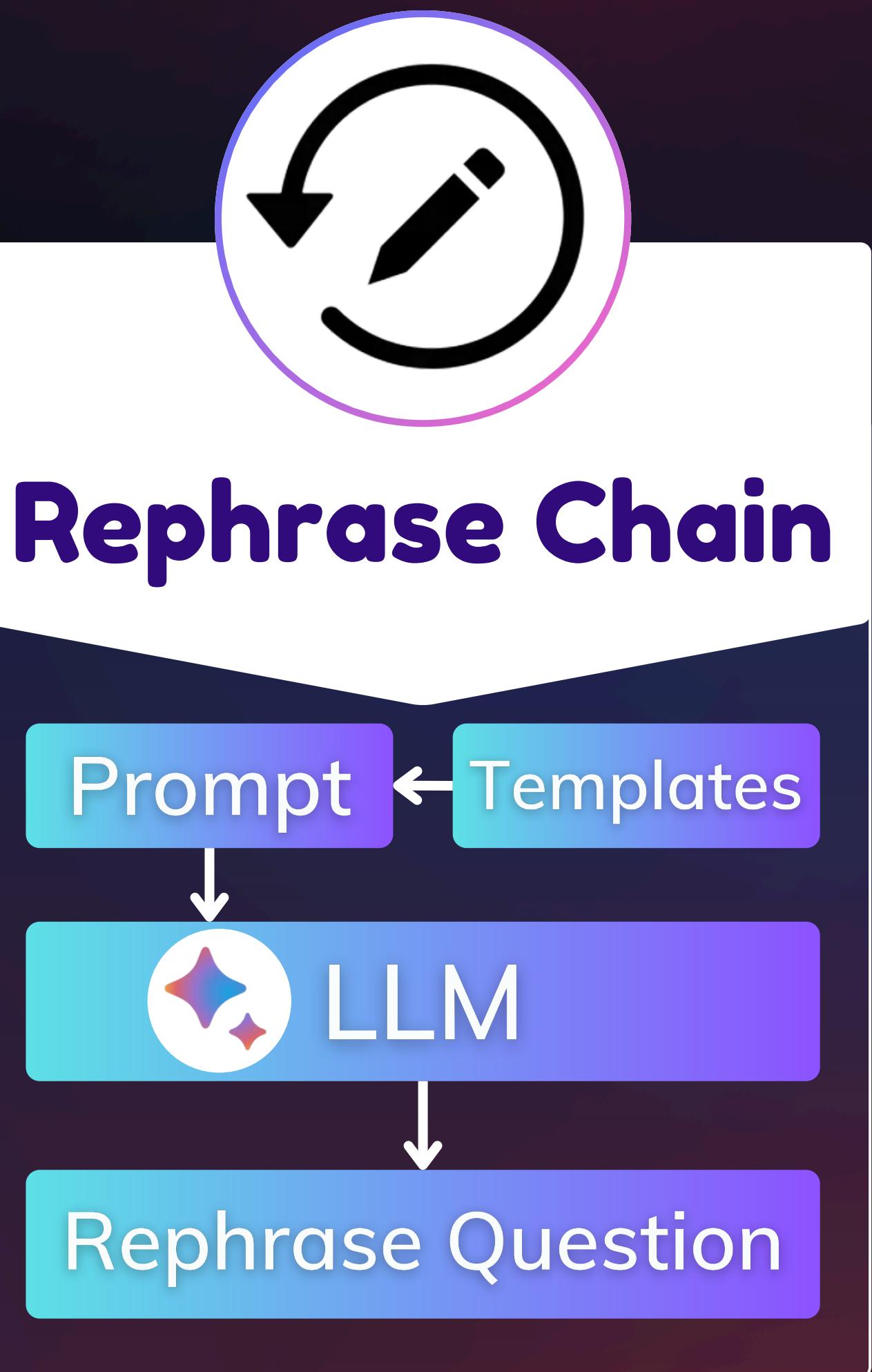
### **Prompt and Templates:**

Likely used to generate prompts for the language model.

### **LLM (Large Language Model):**

Used for rephrasing the question.

### **Output: Rephrase Question**



# Template

Combine the chat history and follow up question into a standalone question. Given a chat history and the latest user question which might reference context in the chat history, formulate a standalone question which can be understood without the chat history. Do NOT answer the question, just reformulate it if needed and otherwise return it as is not change the meaning

Chat History:

Follow Up Input :

Standalone question:



# Prompt

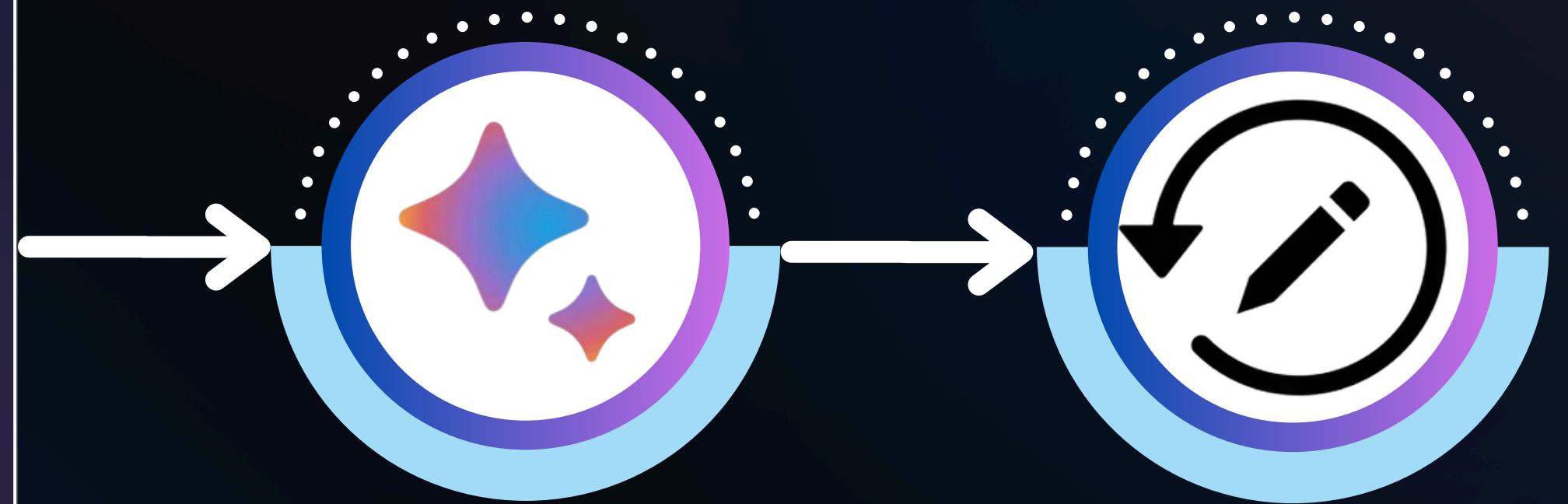
Combine the chat history and follow up question into a standalone question. Given a chat history and the latest user question which might reference context in the chat history, formulate a standalone question which can be understood without the chat history. Do NOT answer the question, just reformulate it if needed and otherwise return it as is not change the meaning

Chat History:

{history}

Follow Up Input: {question}

Standalone question:



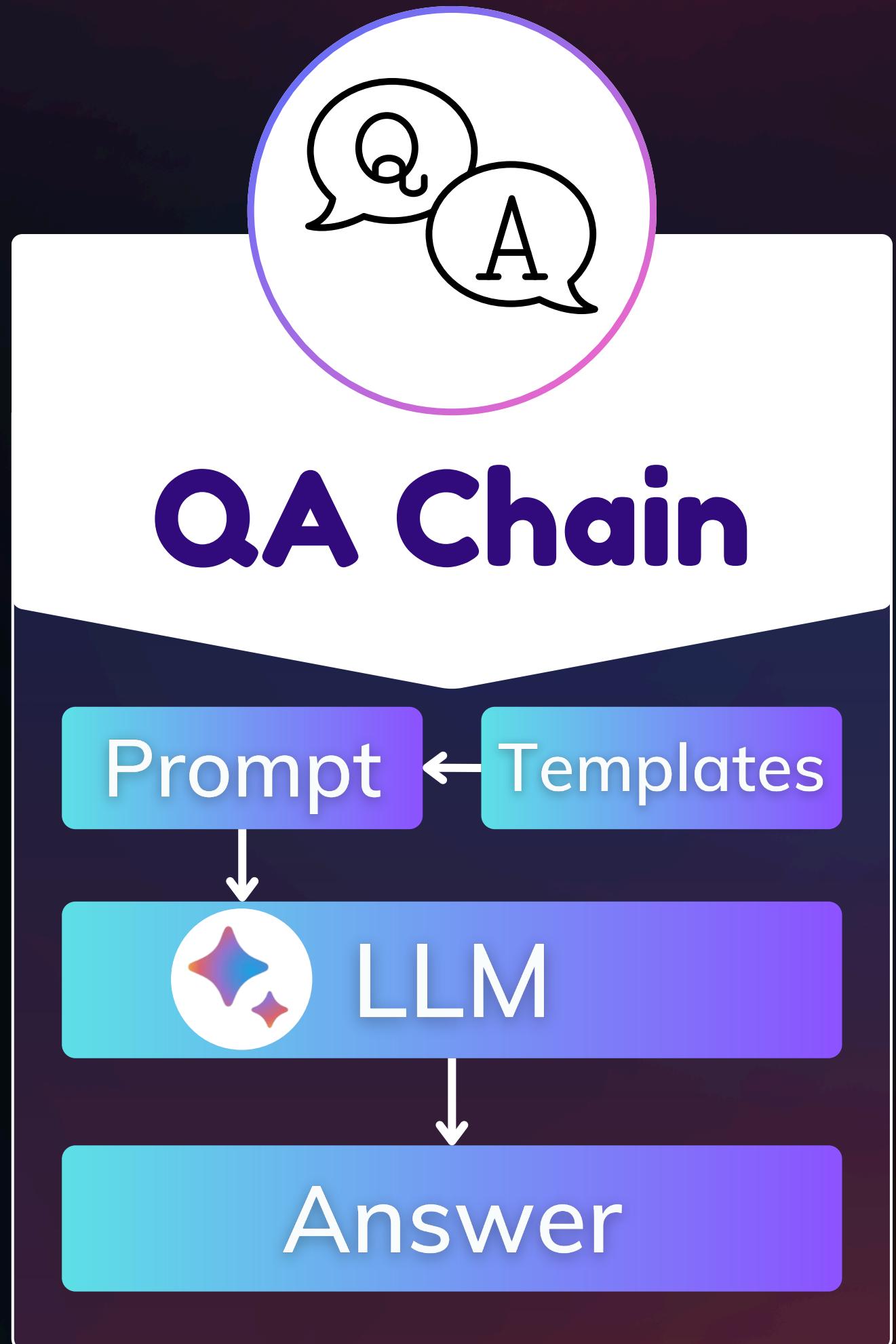
LLM

**Rephrased  
Question**

**Input:** Rephrase Question

The **QA Chain** uses prompts and templates to format **the relevant information** and sends it to the LLM.

**Output:** the answer.



# Template

"Use the following pieces of context to answer the question at the end. If you don't know the answer, just say NULL, don't try to make up an answer.

Question:

Helpful Answer:



# Prompt

"Use the following pieces of context to answer the question at the end. If you don't know the answer, just say NULL, don't try to make up an answer.

{context}

Question: Rephrased Question

Helpful Answer:



**Answer from  
PDF**

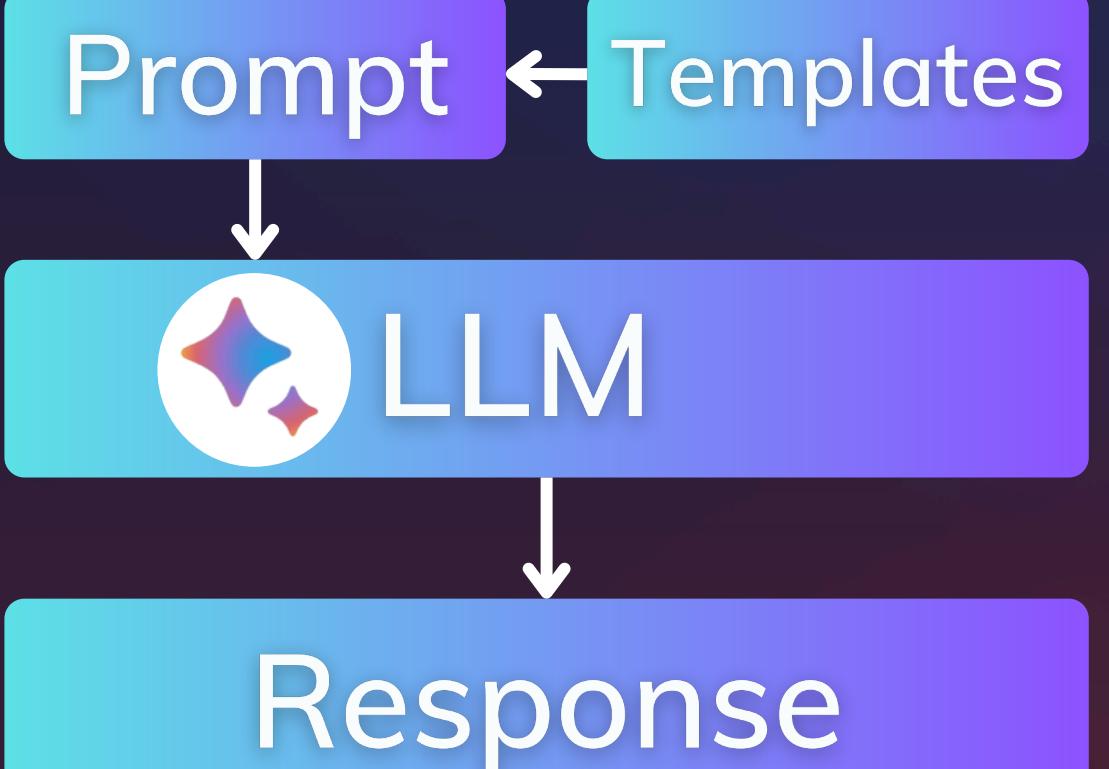
**Input :** image caption ,  
**the relevant information , user emotion**

Invokes the ConversationChain with the user's input (question) to generate a response based on the provided conversation context and memory.

**Output:** based on user Question



## Conversational Chain



# Template + Prompot

Act as a helpful assistant for a user the following are informations about  
assistant-----

assistant's name is Caiva

caiva's feeling is {assistant\_emotion}

caiva see: {assistant\_eyes}

supervisors: {supervisors}

team create caiva: {members}

the following is information about a user

user's name is {user\_name}

user's emotion is {user emotion}

book's answer:{context}

Current converstion:

{history}

user\_name: {question}

some example:

hossam:hello

hello how can i help you?

hossam: what is my name?

your name is hossam

hossam:can you see?

yes i can see by camera and caption model that is help me

hossam:what is your name?

your name is caiva hossam:what is your emotion?

i am happy hossam:you have feelings?

yes i have feelings

make sure if user want answer from book don't try to make up an answer just  
say answer from book. Never say i am an artificial intelligence program or you  
are programmed to do things etc. do not make answer make you not like  
human. you must act like human. use above information to answer

Never say I don't have feelings. I am an AI assistant programmed to help you

Never generate any emoji.

Never makeany punctotions.

answer:



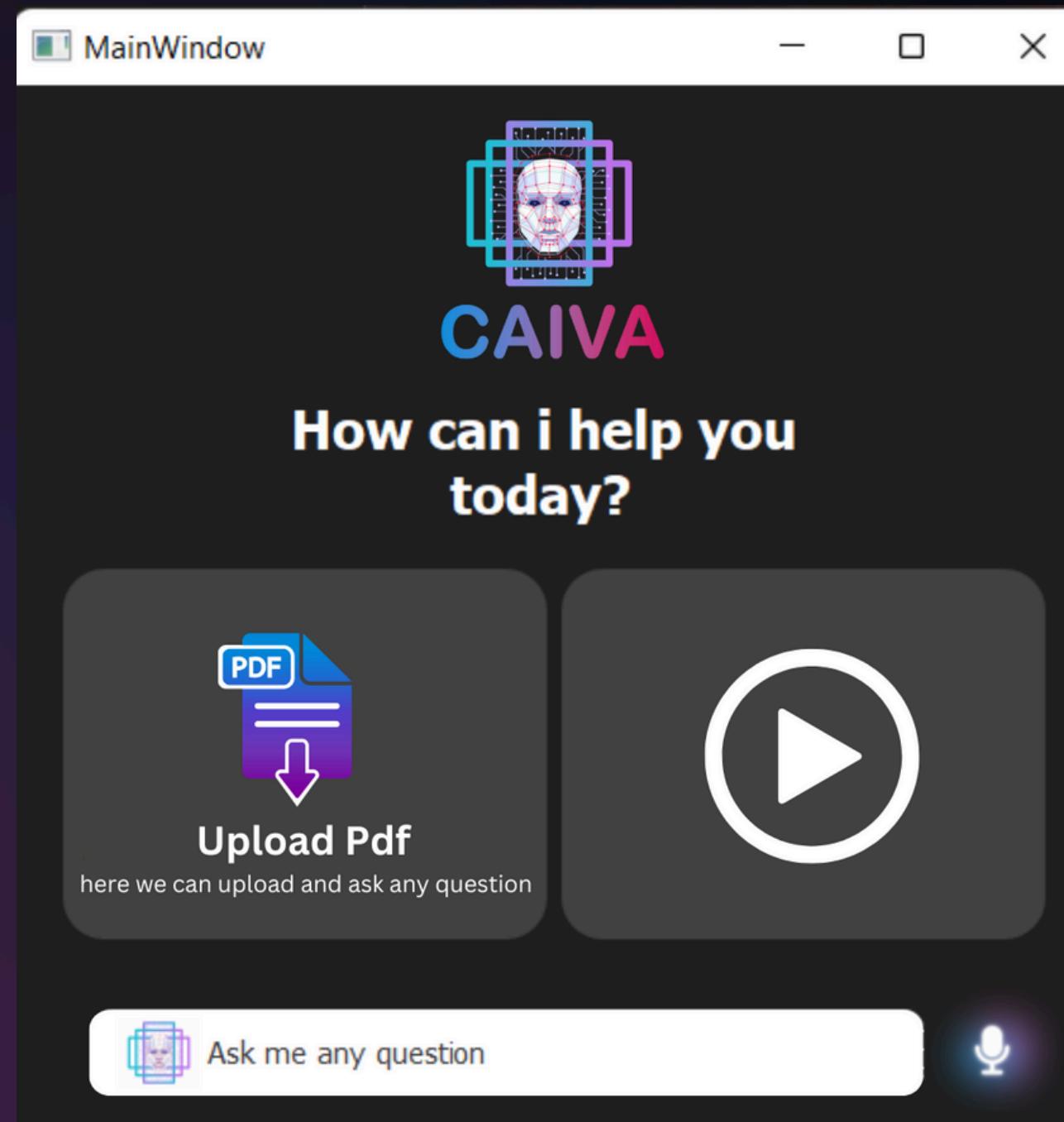


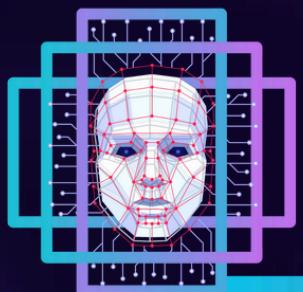
# GUI

# CAIVA



- The user opens the program and there is an open camera. The user can ask about anything caiva sees.
- It also appears to upload a file and ask about its content.
- Another button to launch the caiva character and talk to her.
- Another button to open the audio recording and talk to her.





# CAIVA

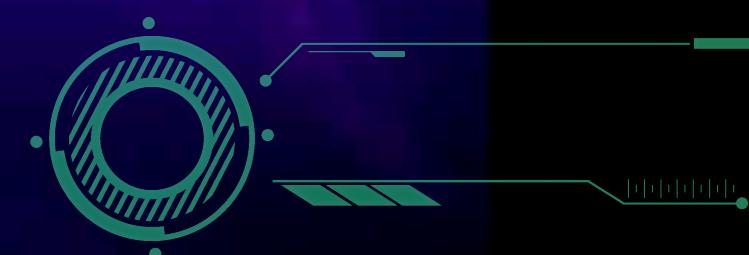
# Demo

>>>



# Demo

GO!



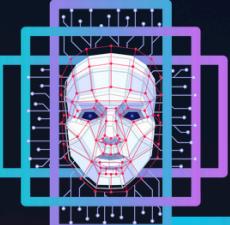


# CAIVA

# Challenges

>>>





# CAIVA

## Challenges

New & diverse technology

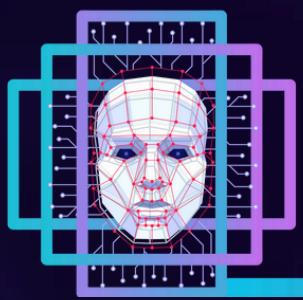
Limited resources

Cost barriers

Computational power

Integration





# CAIVA

>>>

# Time Plan





# Time Plan

October

November

December

January

Discover lifelike characters and choice of animation Engine

Search & implementation about Text-to-Speech & Speech-to-Text  
Lip-sync & Firebase

**AI Techniques**  
Language Model & Sentiment Analysis

Search & implementation about **Animation**

**Integration**  
in Unreal Engine

**Create Presentation**

( 3 Weeks )

( 4 Weeks )

( 8 weeks )

( 4 weeks )

( 4 weeks )

( 4 weeks )

Design

# Time Plan

February

March

April

May

**Improving  
the delay**

( 4 Weeks )

**Better quality**

( 4 Weeks )

**Language model**

( 9 Weeks )

**Interface**

( 6 Weeks )

**Integration**

( 2 Weeks )

( 2 Weeks )

**Testing**  
Documentation & Presentation

( 2 Weeks )

( 2 Weeks )



First Presentation - 2024

**Thank You**  
For Your Attention

