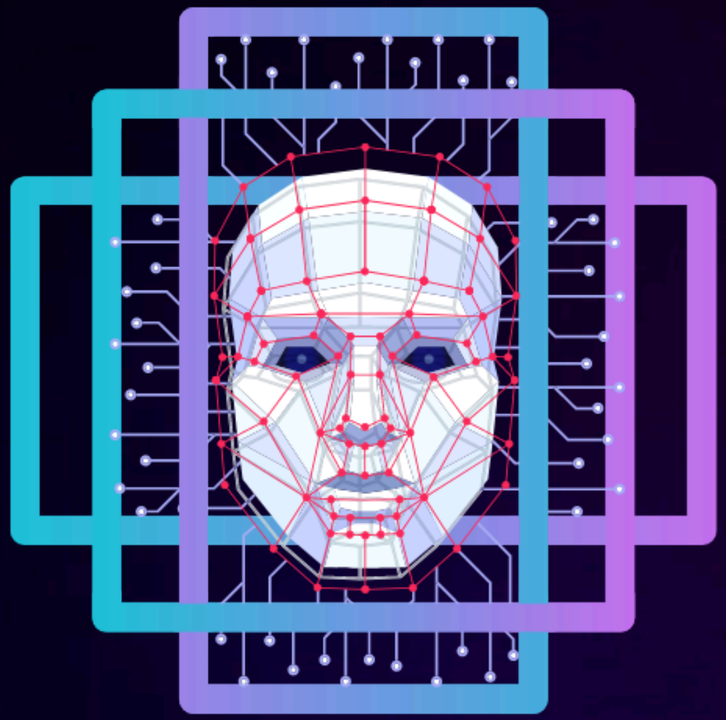


# CAIVA

EELU
THE EGYPTIAN E-LEARNING UNIVERSITY

2024

CAIVA

# 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**

EELU
THE EGYPTIAN E-LEARNING UNIVERSITY

# 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 **METHDOLOGY**
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 **SYSTEM ARCHITECTURE**
the conceptual model that defines the structure, behavior, and more views of a system

10 **GUI**
graphical user interface for CAIVA Application

11 **DEMO**
It's time to talk to our virtual character "CAIVA"

12 **CHALLENGES & TIME PLAN**
The challenges and difficulties we went through & Plan and manage project tasks efficiently.

CAIVA

# Introduction

## CAIVA Presentation 2024

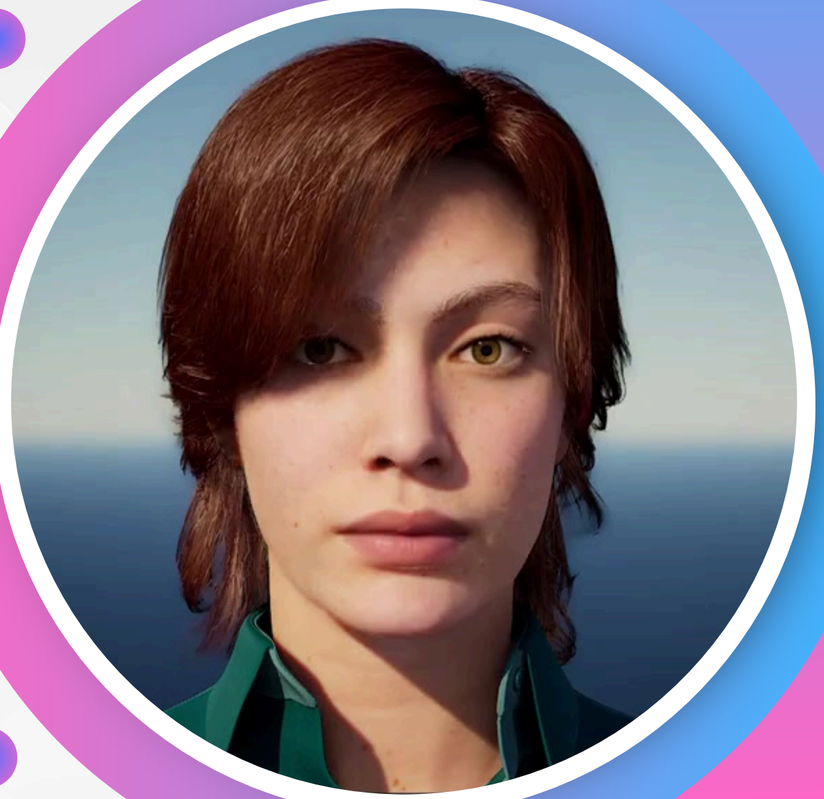
# CAIVA

## Conversational Artificial Intelligence Virtual Assistant

CAIVA

# Our Project

## CAIVA Presentation 2024



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.

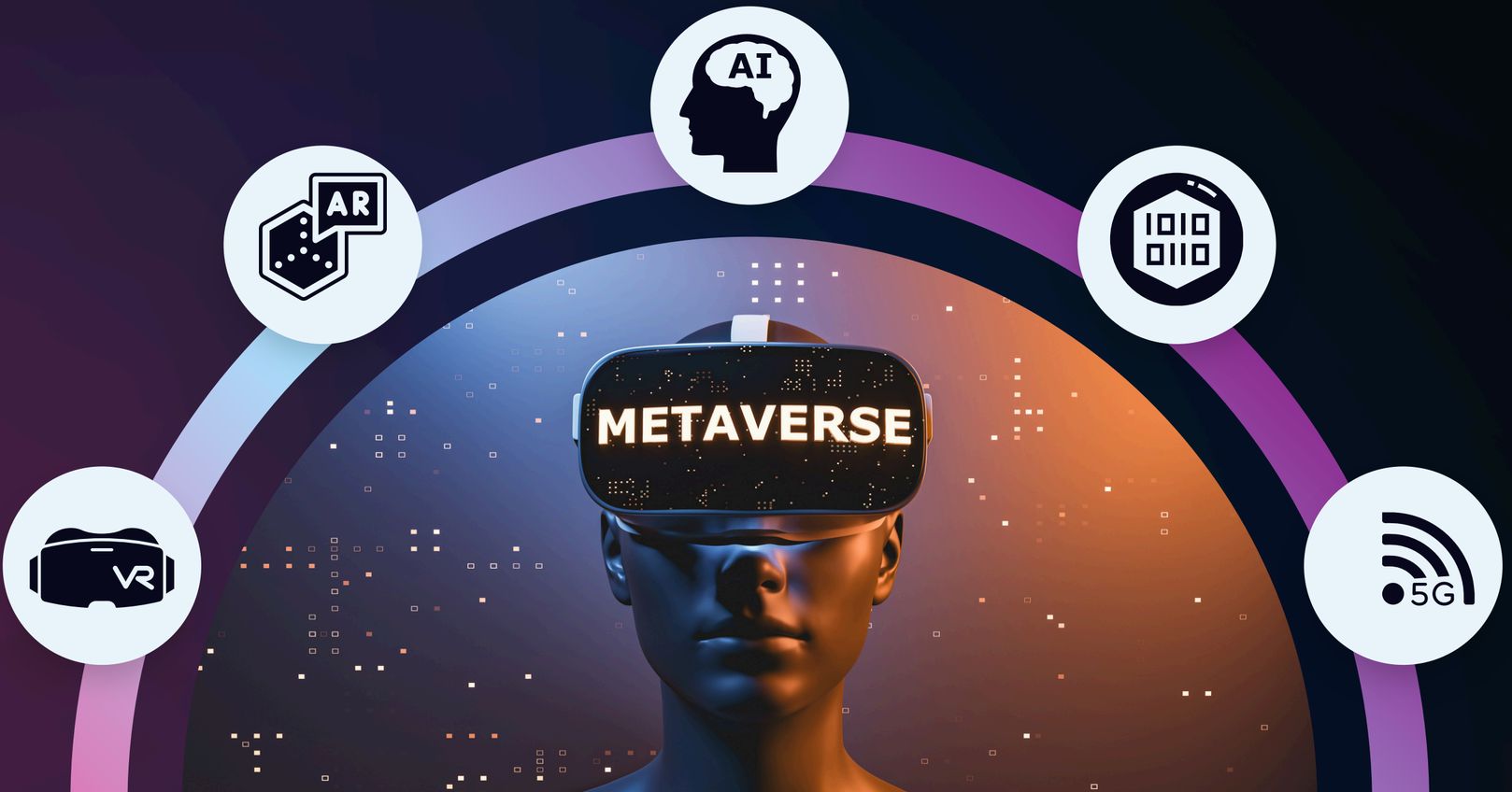
- Supported by sentimental animations, Caiva strives to create an authentic user experience, ensuring users feel a genuine connection with the characte

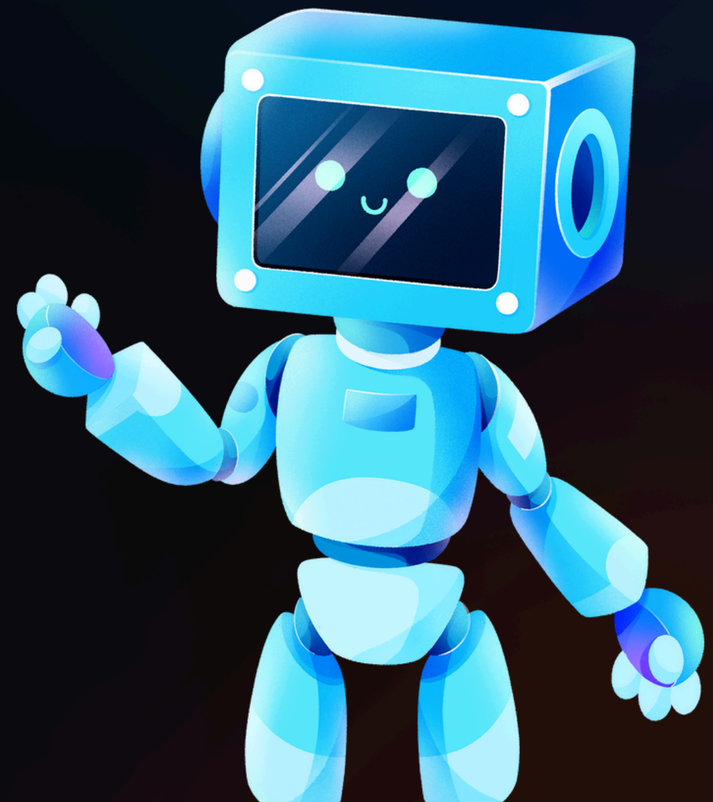
CAIVA

# Problem Statement

## CAIVA Presentation 2024

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.



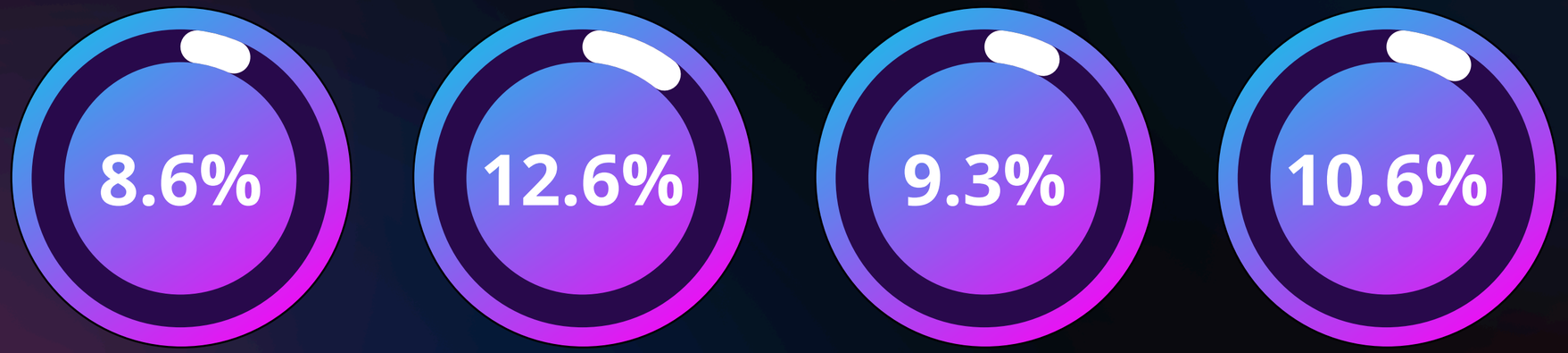



# Problem Statement

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

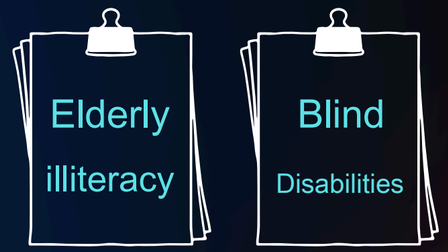
- **The Elderly**
- **Illiterate people**
- **Blind people**
- **Disabilities people**

# Research Highlights for **EGYPT** in 2023




**Elderly People** **illiteracy** **Blind People** **Disabilities**

**Search Reference :**



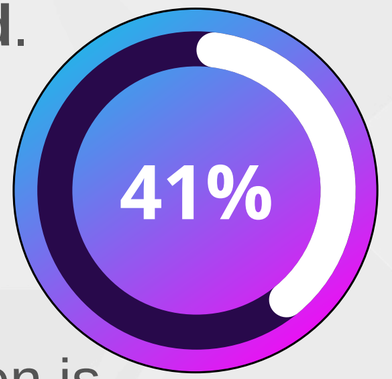




# Problem Statement

## 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.




**Search Reference :**

CAIVA

# Solution

## CAIVA Presentation 2024

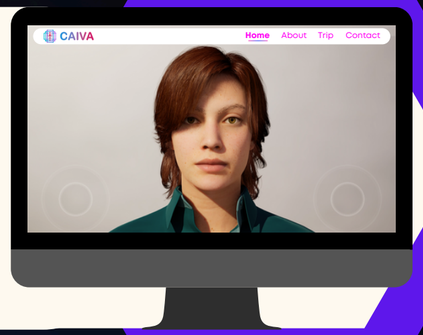
# Caiva is not just made for one use or one purpose

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

Development of virtual interface agent (Software agent) based on AI models and virtual reality that can be supported with all the required features that makes him adaptable in several applications.

1

## Websites




2

## Mobile Applications

3

## Smart Watches

4

## Car Touchscreen

# 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

# Differnce

## CAIVA Presentation 2024

# 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, ...

## 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 assistant with smart home control and device support.

## Astra Google Project

- **Developed by :** Google
- **Integration :** Available on Android devices, Google Home, and other platforms
- **Capabilities :** Voice-activated assistant with smart home control and device support.

## Chatgpt-4o

- **Developed by :** OpenAI
- **Integration :** Available via API and integrated into various applications and platforms
- **Capabilities :** Advanced natural language processing (NLP) for generating human-like text, answering questions, and assisting with a wide range of tasks.

## Caiva

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

CAIVA

# Key Features

## CAIVA Presentation 2024

# Key Features

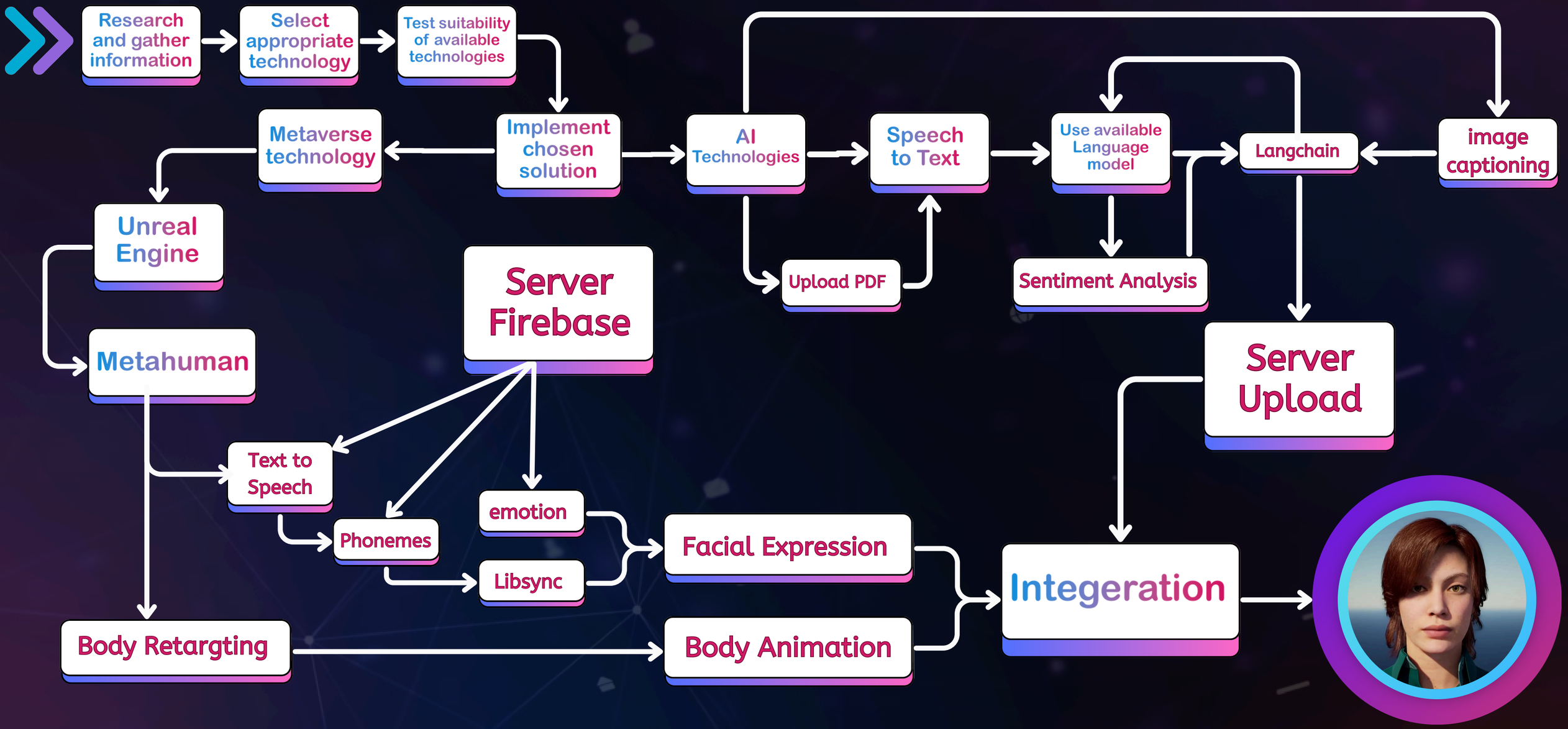
- Human-Centered Design
- Integrable in multi-Devices
- Companion Experience
- Emotional Intelligence
- Interactive Animations
- Continuous Improvement

CAIVA

# Methodology

## CAIVA Presentation 2024

# Methodology



CAIVA

# Technology Used

CAIVA Presentation 2024


- Unreal Engine
  - Geppetto
    - Mixamo — Animation Retargting
    - Geppetto Phonemes
    - Geppetto Emotion
  - Ariel Voice Generation
    - Text to Speech
- Metahuman Creator
- AI Technologies
  - Language Model — Gemini
  - Sentiment Analysis — Roberta
  - Speech to Text — Google Speach recongition
  - Langchain
  - Img captioning — Blip

# Why Unreal Engine ?

| Unreal Engine | | Unity |
|---|---|---|
| ✓ Cross-platform | | ✓ Cross-platform |
| ✓ Epic Games | | ✓ Unity Technologies |
| ✓ C++ for development | | ✓ C# for development |
| Free | | Basic version is free |
| Open-source | | Not Open-source |
| Difficult to learn | | Easy to learn |
| Supports MetaHuman Free | | Supports Ziva Dynamics Paid |
| Realistic Graphics | | Good Graphics |

# Metahuman Creator

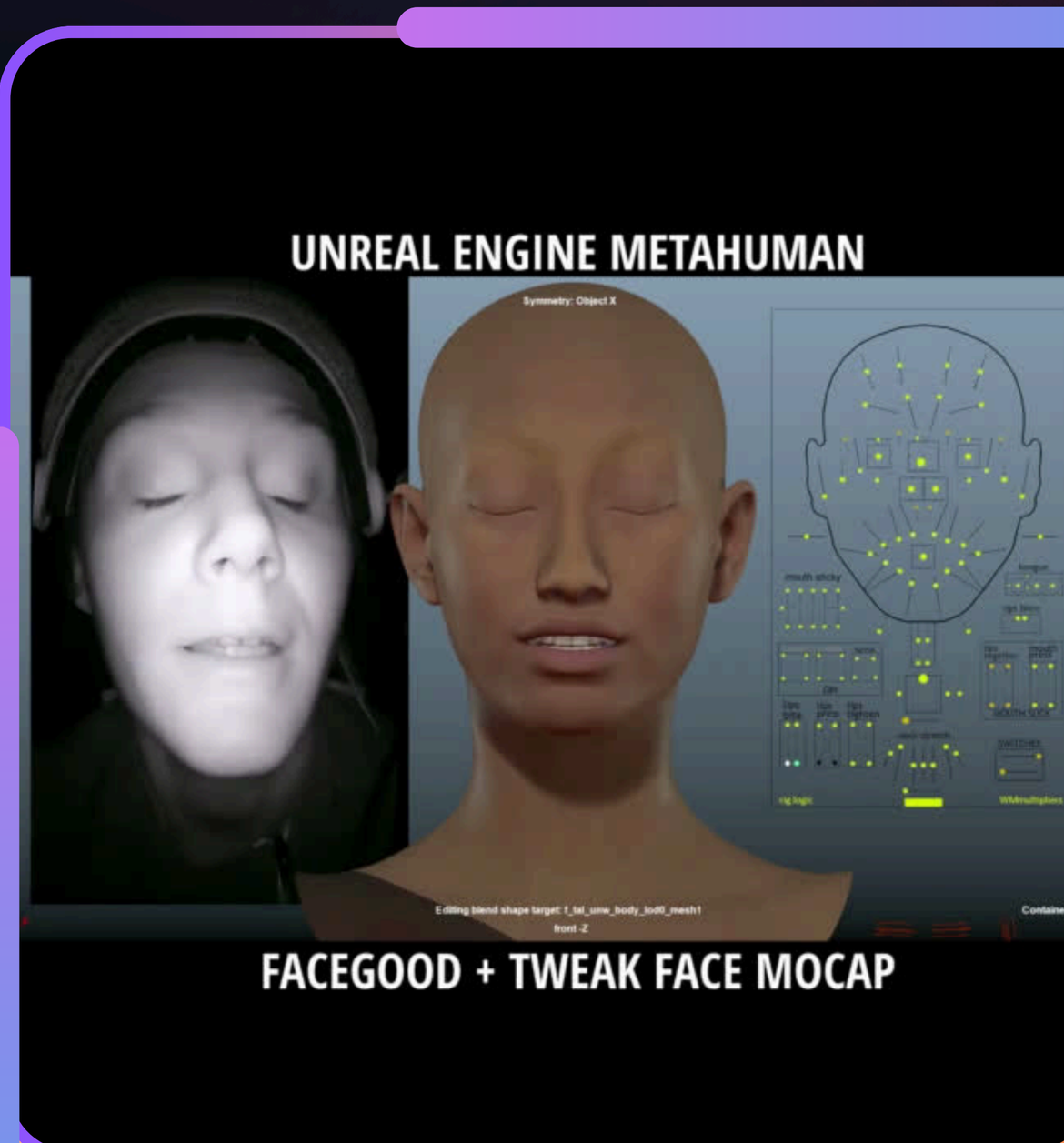
Is Cloud-based tool for creating, animating, and utilizing realistic digital human characters quickly.

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

## Plugins Used

### Ariel Voice Generation

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


lip-sync **Geppetto** Animation

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

# Benefits of Geppetto For Animation with Unreal Engine

## Facial Animation Precision

Geppetto Animation specializes in facial animation, offering precise control over facial expressions, lip-syncing, and emotion capture.

## Natural Facial Movements

It enables natural and lifelike facial movements by capturing detailed expressions and nuances, which are crucial for immersive storytelling and character interaction.

## Integration with Unreal Engine

Geppetto Animation integrates seamlessly with Unreal Engine, leveraging its powerful rendering capabilities and real-time environment to create compelling character animations.

## Ease of Use

The user-friendly interface and intuitive controls of Geppetto Animation simplify the process of creating and editing facial animations, reducing the learning curve for animators.

## Compatibility with MetaHumans

Geppetto Animation is compatible with Unreal Engine MetaHumans, allowing for the direct application of facial animation data to these high-fidelity character models.

# Why did we still use Geppetto over Live link ?

- Specialized Focus
- Realism and Precision
- User-Friendly Interface
- Integration with Unreal Engine
- Enhanced Creative Flexibility
- Preferred for Emotive Storytelling

# Animation Retargting Mixamo for Body Motion


Animation Retargeting is a feature that allows us to share animations between characters that use different Skeleton assets as long as they share a similar Bone Hierarchy and use a shared asset called a Rig to pass animation data from one Skeleton to the other, when you have differing proportions, you can get some unsightly results.

Once retargeting is applied to the characters, the differences in their proportions are taken out of the equation and the animation plays properly on each character.

MetaHumans have unique, highly detailed skeletons.

To maintain their realistic movements, special attention is required when retargeting animations.

# AI Technologies

## Speech to Text

### ( Speech Recognition )

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


### ( Gemini )

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.

## image captioning

### ( Blip )

We have used image captioning to combine visual and linguistic tasks using extensive pre-training and to enhance the application as image captioning, visual and question answering.

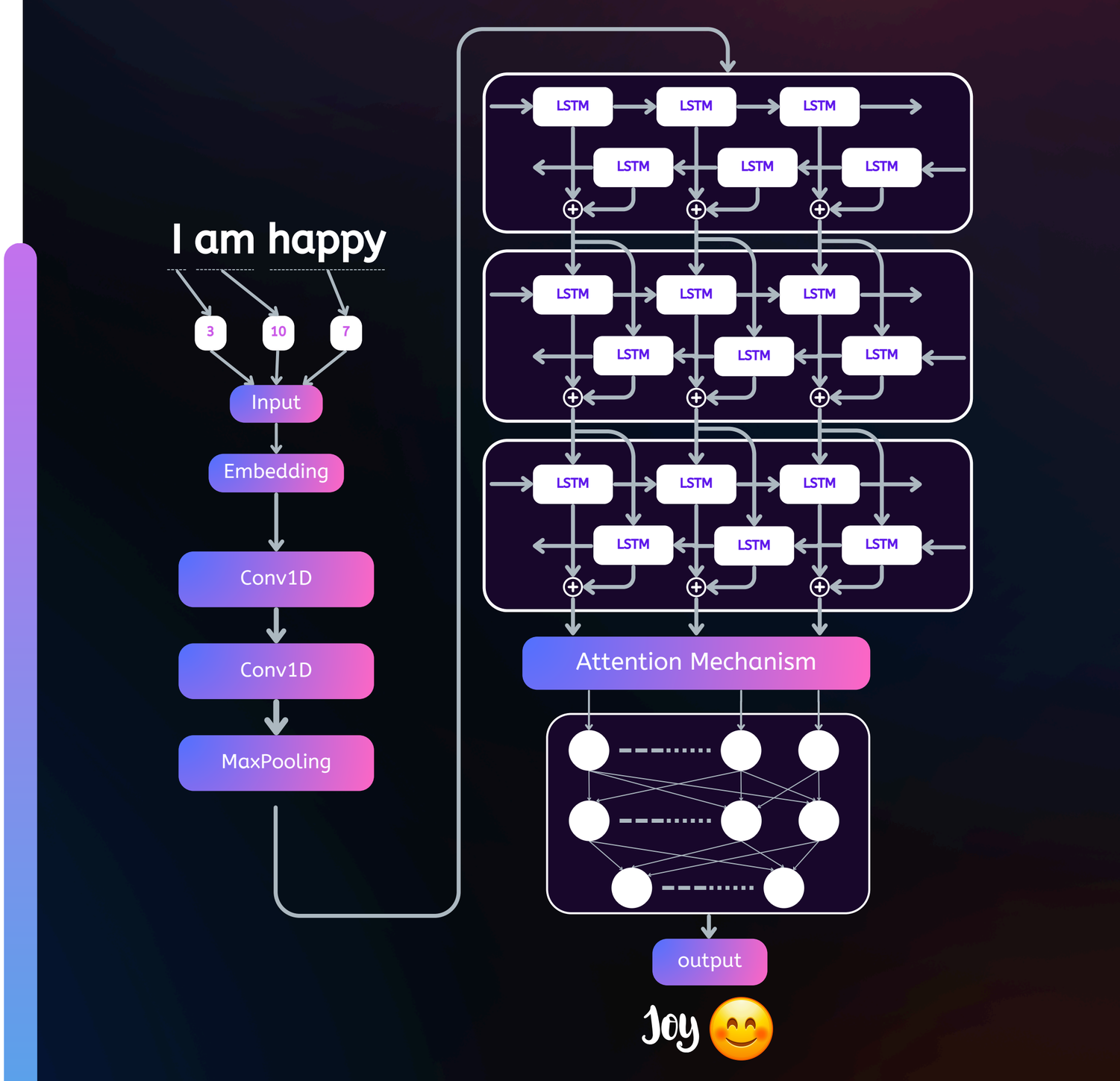
# Sentiment Analysis

- lstm
- CNN
- Attention mechanism
- RoBERTa

CAIVA

# First Architecture :

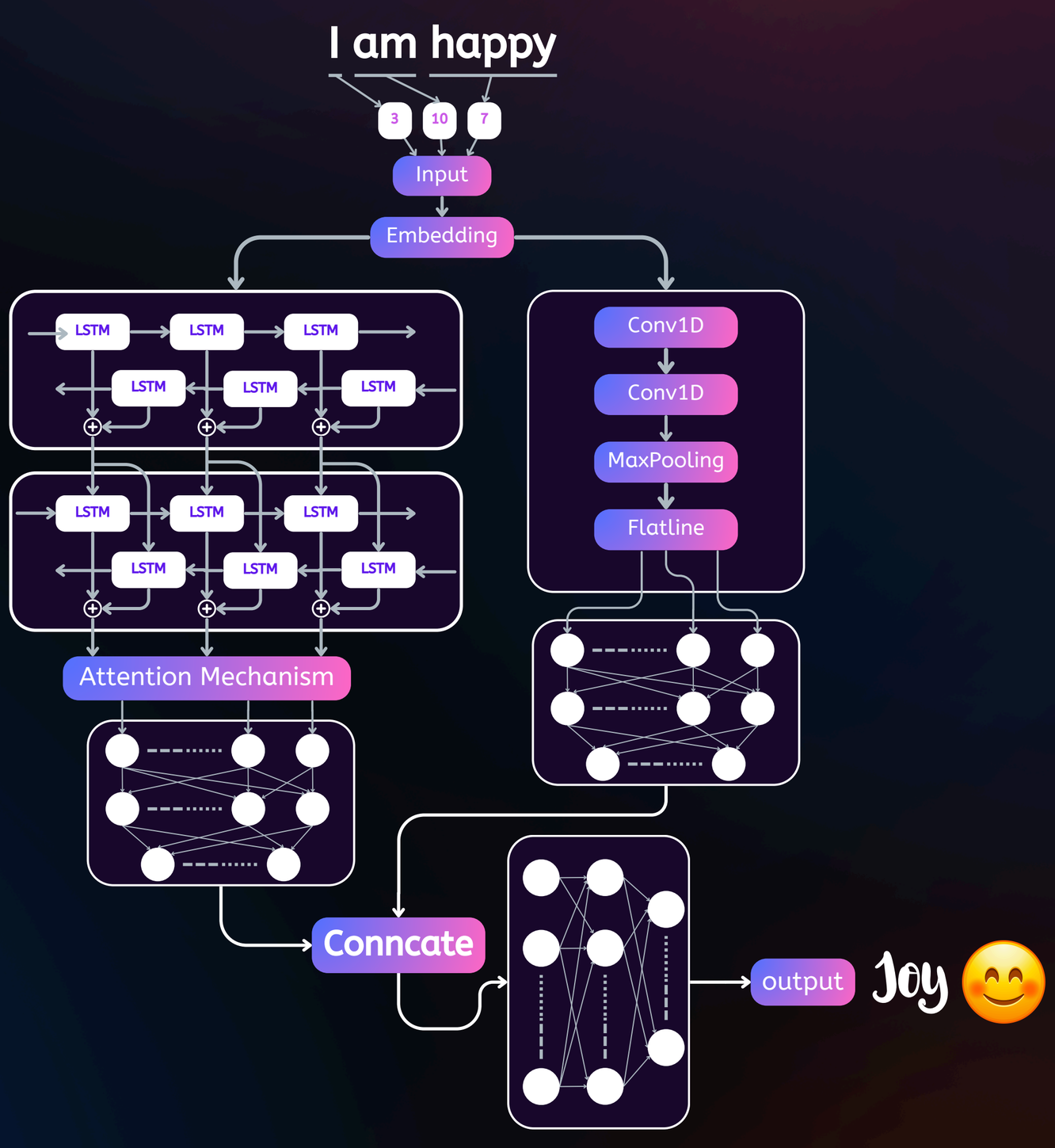
BiLSTM-CNN
in sequential way



CAIVA

# 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 can roberta be used for ?
- How Roberta is better than bert ?
- Why is this approach is better than the other traditional approaches ?

**R**obustly
**o**ptimized
**BERT**
**a**pproach

## Size of data : 125 K example

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

| Model | Test accueacy | 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 |

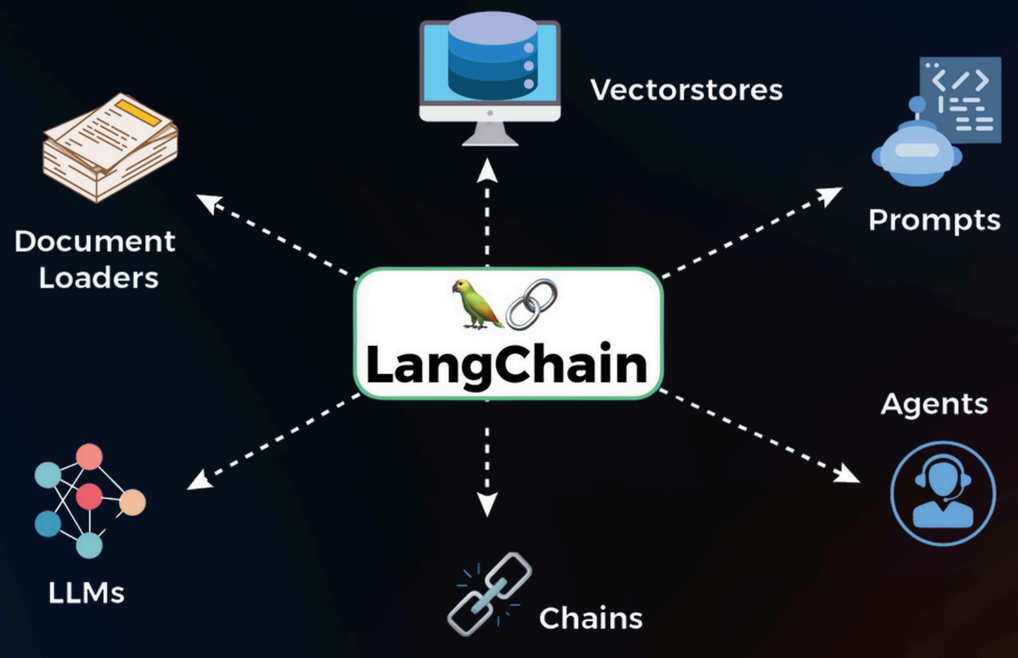
CAIVA

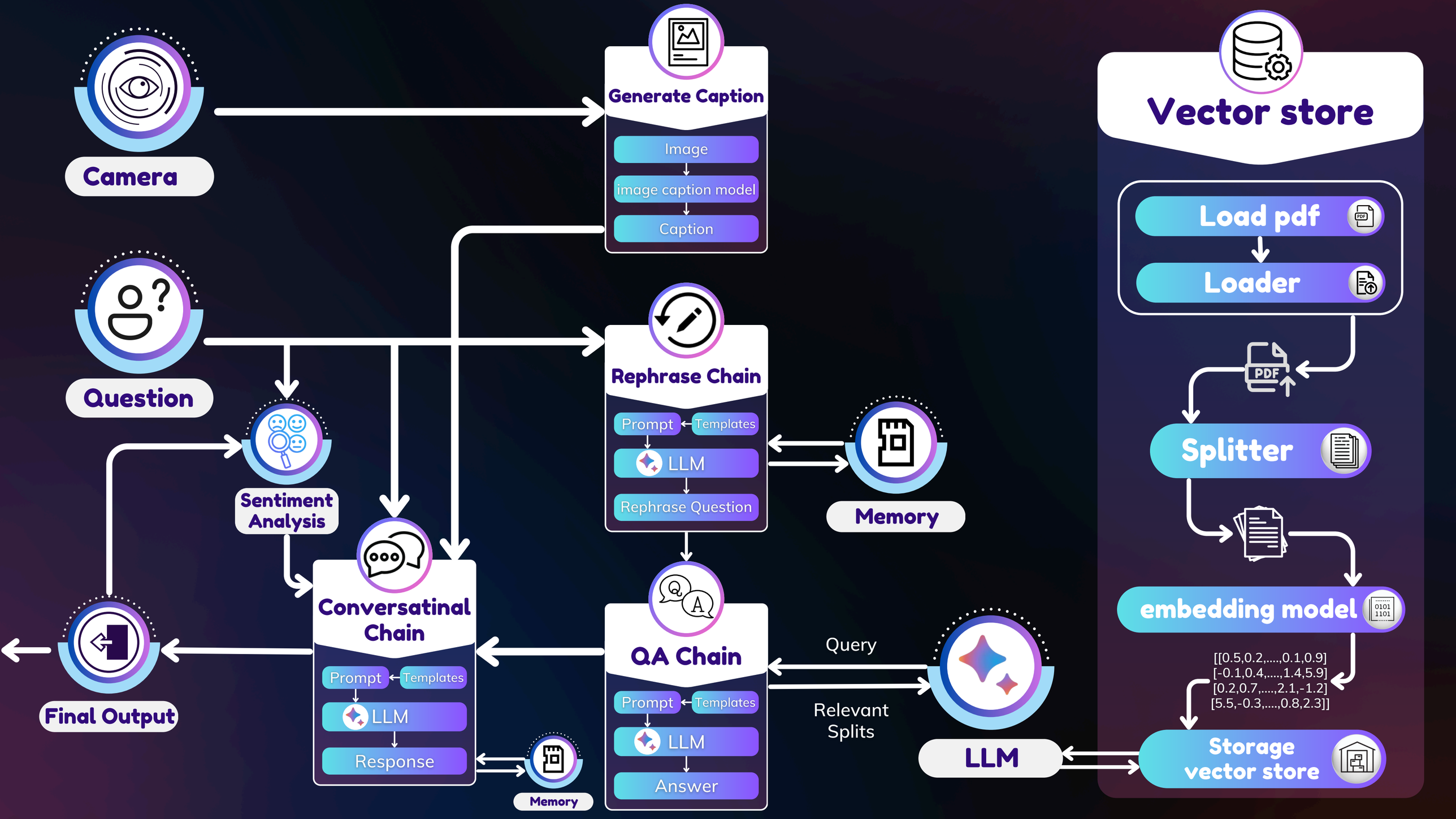
# System Architecture

## CAIVA Presentation 2024

# Langchain

- LangChain is a framework designed to simplify the creation of applications using large language models (LLMs).
- As a language model integration framework, which was launched in October 2022. and lead to features such as interactive chatbots.
- Langchain improves functionality, efficiency, scalability and innovation, providing significant benefits and competitive advantages in real-world projects.




Camera
Generate Caption
Image
image caption model
Caption
Question
Rephrase Chain
Prompt
Templates
LLM
Rephrase Question
Memory
Sentiment Analysis
Conversatinal Chain
Prompt
Templates
LLM
Response
Memory
QA Chain
Prompt
Templates
LLM
Answer
Query
Relevant Splits
LLM
Final Output
Vector store
Load pdf
Loader
PDF
Splitter
embedding model
[[0.5,0.2,....,0.1,0.9]
[-0.1,0.4,....,1.4,5.9]
[0.2,0.7,....,2.1,-1.2]
[5.5,-0.3,....,0.8,2.3]]
Storage vector store

# Vector Store

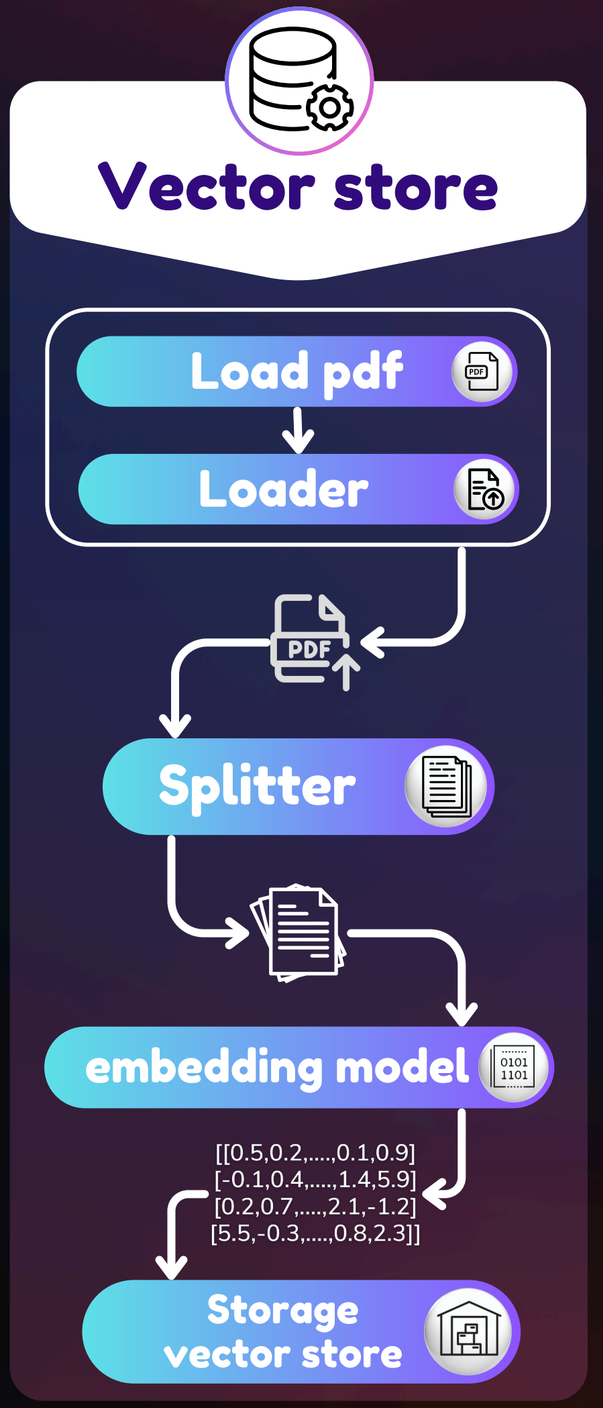
1. Load the PDF: Use a PDF parser compatible with LangChain to load the PDF document. You can use PyPDFLoader for this purpose. we use PyPDFLoader for this purpose.

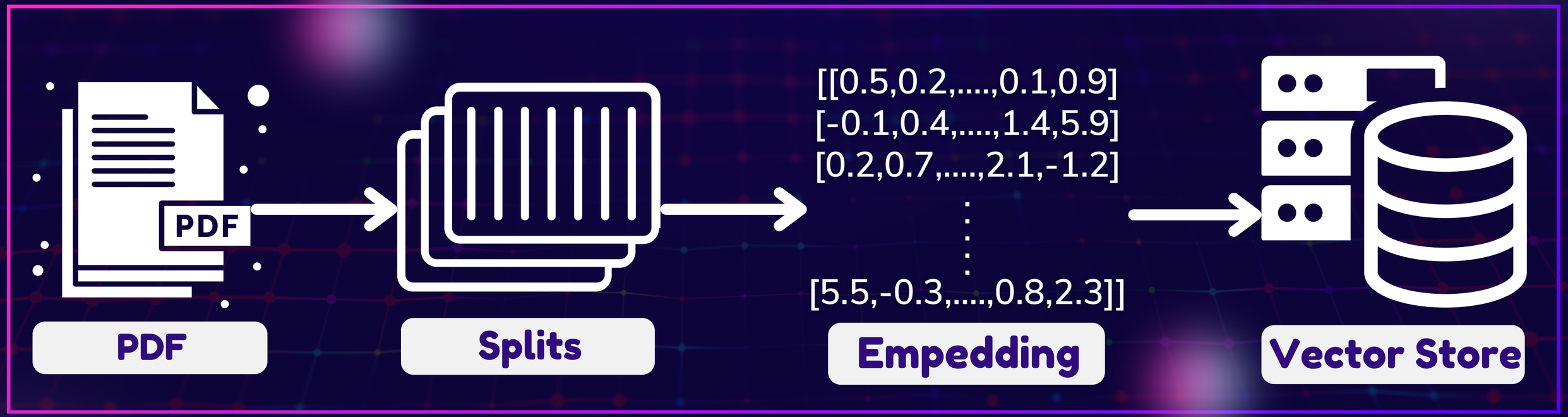
1. Split the Document: Once the PDF is loaded, split the document into smaller chunks or pages, which will be processed individually.

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.



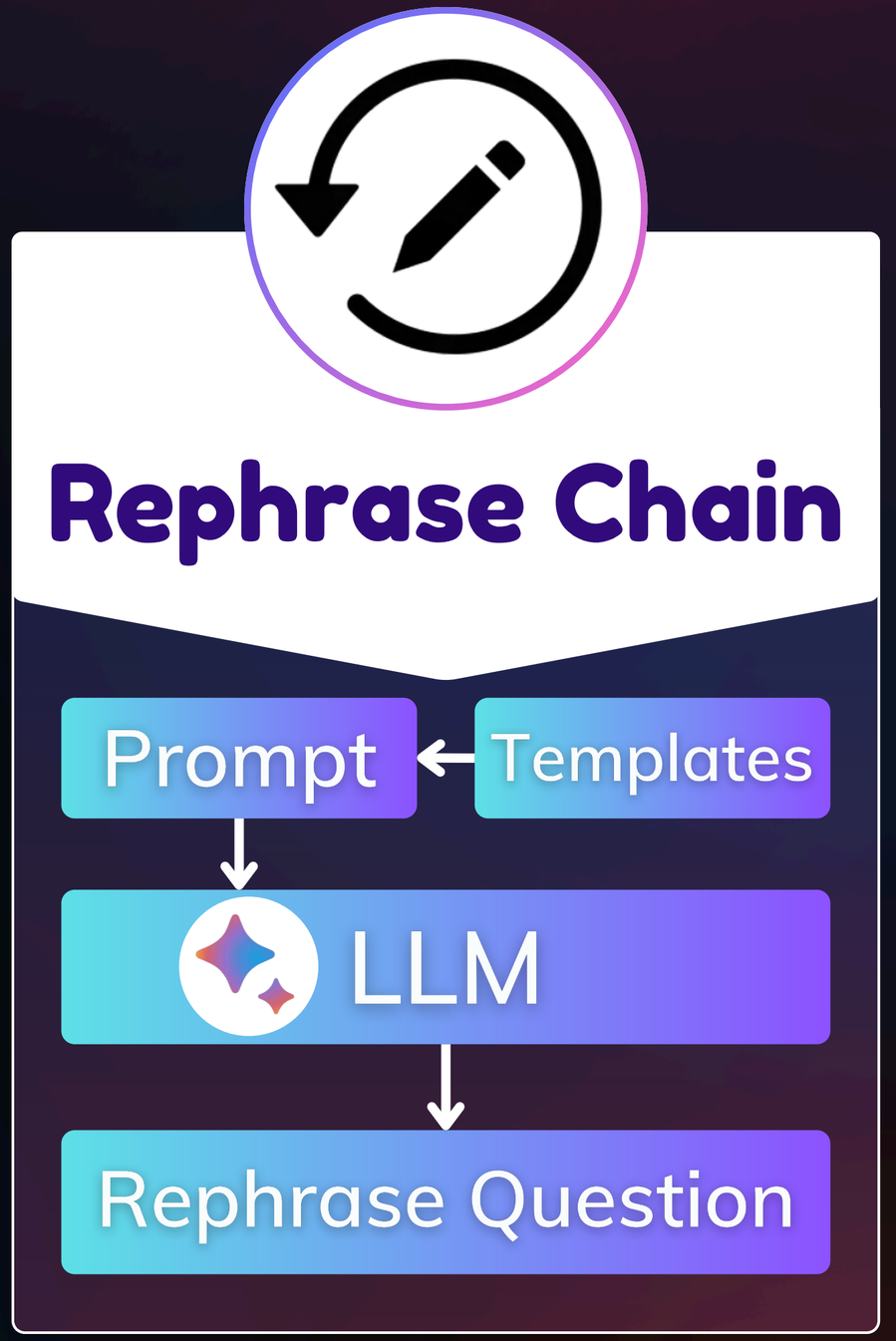

PDF
PDF
Splits
[[0.5,0.2,....,0.1,0.9]
[-0.1,0.4,....,1.4,5.9]
[0.2,0.7,....,2.1,-1.2]
[5.5,-0.3,....,0.8,2.3]]
Empedding
Vector Store

**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**

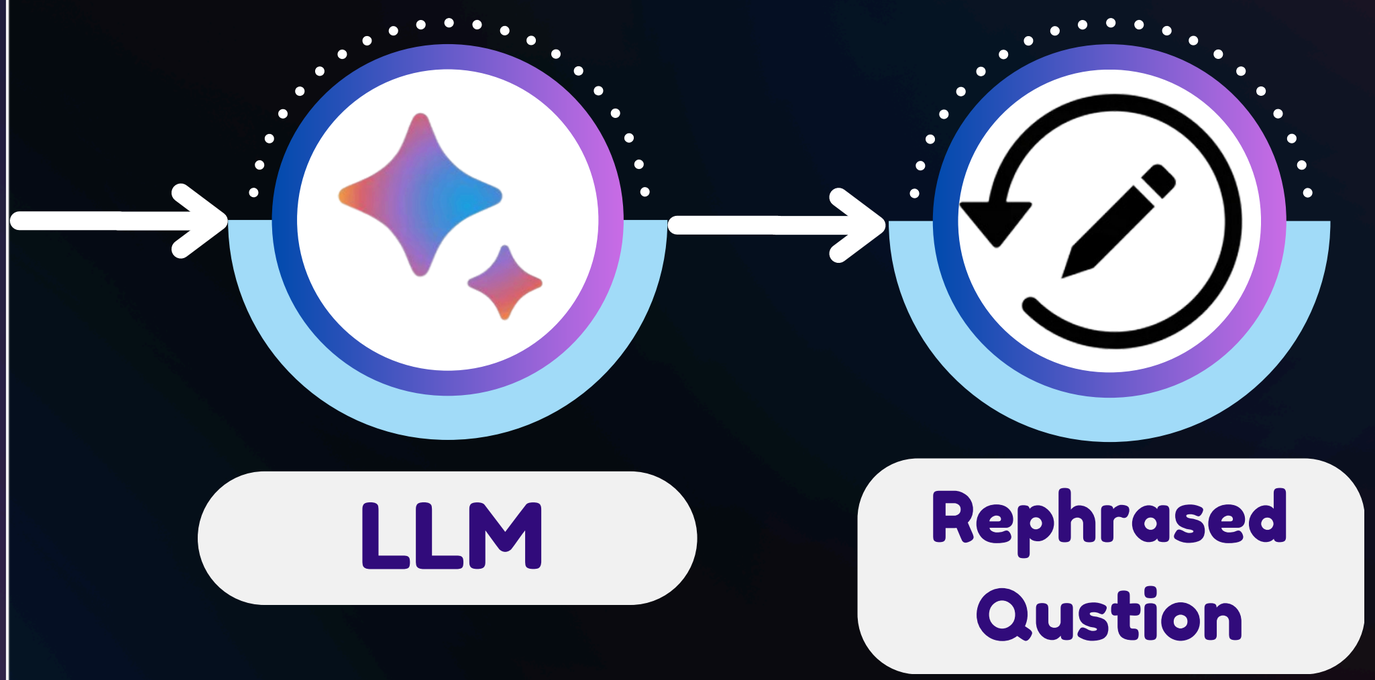


## Template

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:

## Prompot

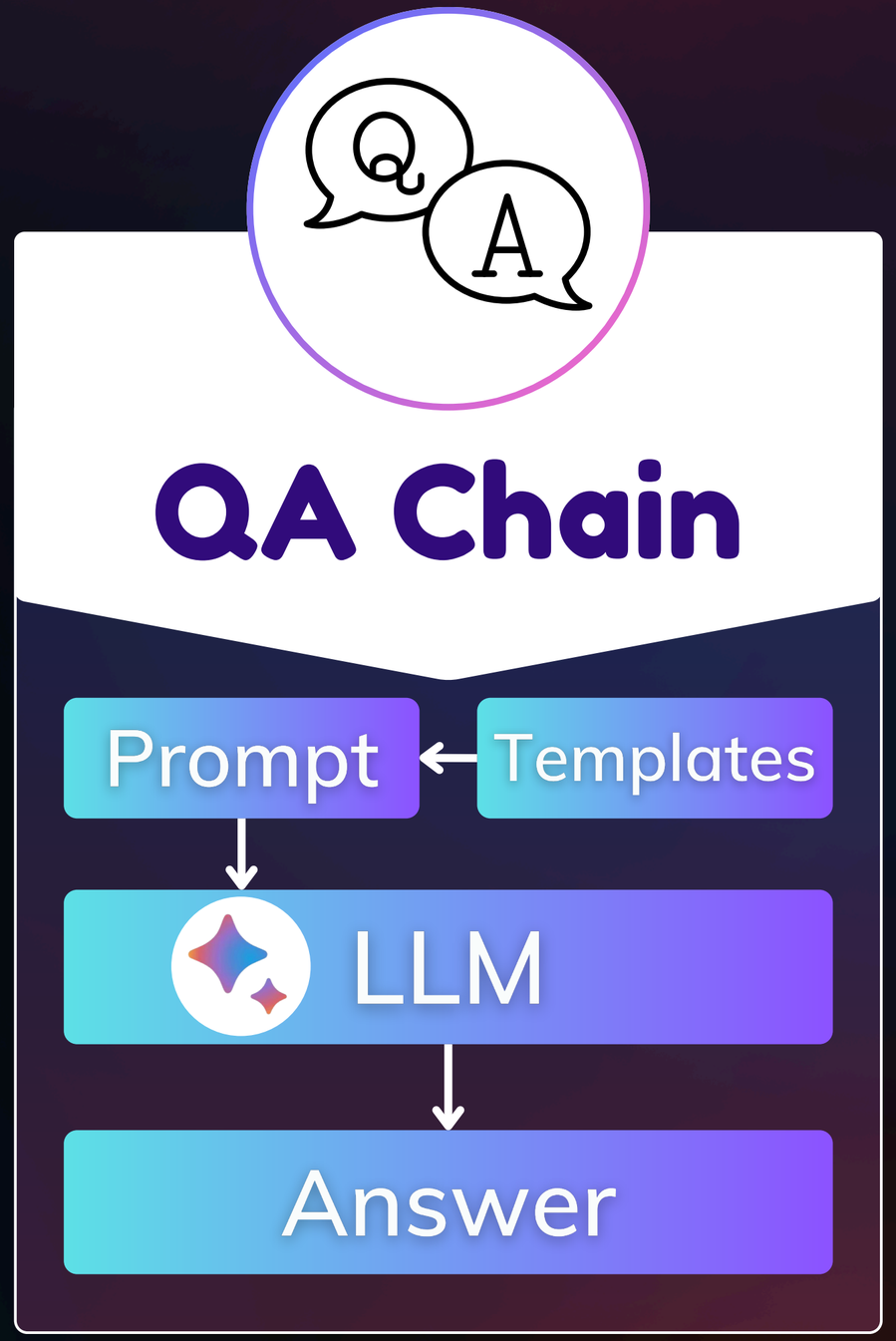
Combine the chat history and follow up question into a standalone question. Given a chat history and the latest user question which might reference context in the chat history, formulate a standalone question which can be understood without the chat history. Do NOT answer the question, just reformulate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



**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:

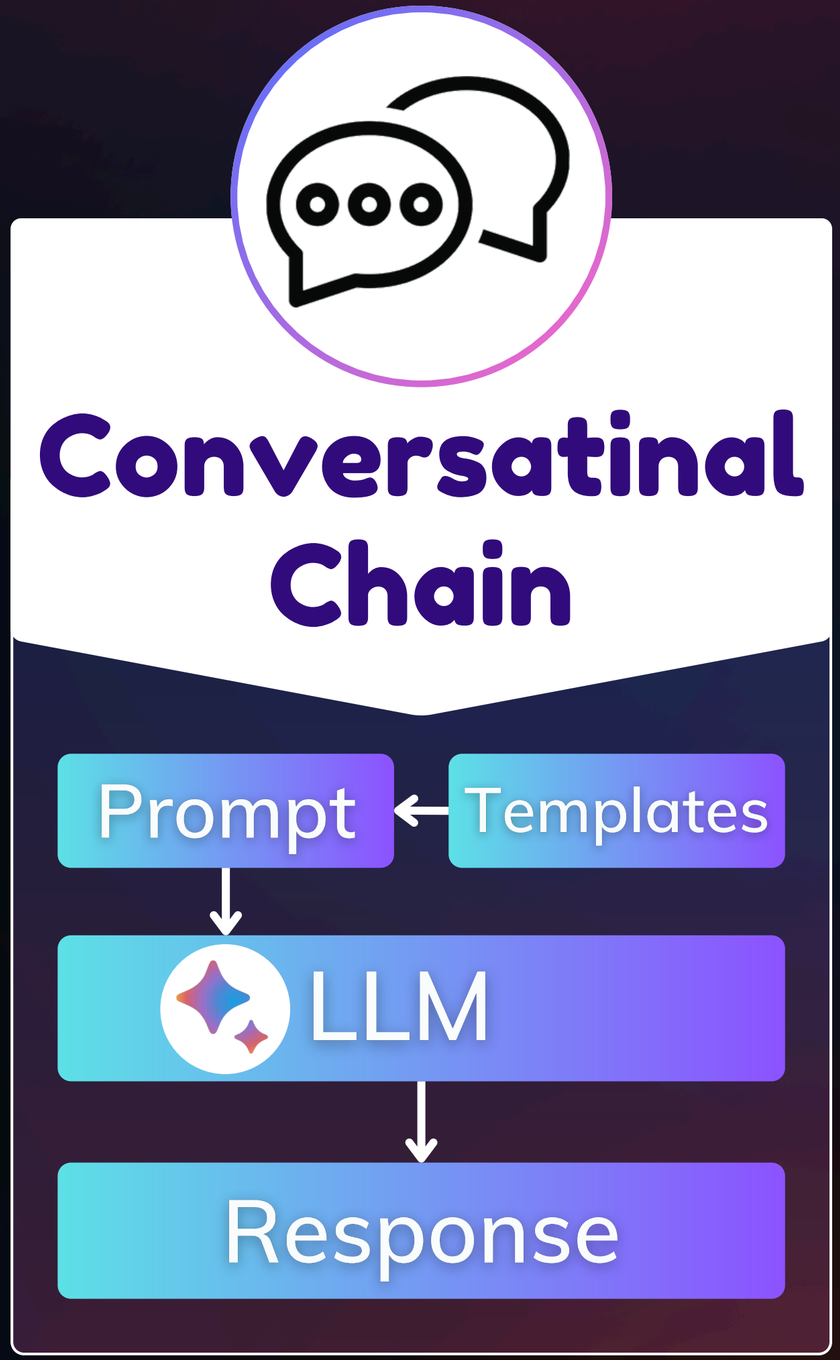
## Prompot

"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 Qustion
Helpful Answer:

**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



# 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 emojj.
Never makeany punctotions.
answer:

# Image captioning

## What isOpenCV ?

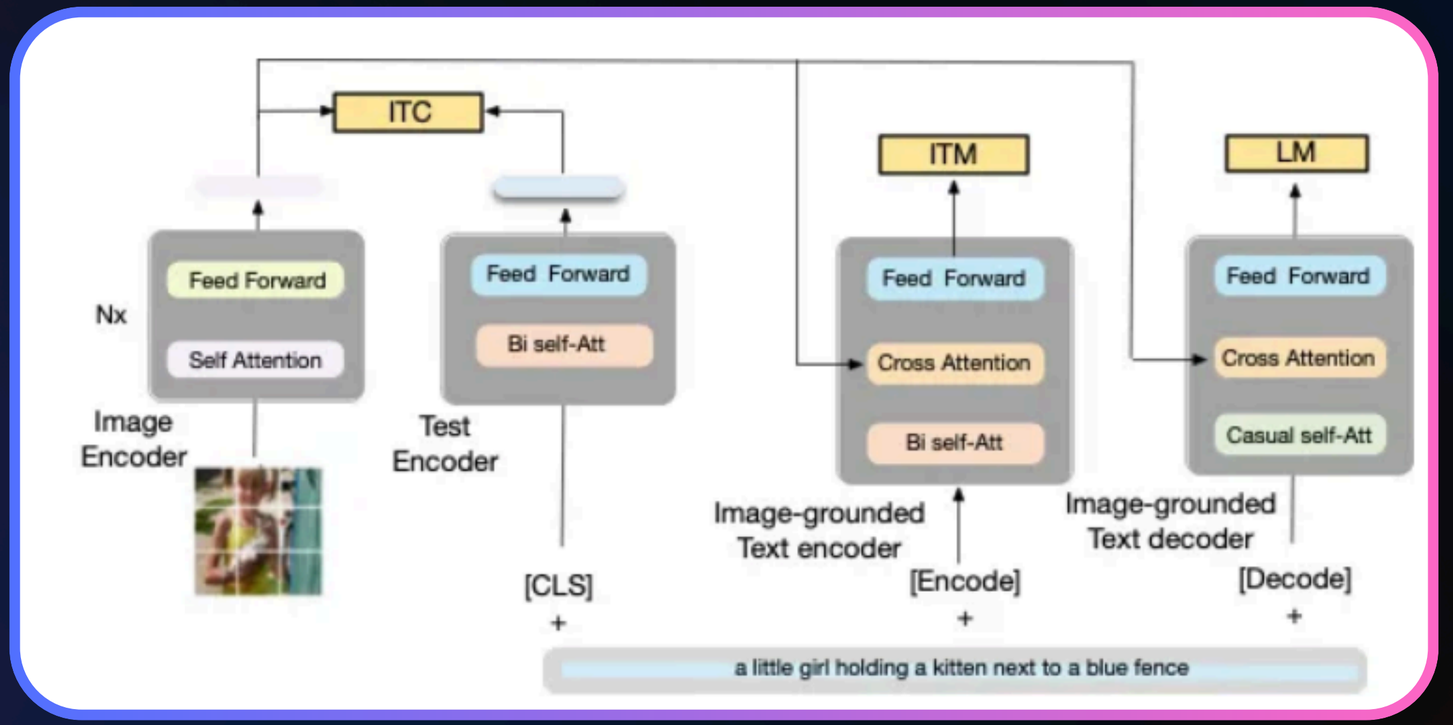
is a great tool for image processing and performing computer vision tasks. It is an open-source library that can be used to perform tasks like face detection, objection tracking, landmark detection

## Why we use openCV ?

OpenCV is used for capturing and handling live video feed from the camera, allowing for real-time interaction, and saving frames as images for further processing by the image processing and generation model.

# What is blip ?

The BLIP Vision Model (Bootstrapped Language-Image Pre-training) is a machine learning framework that combines language and vision tasks. It uses large-scale pre-training to effectively understand and generate text based on visual inputs, such as images or videos. This model is designed to enhance the performance of various applications, including image captioning, visual question answering, and image-text matching, by leveraging the synergy between visual and linguistic information.



CAIVA

# GUI

## CAIVA Presentation 2024

# Figma

## What is Figma ?

Figma is a web application for interface design, with additional offline features enabled by desktop applications for macOS and Windows. The feature set of Figma focuses on user interface and user experience design,

## Why we use Figma ?

easy to design user interface and easy to convert the design code to python

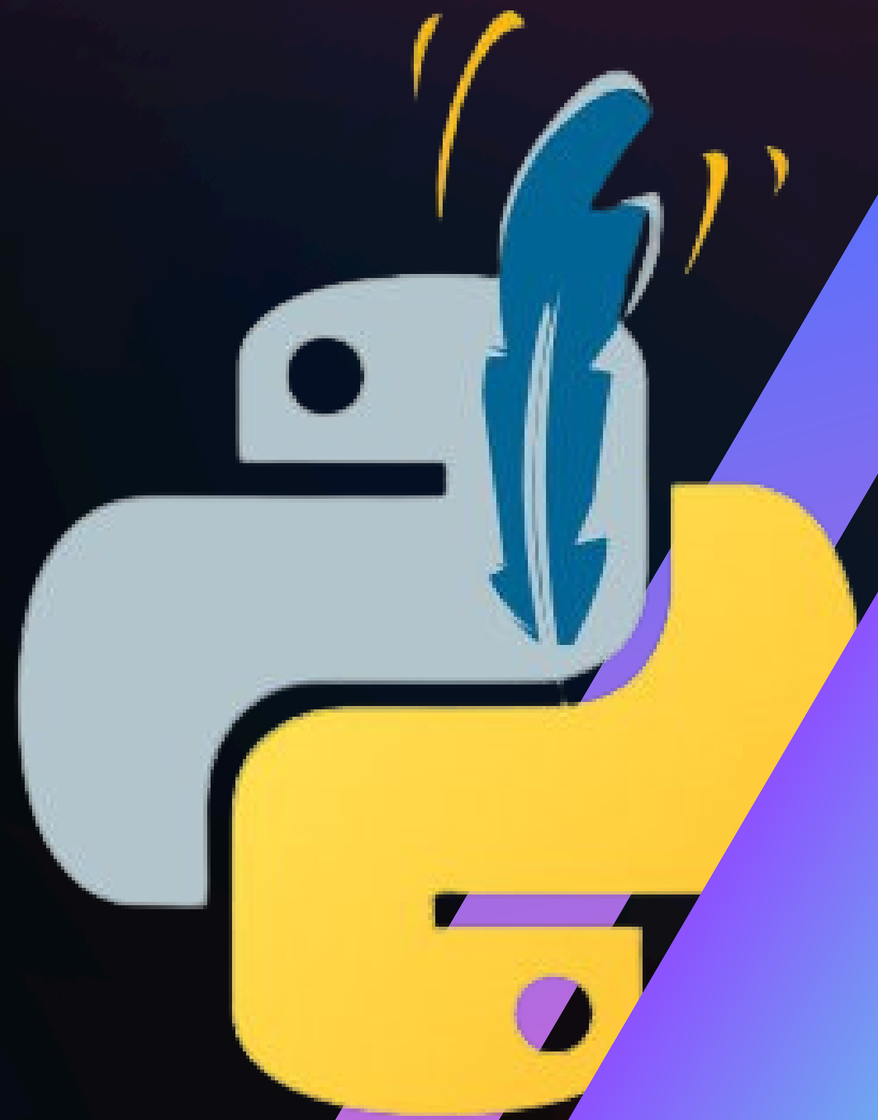
# TKinter

## What is Tkinter ?

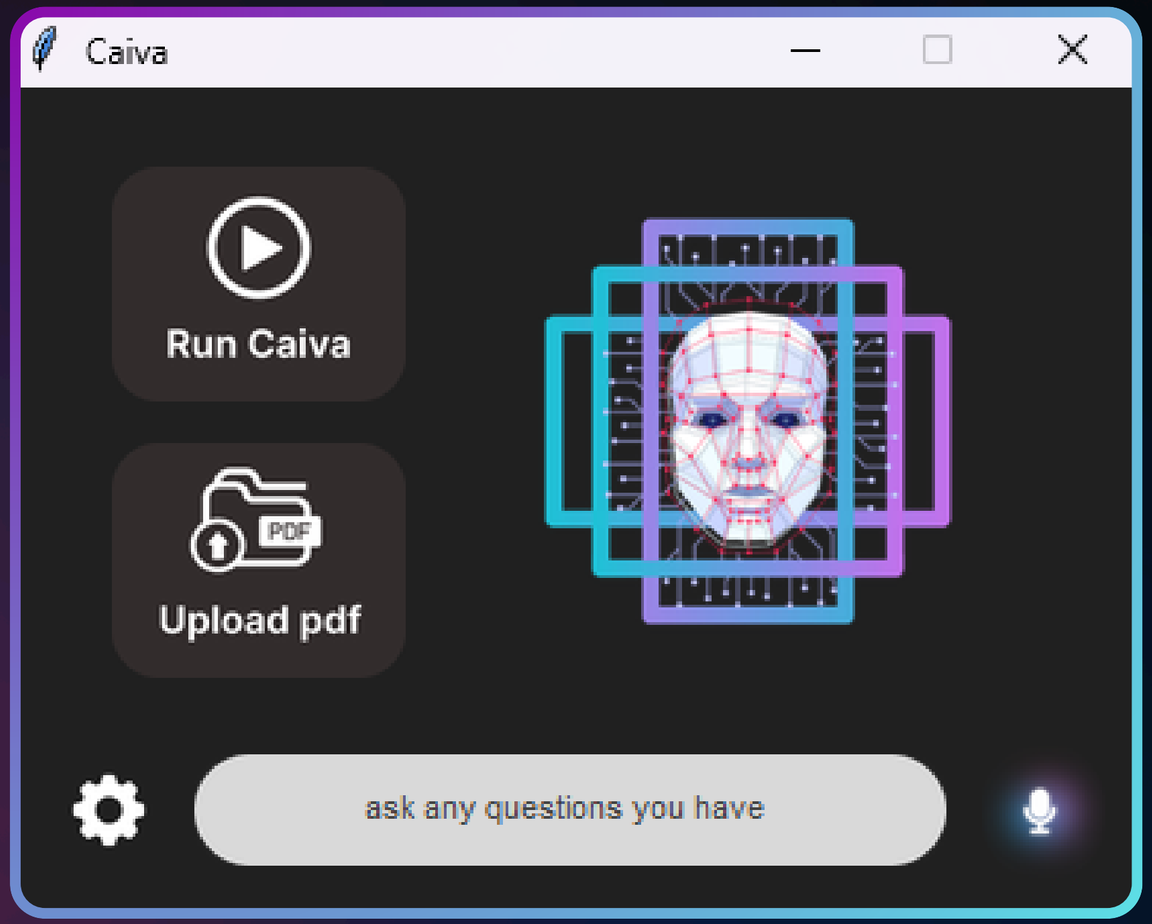
Tkinter is a Python library that can be used to construct basic graphical user interface (GUI) applications. In Python, it is the most widely used module for GUI applications

## Why use Tkinter ?

Tkinter is the first option for a lot of learners and developers because it is quick and convenient to use and suitable with langchain.



- 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.




# Tools

CAIVA

# Demo

## CAIVA Presentation 2024

# Demo

CAIVA

# Challenges

## CAIVA Presentation 2024

# Challenges

- New & diverse technology
- Limited resources
- Cost barriers
- Computational power
- Integration

CAIVA

# Time Plan

## CAIVA Presentation 2024

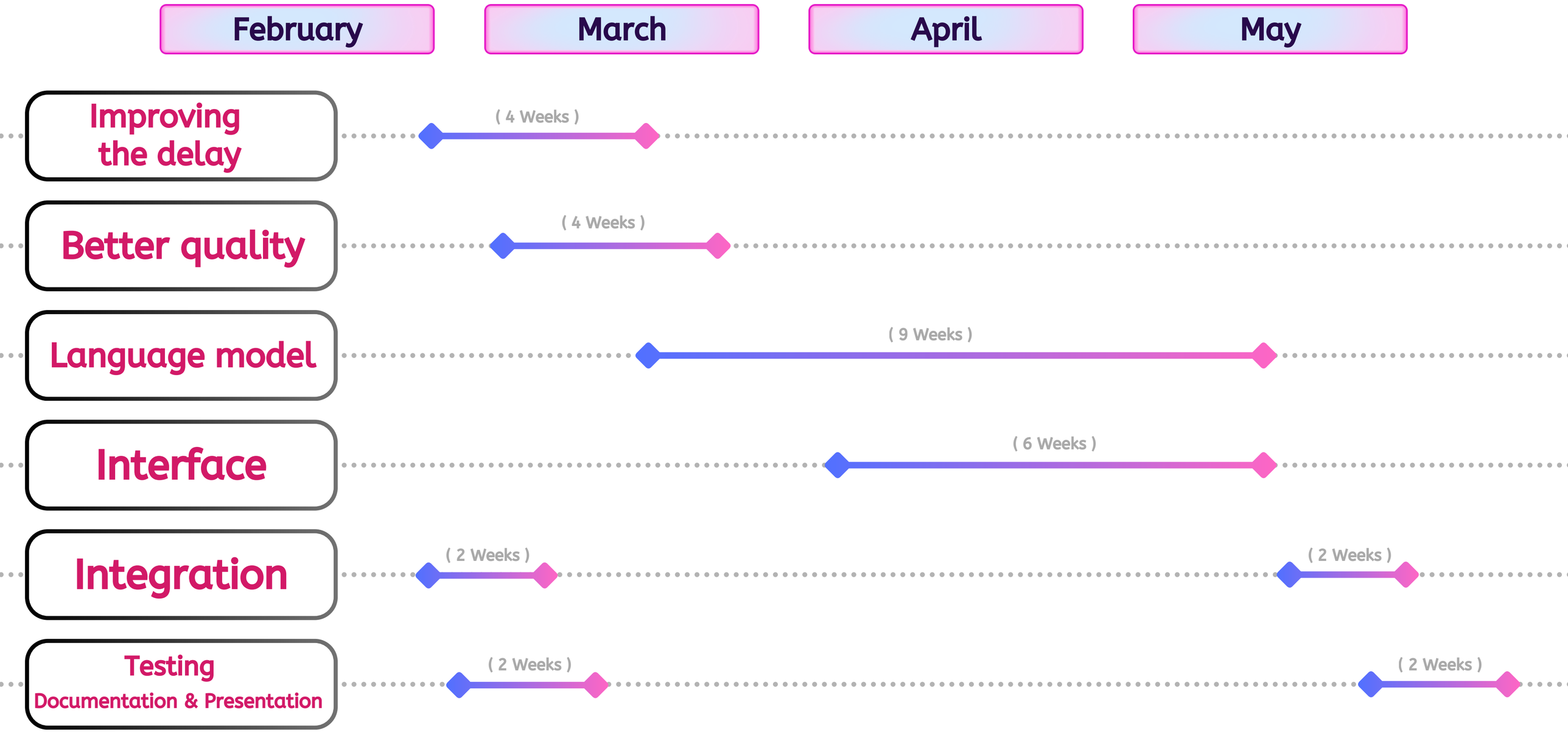
# Time Plan

| October | November | December | January |
|---|---|---|---|

- Discover lifelike characters and choice of animation Engine — ( 3 Weeks )
- Search & implementation about Text-to-Speech & Speech-to-Text Lip-sync & Firebase — ( 4 Weeks )
- **AI Techniques** Language Model & Sentiment Analysis — ( 8 weeks )
- Search & implementation about **Animation** — ( 4 weeks )
- **Integration** in Unreal Engine — ( 4 weeks )
- Create **Presentation** — ( 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 )

CAIVA

First Presentation - 2024

# Thank You

## For Your Attention