**Final Year Project**

**TALKTALES**

FYP Team

Ahmed Hayat 20I-0695

Sufian Ahmed Shah 20I-2654

Fahad Maqbool 20I-0693

Supervised by

Sir Umair Arshad

FAST School of Computing

National University of Computer and Emerging Sciences

Islamabad, Pakistan

2023

**Anti-Plagiarism Declaration**

This is to declare that the above FYP report produced under the:

Title: Talk Tales

is the sole contribution of the author(s) and no part hereof has been reproduced on as it is

basis (cut and paste) which can be considered as Plagiarism. All referenced parts have

been used to argue the idea and have been cited properly. I/We will be responsible and

liable for any consequence if violation of this declaration is determined.

Date: 11 October 2023

Student 1

Name: Fahad Maqbool

Signature: Fahad

Student 2

Name: Sufian Ahmed Shah

Signature: Sufian

Student 3

Name: Ahmed Hayat

Signature: Ahmed

Supervisor (Faculty)

Name: Sir Umair Arshad

Signature: Umair

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Authors’ Declaration**

This states Authors’ declaration that the work presented in the report is their own, and has

not been submitted/presented previously to any other institution or organization.

**Executive Summary:**

This is an AI-Powered 2D News Narration project in which AI is used as a core to advance in new experiences of delivering News using AI. Using top known technology and models such as LLM models, animation models, lip syncing models and various more advanced models, this project has a set an aim to innovate the system of how news are consumed and interacted with.

**Project Overview:**

In the current era, everywhere AI is advancing and showing its true capabilities. Our Project offers an interactive and engaging field where a virtual narrator is created and is ready to deliver news content realistically and dynamically. Users can interact with this virtually created narrator by providing prompt through a very user-friendly interface that spans news categories. These categories include different news types, sources and etc.

**Key Features:**

**News Content Selection:** The users can use a Dropbox to select their preferred news categories from which the news content will be generated.-

**Virtual Narrator:** A 2D virtual Narrator is ready to deliver any sort of news content realistically using gestures and real time facial expressions.

**Background Customization:** Users will be given some options of backgrounds from which they can choose a background of their liking which will be set in the 2d virtual environment.

**Subtitle Support:** Subtitles will be displayed as well so that listeners can have some sort of assistance to understand the news being delivered more efficiently.

**Language Learning:** As subtitles are present with the news being delivered so new language learners can learn the correct pronunciation of words.

**News Content:** The project is ready to deliver any sort of news content realistically in a virtual 2D narration and all the News content is updated with live news.

**Applications:**

* **Educational Purpose News:** People can Learn through news which is educational like delivering news on what new Machine Learning models.
* **Global News Delivery:** News associated to a country or global news can delivered based on user preferences.
* **Entertainment News Industry:** News based on entertainment such as what new products are launched etc.
* **Language Learning:** As news is being delivered verbally and subtitles support is present, people can learn language as it help them in learning how to pronounce words in a certain language.

**Ethical Considerations:**

* Content filtering mechanisms ensure unbiased and appropriate content generation.
* Robust privacy measures protect user data, and ethical content generation practices are implemented.

Table of Contents

[1. Introduction 1](#_Toc148296535)

[1. Create and execute an AI-Powered 2D Environment: 1](#_Toc148296536)

[2. Narrative Adaptability: 1](#_Toc148296537)

[3. Seamless Content Generation: 2](#_Toc148296538)

[4. User-Centered Innovation: 2](#_Toc148296539)

[5. Technical Excellence: 2](#_Toc148296540)

[2. Research on existing products 2](#_Toc148296541)

[3 . Project Vision 3](#_Toc148296542)

[3.1. Problem Statement 3](#_Toc148296543)

[3.2. Business Opportunity 3](#_Toc148296544)

[3.3. Objectives 3](#_Toc148296545)

[3.4. Project Scope 4](#_Toc148296546)

[3.5. Constraints 5](#_Toc148296547)

[3.6. Stakeholders Description 5](#_Toc148296548)

[3.6.1. Stakeholders Summary 5](#_Toc148296549)

[3.6.2. Key High-Level Goals and Problems of Stakeholders 6](#_Toc148296550)

[4. Software Requirements Specifications 6](#_Toc148296551)

[4.1. List of Features 6](#_Toc148296552)

[1. Interface for Users (UI): 6](#_Toc148296553)

[2. Integration of AI Models: 7](#_Toc148296554)

[3. Design of Virtual Personas: 7](#_Toc148296555)

[4. Creation and Integration of Content: 7](#_Toc148296556)

[5. Refinement of User Interaction: 7](#_Toc148296557)

[4.2. Functional Requirements 8](#_Toc148296558)

[1. User authentication: 8](#_Toc148296559)

[2. User Input Processing: 8](#_Toc148296560)

[3. AI Model Integration: 8](#_Toc148296561)

[4. Content Delivery: 8](#_Toc148296562)

[6. Virtual Persona: 8](#_Toc148296563)

[7. Background Replacement: 8](#_Toc148296564)

[4.3. Quality Attributes 9](#_Toc148296565)

[1. User Experience (UX): 9](#_Toc148296566)

[2. Performance: 9](#_Toc148296567)

[3. Scalability: 9](#_Toc148296568)

[4. Security: 9](#_Toc148296569)

[4.4. Non-Functional Requirements 9](#_Toc148296570)

[5. Iteration Plan: 10](#_Toc148296571)

[5.1. Iteration 1: 10](#_Toc148296572)

[Iteration 1 Goals: 10](#_Toc148296573)

[Details: 11](#_Toc148296574)

[5.2. Iteration 2: 12](#_Toc148296575)

[Iteration 2 Goals: 12](#_Toc148296576)

[Details: 12](#_Toc148296577)

[5.3. Iteration 3: 13](#_Toc148296578)

[Iteration 3 Goals: 13](#_Toc148296579)

[Details: 13](#_Toc148296580)

[5.4. Iteration 4: 13](#_Toc148296581)

[Iteration 4 Goals: 13](#_Toc148296582)

[Details: 14](#_Toc148296583)

[6. Architecture: 15](#_Toc148296584)

[7. Implementation: 15](#_Toc148296585)

[Frontend: 15](#_Toc148296586)

[Backend: 16](#_Toc148296587)

[Script Synopsis 16](#_Toc148296588)

[Voice Production 16](#_Toc148296589)

[Avatar Time Alignment: 16](#_Toc148296590)

[8. Use Cases: 17](#_Toc148296591)

[8.1. High Level Use Cases: 17](#_Toc148296592)

[1.Registration 17](#_Toc148296593)

[2.Avatar Selection 17](#_Toc148296594)

[3.News Selection 18](#_Toc148296595)

[4.Background Selection 18](#_Toc148296596)

[5.Language Selection 18](#_Toc148296597)

[6.Script Edit 19](#_Toc148296598)

[8.2. Extended Use Cases: 19](#_Toc148296599)

[1.Registration 19](#_Toc148296600)

[2.Avatar Selection 20](#_Toc148296601)

[3.News Selection 21](#_Toc148296602)

[4.Background Selection 22](#_Toc148296603)

[5.Language Selection 23](#_Toc148296604)

[6.Script Edit 24](#_Toc148296605)

[9. Use Case Diagram: 25](#_Toc148296606)

[10. System Sequence Diagram: 26](#_Toc148296607)

[1.Registration 26](#_Toc148296608)

[2.Avatar Selection 27](#_Toc148296609)

[3.News Selection 28](#_Toc148296610)

[5.Background Selection 29](#_Toc148296611)

[6.Voice Selection 30](#_Toc148296612)

[7.Voice Generation 31](#_Toc148296613)

[8.Authentication 32](#_Toc148296614)

[9.Script Generation 33](#_Toc148296615)

[10.News Domain Selection 34](#_Toc148296616)

[11. Class Diagram: 35](#_Toc148296617)

[12. JSON Schema: 35](#_Toc148296618)

[13. Activity Diagram: 37](#_Toc148296619)

[37](#_Toc148296620)

[14. User Manual: 38](#_Toc148296621)

[Frequently Asked Question (FAQS): 38](#_Toc148296622)

# Introduction

A new era of interactive and immersive content distribution has been ushered in by the combination of natural language synthesis and processing in the ever-evolving field of artificial intelligence (AI). This project sets out on a revolutionary quest to fully utilize the potential of state-of-the-art AI capabilities, as demonstrated by outstanding models like Bard and ChatGPT. By showcasing AI's competence in a variety of tasks—from data retrieval to proposal writing to content creation based on user interactions—these models have set the stage.

By concentrating on the areas of speech synthesis, text comprehension, picture processing, and interactive content production, our research seeks to use the current state of AI technology. When these components come together, a cutting-edge platform—an AI-powered 2D News and Information Narrator—will be born. With its lively demeanor, this virtual narrator will work as a medium for news updates, narratives, information sharing, and critical evaluations. All of this will be done while deftly adding pertinent images and videos to improve the user experience.

The focus of this endeavor is the AI's ability to provide engaging material in a variety of areas. For example, our virtual narrator creates eloquent tales for real-time news updates when a user selects the "news" category. When users explore the "stories" category, the AI will generate creative narratives with compelling storylines. Similar to this, those looking for "information" will find educational material, while users looking for "reviews" will find enlightening analyses and assessments.

This project, encompassing a diverse range of AI competencies, not only promises to redefine the way we consume content but also to revolutionize how we interact with it. At its core, this venture aims to:

## Create and execute an AI-Powered 2D Environment:

Our main objective is to create a 2D environment that can serve as a flexible narrator for a variety of content types. To ensure an enhanced narrative experience, this entails not only the production of textual content but also the incorporation of pertinent images and videos.

## Narrative Adaptability:

Each user will have a personalized, immersive, and engaging news telling experience as the virtual narrator's character will dynamically adjust to various content genres and user interactions.

## Seamless Content Generation:

This project's main strength is the AI's capacity to produce interesting content instantly. Real-time user prompts and choices are at the core of this content creation process, guaranteeing that every interaction is interesting and unique.

## User-Centered Innovation:

We are dedicated to prioritizing the user experience in all that we do. This includes creating an interface that is easy to use, perfecting avatar identities, and making sure that the technology is still understandable and accessible to many people.

## Technical Excellence:

We recognize that this project is based on complex technical hurdles. Our method is based on strong technological underpinnings and includes everything from model selection through backend integration, animation, and content creation.

We will delve into the nuances of this project in the pages that follow, going over each element in detail and showing how the smooth integration of AI technology will completely change how we interact with news, information, stories, and reviews. This project seeks to push the limits of what is possible with contemporary AI by establishing new benchmarks in content narration and interaction through thorough study, careful design, and iterative improvement.

# 2. Research on existing products

Whilst conducting our research, Three major players have been identified from our analysis of the market's current product offerings: DI-D, Synthesia, and Kredio. These products all have the same goal of providing an easy-to-use platform where users may choose their chosen speaker, avatar, and script for automated video narration. The landscape of virtual storytelling and information distribution has changed because of this creative method of content development. These technologies offer a simple way to quickly convert text-based scripts into interesting video presentations, providing a glimpse into the future of AI-powered content creation as customers want personalized and dynamic video content. We want to better understand the industry dynamics and find areas where our AI-Powered 2D News Narration system may be improved by looking into these platforms.

# 3 . Project Vision

The project imagines a day when advanced artificial intelligence (AI) technology will make consuming material in the future an immersive, personalized, and interactive experience. Our goal is to develop a 2D news and information narrative powered by artificial intelligence that will let users interact with a virtual character and improve their access to news, stories, information, and reviews.

## 3.1. Problem Statement

Users in the current digital environment encounter several significant obstacles when consuming content. Information overload, passive content delivery, a lack of personalization, and problems with accessibility and diversity are some of these difficulties. On the other side, content producers struggle with the growing rivalry for users' attention as well as the technological challenges of using AI technology for dynamic content creation. All of these difficulties point to the necessity of a creative solution that completely rethinks the paradigms of user interaction and content distribution.

## 3.2. Business Opportunity

There is a huge business opportunity at the intersection of AI technologies and content consumption. The need for dynamic and captivating information distribution systems that accommodate personal tastes and inclinations is on the rise. This project takes use of the chance to develop a distinctive platform that may be applied to news distribution, education, entertainment, and language learning, among other fields. This initiative is not only technologically revolutionary but also financially feasible due to the significant potential for revenue creation through partnerships, subscriptions, and technology licensing.

## 3.3. Objectives

The primary objectives of this project are as follows:

* + - Create and put into action a 2D AI environment that acts as a virtual narrator for a variety of content types.
    - Create a system that dynamically creates material, such as news updates, articles, reviews, and information, and pairs it with pertinent images and videos.
    - Make sure the story is flexible enough to adapt to different kinds of information and user interactions, providing an immersive experience for the user.
    - Provide content instantly, meeting the demand for frequent updates and interactive interaction.
    - Offer an intuitive user interface that promotes inclusion and accessibility.
    - To provide a flawless user experience, optimize technical elements such as AI model selection, integration, and performance.
    - Always develop and improve to keep up with changing user demands and technology breakthroughs.

## 3.4. Project Scope

The project scope encompasses the following key areas:

* + - Conducting research and choosing suitable AI models for image processing, text comprehension, speech synthesis, and interactive content production.
    - Frontend development of an interface for users to choose content categories and enter prompts.
    - Backend integration for content generation, user input processing, and dynamic 2D scenario creation with virtual narrators.
    - Creating an animated 2D virtual character for the narrator, complete with facial emotion development and avatar construction.
    - Content creation and integration, making sure that speech, text, images, and videos are all coherent and relevant.
    - Background creation according to user inclinations and chosen niche.
    - Enhancement of user engagement with realistic animation, customization, and emotion recognition.
    - Putting into practice technological factors including resource optimization, deep learning frameworks, fine-tuning, and training data.
    - To improve user experience, extensive testing, feedback collection, and iterative refinement are used.

## 3.5. Constraints

The project is limited in some ways, including:

* + - Technical restrictions on the computing power and capabilities of AI models.
    - Limited resources for model training and integration in terms of money, time, and experience.
    - Ensuring that interactions occur in real-time or very close to it by making sure the background replacement process moves quickly enough.
    - Handling possible difficulties with user customization and emotion recognition.
    - Respecting privacy issues and ethical considerations when creating material and interacting with users.

## 3.6. Stakeholders Description

### 3.6.1. Stakeholders Summary

Several parties are involved in the project, including:

* + - End Users: Those looking for news, tales, information, and reviews are among the main consumers of the AI-powered 2D narrator.
    - Content Creators: Organizations looking to use the platform for interaction and content delivery.
    - Educational Institutions: Businesses that might distribute educational information via the platform.
    - Media Houses: Organizations in the news and entertainment sectors seeking to implement cutting-edge methods of distributing material.
    - Language learners: People looking for assistance with language acquisition and information that is appropriate for their setting.
    - Data scientists and AI developers: Professionals in charge of choosing, integrating, and training models.
    - Regulatory Authorities: Organizations in charge of the moral and legal ramifications of AI content creation.
    - Stakeholders and Investors: People who are concerned about the project's success and financial sustainability.

### 3.6.2. Key High-Level Goals and Problems of Stakeholders

* + - End Users: Demand accessible material, a user-friendly interface, and tailored and interesting content delivery.
    - Content Providers: Try to use the platform to increase user interaction and income production.
    - Educational Institutions: Need a useful tool to provide diversified learners with instructional content.
    - Media companies: Look for cutting-edge ways to enthrall viewers with news and entertainment material.
    - Language learners: Request pronunciation assistance and content that is appropriate for the setting.
    - Data scientists and AI developers should strive to efficiently create and incorporate AI models in order to accomplish project goals.
    - Regulating Authorities: Make sure AI-powered content creation complies with ethical and legal standards.
    - Stakeholders and investors: anticipate a financially sound initiative with room to grow and expand.

# 4. Software Requirements Specifications

## 4.1. List of Features

### 1. Interface for Users (UI):

* + - Interactive Interface: Users will be able to input prompts and choose content categories using an intuitive interface provided by the system.
    - Niche Selection: To personalize the background of the virtual storyteller, users can select different niches.
    - Real-time Interaction: Users and the virtual narrator will be able to interact in real-time thanks to the user interface.

### 2. Integration of AI Models:

* + - Speech Synthesis: Using models for speech synthesis, generated text can be turned into speech.
    - Text Comprehension: Applying text comprehension models to interpret context and user instructions.
    - Image Processing: Combining models for image processing to manage visual content and backdrop replacement.
    - Interactive Content Generation: Using AI models to create content that changes in real time in response to user inputs.

### 3. Design of Virtual Personas:

* + - Avatar Creation: creating visually appealing avatars.
    - Animation: Adding motions and face expressions to the virtual narrator's animations.
    - Emotion Detection: To modify the persona's expressions, AI algorithms for emotion recognition are used.

### 4. Creation and Integration of Content:

* + - Content Categorization: Divide user-selected material into sections such as news, information, tales, and reviews.
    - Dynamic Content Generation: the process of creating written content in real time and incorporating images and videos into it.
    - Background replacement: altering backdrops according to user choices and chosen niches.
    - Narrative Flexibility: Making sure the avatar of the virtual narrator fits the chosen genre of content.

### 5. Refinement of User Interaction:

* + - Realistic Animation: Improving avatar movements to make them more organic and flowing.
    - Personalization: Adapting the responses of the virtual narrator to the user's interactions.
    - Emotion Recognition: Modifying the character’s emotions so they look really realistic.

## 4.2. Functional Requirements

### 1. User authentication:

* To access customized features, users can register and log in.

### 2. User Input Processing:

* + - The system handles choices for content categories and prompts supplied by the user.

### 3. AI Model Integration:

* + - Speech synthesis, text comprehension, picture processing, and content creation are all accomplished by AI models that are integrated.

### 4. Content Delivery:

* + - The system provides users with real-time access to created text, audio, images, and videos.

### Virtual Persona:

* + - The avatar changes based on user interactions and content categories.
    - Textual descriptions and emotions are used to generate facial expressions and animation**.**

### Background Replacement:

* + - A seamless transition between the new and avatar backdrops is guaranteed. The system updates backgrounds according to user niche preferences.

## 4.3. Quality Attributes

### User Experience (UX):

Through interactive content delivery and persona animations, the system seeks to offer an immersive and captivating user experience. A wide variety of people can easily navigate and use user interfaces because of its design.

### Performance:

An essential quality attribute is the ability to create and interact with material in real-time or almost real-time. The system should effectively manage several concurrent user requests.

### Scalability:

To handle an expanding user base and rising content demand, the platform should be built to be scalable.

### Security:

Make sure that user accounts and personalization data are handled with care, as well as data security and privacy. Keep an eye out for any weaknesses and defend against illegal access.

## 4.4. Non-Functional Requirements

Model Accuracy: AI models ought to offer precise image processing, text understanding, and speech synthesis. Content creation needs to be logical and appropriate for the situation.

Robustness: To ensure a smooth experience, the system should be able to adapt gracefully to changes in user prompts and content categories.

Usability: There should be clear prompts and feedback mechanisms, and the user interface should be simple to use. It is imperative to provide accessibility features that accommodate people with disabilities.

Resource Optimization: Make sure that computational resources are used effectively, taking into account the creation of content in real time and the training of models.

Ethical Considerations: The system ought to conform to ethical standards, steering clear of the creation of biased or improper content. Comply with applicable legislation to protect user privacy and data.

Documentation: Including user manuals, developers and users should have access to thoroughdocumentation.

# 5. Iteration Plan:

****

## 5.1. Iteration 1:

### Iteration 1 Goals:

Data Collection: Compile the text, picture, and audio datasets that are required.

Preprocessing: Get the gathered data ready and cleaned so that the model may be trained.

Frontend Development: Design the first user interface that allows for content selection and prompt input.

Backend using Flask API: Configure the Flask API backend architecture to handle requests and accept input from users.

Training Large Language Models (LLMs): GPT-3 and other LLMs are trained and refined to produce textual material in response to user cues.

Content Integration: Use LLMs to put the reasoning into practice to create coherent content dynamically.

Voice Generation: To turn produced text into speech, incorporate text-to-speech (TTS) capabilities.

Avatar Creation: Start creating the virtual narrator's 2D avatars.

Lip Syncing: Create lip-syncing systems to match speech and avatar movements.

### Details:

Data Collection: This stage entails gathering background photos and datasets that will be used as LLM training data. Articles, tales, reviews, and educational text with related images can all be found in datasets.

Preprocessing: To clean and prepare the gathered data, preprocessing is necessary. This covers formatting audio data, resizing images, and cleaning text.

Frontend Development: Provide a rudimentary front-end interface where users can choose content categories (news, stories, information, reviews) and enter prompts.

Backend with Flask API: Create the backend framework with Flask API to facilitate communication between the AI models and the frontend.

LLM Training: Use the gathered text data to train and fine-tune LLMs so they can produce material that is both coherent and contextually relevant.

Content Integration: Put the logic in place so that it will produce content in response to user input and show it in the user interface. Make sure the content that is produced is interesting and logical.

Voice Generation: Include a TTS system to speak the generated text aloud to consumers so they may experience it audibly.

Avatar Creation: Get started on creating the two-dimensional avatars that will stand in for the virtual narrator.

Lip Syncing: To improve realism, create systems that will match the avatar's lip movements with the speech that is created.

## 5.2. Iteration 2:

### Iteration 2 Goals:

Frontend Improvement: Upgrade the user interface to make it more interactive and user-friendly.

Backend using Flask API: Keep improving the backend framework to accommodate more features.

Urdu Text-to-Speech (TTS): To increase the number of available languages, add support for Urdu TTS.

Background Generation and Integration: Provide the capability to smoothly integrate and replace backdrops according to user preferences.

Image Creation and Scene Mapping: Create images and align them with the text to make sure they are consistent and relevant.

Subtitle Integration: Add subtitles to improve language support and accessibility.

### Details:

Frontend Improvement: Make the frontend more intuitive and user-friendly by iteratively improving it. Apply user feedback to improve the UI.

Backend using Flask API: Keep improving the infrastructure in the backend, making it more capable of handling complicated interactions and more content creation.

Text-to-Speech (TTS) in Urdu: To serve a larger user base, incorporate Urdu TTS features to extend language support.

Creation and Integration of Background: Create background replacement algorithms according to user niche preferences. Make sure the avatars merge seamlessly with the fresh backdrops.

Image Creation and Scene Mapping: Produce visually appealing scenes that complement the story by generating images and assigning them to relevant content categories.

Subtitle Integration: To improve accessibility and assist users with hearing problems, incorporate subtitles into the content delivery system.

## 5.3. Iteration 3:

### Iteration 3 Goals:

Creating Videos and Mapping Scenes: Expand the platform's video generation and mapping capabilities to create more engaging user experiences.

Frontend Enhancement: Persist in refining the frontend interface in response to user input.

Backend using Flask API: Improve the scalability and performance of the backend infrastructure.

### Details:

Video Creation and Scene Mapping: Add video creation and scene mapping to the platform to improve its content generation capabilities. This makes it possible to tell stories that are more engaging.

Frontend Improvement: Keep making adjustments to the frontend in response to user input and usability testing. Make sure the UI stays user-friendly and captivating.

Backend using Flask API: Keep refining the backend architecture to manage the growing intricacy of producing and distributing video content.

## 5.4. Iteration 4:

### Iteration 4 Goals:

Frontend Enhancement: Complete the front end design for the best possible user experience.

Backend using Flask API: Evaluate and optimize the backend architecture thoroughly.

Testing and Feedback: To find and fix any problems, conduct thorough testing, including user testing.

Final Demo Preparation: Make sure that every component functions flawlessly as you get ready for the project's final demonstration.

### Details:

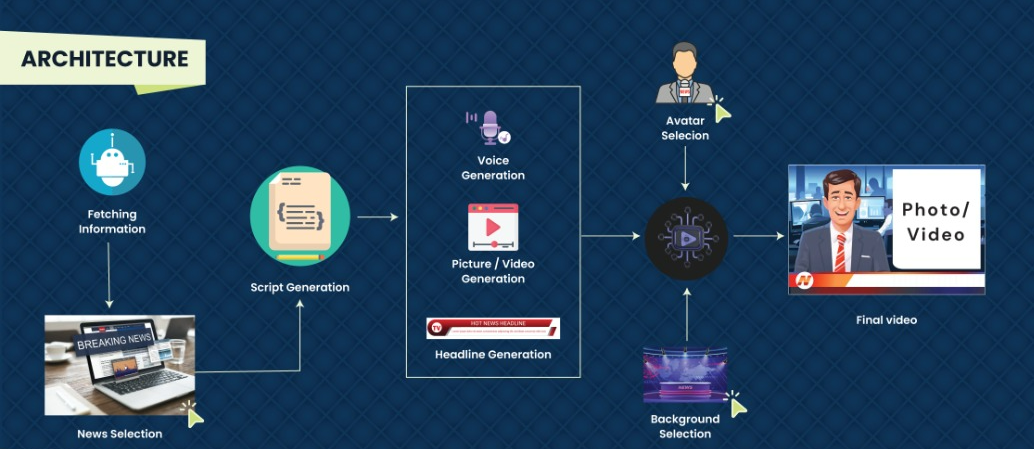
Frontend Improvement: To guarantee that the front-end interface lives up to user expectations, finalize it while taking into account all improvements and suggestions from earlier versions.

Reverse using Flask API: Make sure the backend infrastructure is capable of handling real-time interactions, video generation, and a range of user scenarios by conducting thorough testing and optimization.

Testing & Feedback: Conduct comprehensive testing, encompassing usability, performance, and functionality assessments. Get input from users to find any problems that still exist and improve the user experience.

Final Demo Preparation: Get ready for the project's final demonstration, making sure that every element functions flawlessly and is well-integrated to highlight the platform's potential to stakeholders.

# 6. Architecture:



# 7. Implementation:

## Frontend:

We have used react js for front-end development in which we have created components.

1. The user is displayed a input field where they can enter any phrase. On the basis of their phrase news articles will be fetched to the user.
2. Avatar Selection: The front end offers the user the chance to select an avatar.
3. Language Selection: The frontend offers the user to select a language.
4. Background Selection: The front end will show a list of backgrounds to choose from and that will be integrated in the video as a result.
5. News Selection: the news links are transferred to the front end and then the user selects the articles link that the user wants to generate a headline w.r.t it.
6. direct link: input field is provided where the user will provide a link from where news will be extracted.
7. Flask API paired with Ngrok: In this case, data collection from the frontend is done via the Flask API. Sending data to a server connected to Ngrok, specifically on Colab, is the responsibility of the Flask API. The only purpose of Ngrok is to transfer data between Colab and the nearby Flask server. In this case, the procedure comprises locally scraping articles and sending them over Ngrok, which is linked to a Flask server on Colab. A model, in this example the Vistral 7b model, is running on the Colab server. After analyzing the provided data, the model produces an objective script that is returned. In summary, Ngrok enables secure transfer between Colab and the local Flask server, Flask is used to communicate between the frontend and the server, and the Colab server processes and generates unbiased scripts based on the locally scraped articles using the Vistral 7b model.
8. Video Display: After the LLM produces the script, lip syncing on the video and after mapping of background and image the final video is displayed on a separate frontend page of the website.

## Backend:

How to Select the Best Script Generator

Initially, we generated scripts using Falcon 7b, a huge language model. But Falcon 7b's script quality fell short of our expectations, and its delivery of results was sluggish. We investigated prompt engineering and fine-tuning strategies to make this better. These efforts did not yield the necessary quality from Falcon 7b.

Consequently, we made the decision to move to Llama 2 7b, which produced remarkable scripts and had a quicker inference time. We refined prompts and adjusted Llama 2 7b to remove biases and enhance the quality of the script.

## Script Synopsis

We utilized Llama 2 7b as our script generator to write scripts that were based on the news articles. We concentrated on creating just the narrative for the news anchor, ensuring clarity and news delivery.

I'll go into detail about our experience testing using script generators in this narrative. We started out using Falcon 7b, a reliable language model. Sadly, Falcon 7b's script quality fell short of our expectations and its performance was slow. Even with rapid engineering and fine-tuning efforts, Falcon 7b fell short of our expectations.

We then decided it would be strategically advantageous to switch to Llama 2 7b. This change paid off, as Llama 2 7b showed faster inference times in addition to producing impressive scripts. To remove biases, we improved the prompts and customized Llama 2 7b, which improved the script quality in the end.

As we have mentioned, at first, we used Falcon for comprehensive testing, yielding precise findings. Falcon's import context performed poorly, nevertheless, when it came to performing bias reduction for the three-article challenge. After learning about Mistral 7B, a model that produces high-quality results with a significant import context, we thought about using it for the three-article challenge. We will adjust or expedite engineer work if its inference time coincides well. If not, we will investigate substitute models to guarantee accurate inference time and quality, thereby fine-tuning our range of options.

## Voice Production

We required a Text-to-Speech (TTS) model in order to make the script come to life. After testing out a number of models, we discovered that Seamless 4t Large performed the best for English TTS in the first run.

## Avatar Time Alignment:

Avatar synchronization with the generated audio was the last missing component. To create a more engaging news presentation, we used the WavToLip tool to make sure the avatars' lip movements matched the spoken words.

To summarize, we used WavToLip for avatar synchronization, Seamless 4t Large for voice generation, and Llama 2 7b for script generation in our solution. We were able to produce interesting and educational news presentations by following these procedures.

# 8. Use Cases:

## 8.1. High Level Use Cases:

### 1.Registration

|  |  |
| --- | --- |
| **Use Case** | **Registration** |
| **Actors** | **User, System** |
| **Type** | **Primary, offstage** |
| **Description** | **The user provides specific details required for registration in a new form page. After completing it, the user successfully gets registered.** |

### 2.Avatar Selection

|  |  |
| --- | --- |
| **Use Case** | **Avatar Selection** |
| **Actors** | **User, System** |
| **Type** | **Primary, offstage** |
| **Description** | **The user is provided with some options of avatars from which the user chooses the avatar of their liking.** |

### 3.News Selection

|  |  |
| --- | --- |
| **Use Case** | **News Selection** |
| **Actors** | **User, System** |
| **Type** | **Primary, offstage** |
| **Description** | **The User is provided with input field where the enter a phrase to search news on for.** |

### 4.Background Selection

|  |  |
| --- | --- |
| **Use Case** | **Background Selection** |
| **Actors** | **User, System** |
| **Type** | **Primary, offstage** |
| **Description** | **The user is provided with some options of backgrounds from which the user chooses the background of their liking.** |

### 5.Language Selection

|  |  |
| --- | --- |
| **Use Case** | **Language Selection** |
| **Actors** | **User, System** |
| **Type** | **Primary, offstage** |
| **Description** | **The user is provided with some options of languages for avatars from which the user chooses the language of avatar of their liking.** |

### 6.Script Edit

|  |  |
| --- | --- |
| **Use Case** | **Script Edit** |
| **Actors** | **User, System** |
| **Type** | **Primary, offstage** |
| **Description** | **The user is given option to edit the script if they do not like it or want to remove something from it for the general public.** |

## 8.2. Extended Use Cases:

### 1.Registration

|  |  |  |
| --- | --- | --- |
| **Use Case Name** | **Register an Account** | |
| **Scope** | **Talk Tales** | |
| **Level** | **User’s goal** | |
| **Primary Actor** | **User** | |
| **Stakeholders and**  **Interests** | **User** | |
| **Preconditions** | **The user does not have an account.** | |
| **Success Guarantee** | **The user will be able to successfully register his/her account.** | |
| **Main Success**  **Scenario** | **User Actions**  **1: The user opens the website and clicks the register button.**  **3: The user enters his name, contact number, username,** | **System Response**  **2: The system asks for some details from the user.**  **4: The system successfully registers the user.** |
|  | **password, email, etc., and submits it.** |  |
| **Extensions** | **If the password is less than 8 characters then an error message will occur.** | |

### 2.Avatar Selection

|  |  |  |
| --- | --- | --- |
| **Use Case Name** | **Avatar Selection** | |
| **Scope** | **Talk Tales** | |
| **Level** | **User’s goal** | |
| **Primary Actor** | **User** | |
| **Stakeholders and**  **Interests** | **User** | |
| **Preconditions** | **The user is already logged in.** | |
| **Success Guarantee** | **The user will be able to successfully Select their Avatar.** | |
| **Main Success**  **Scenario** | **User Actions**  **1: The user opens the website and is shown options of Avatars**  **3: The user asks for the selected Avatar to be used in the video.** | **System Response**  **2: Asks User to select an Avatar**  **4: The system successfully generates the avatar ready to be used in Video.** |
| **Extensions** | **If there is an error in generating the avatar, the user is alerted with a pop up message.** | |

### 3.News Selection

|  |  |  |
| --- | --- | --- |
| **Use Case Name** | **News Selection** | |
| **Scope** | **Talk Tales** | |
| **Level** | **User’s goal** | |
| **Primary Actor** | **User** | |
| **Stakeholders and**  **Interests** | **User** | |
| **Preconditions** | **The user is already logged in.** | |
| **Success Guarantee** | **The user will be able to successfully select the news they want to be delivered.** | |
| **Main Success**  **Scenario** | **User Actions**  **1: The user opens the website and is shown options of News.**  **3: The user selects the options and submits them to the system.** | **2: Asks User to search news they want through a search bar.**  **4: The system successfully accepts the news selection information and will use it while scraping the news.** |
| **Extensions** | **If the user does not select all the required information, an alert message is shown to please select all the required field options.** | |

### 4.Background Selection

|  |  |  |
| --- | --- | --- |
| **Use Case Name** | **Background Selection** | |
| **Scope** | **Talk Tales** | |
| **Level** | **User’s goal** | |
| **Primary Actor** | **User** | |
| **Stakeholders and**  **Interests** | **User** | |
| **Preconditions** | **The user is already logged in.** | |
| **Success Guarantee** | **The user will be able to successfully be able to select a background.** | |
| **Main Success**  **Scenario** | **User Actions**  **1: The user opens the website and is shown options of Backgrounds.**  **3: The user asks for the selected Background to be used in the video.** | **System Actions**  **2: Asks User to select a Background.**  **4: The system successfully accepts the background ready to be used in Video.** |
| **Extensions** | **The user can only select one background.** | |

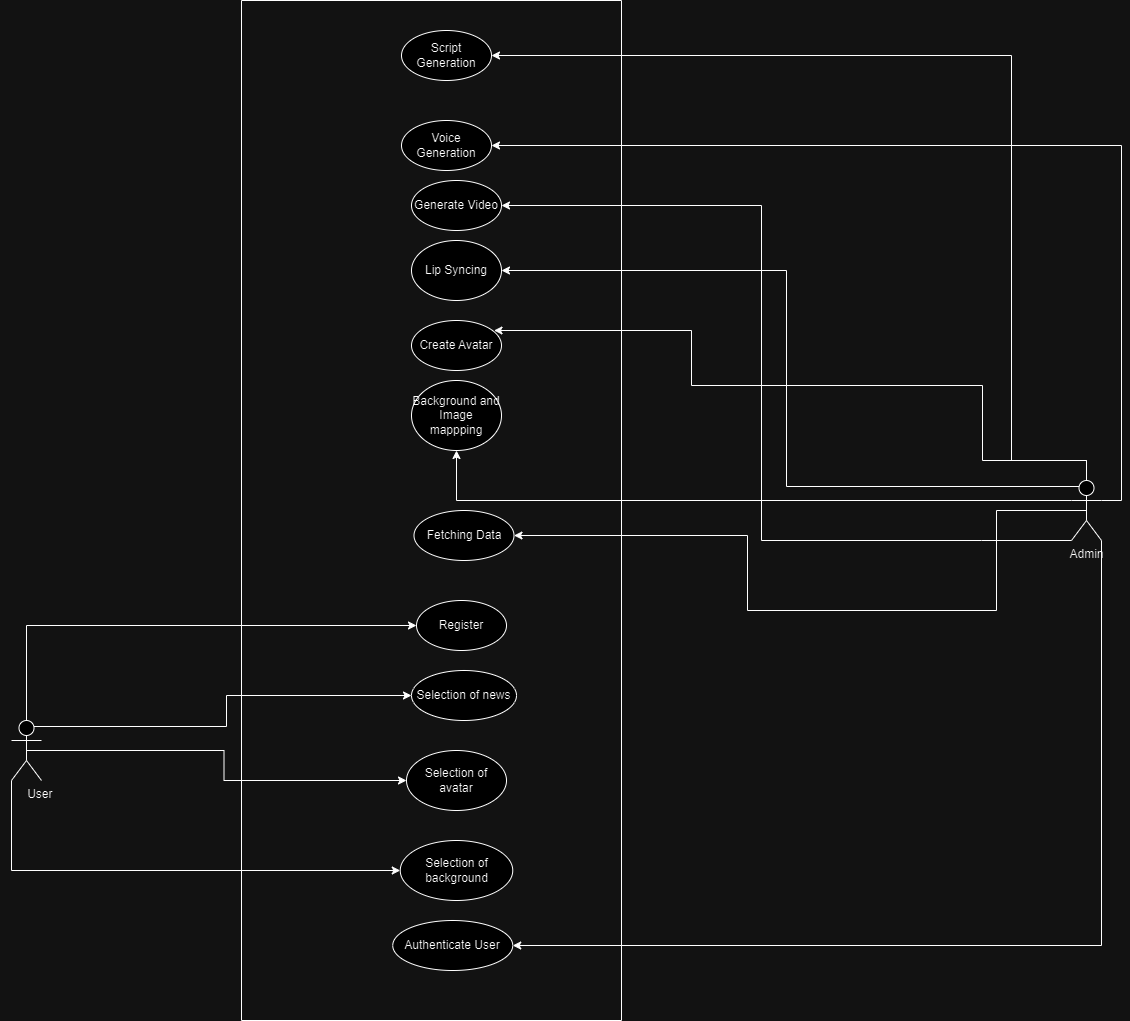
### 5.Language Selection

|  |  |  |
| --- | --- | --- |
| **Use Case Name** | **Voice Selection** | |
| **Scope** | **Talk Tales** | |
| **Level** | **User’s goal** | |
| **Primary Actor** | **User** | |
| **Stakeholders and**  **Interests** | **User** | |
| **Preconditions** | **The user is already logged in.** | |
| **Success Guarantee** | **The user will be able to successfully select a language used for AI voice.** | |
| **Main Success**  **Scenario** | **User Actions**  **1: The user opens the website and is shown options of voice.**  **3: The user asks for the selected language to be used in the video.** | **System Response**  **2: Asks User to select any languages available in the system.**  **4: The system successfully accepts the language ready to be used in Video.** |
| **Extensions** | **The user can only select one voice.** | |

### 6.Script Edit

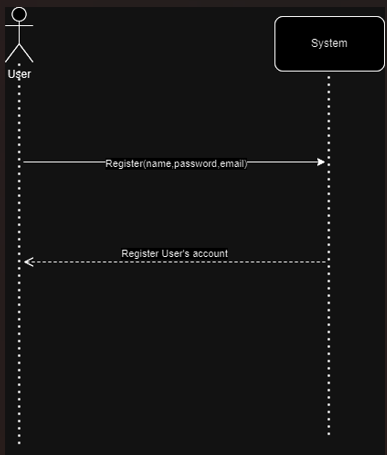
|  |  |  |
| --- | --- | --- |
| **Use Case Name** | **Script Edit** | |
| **Scope** | **Talk Tales** | |
| **Level** | **User’s goal** | |
| **Primary Actor** | **User** | |
| **Stakeholders and**  **Interests** | **User** | |
| **Preconditions** | **The user is already logged in.**  **The user has requested for news and a script has been generated.** | |
| **Success Guarantee** | **The user will be able to successfully edit the news script according to their liking.** | |
| **Main Success**  **Scenario** | **User Actions**  **1: The user opens the website and is shown with the generated script using AI.**  **3: The User selects the option and edits the script according to their liking.** | **System Response**  **2: The system displays the option to edit the script.**  **4: The system successfully accepts the request and uses the edited script of the user to be used in the video.** |
| **Extensions** |  | |

# 9. Use Case Diagram:

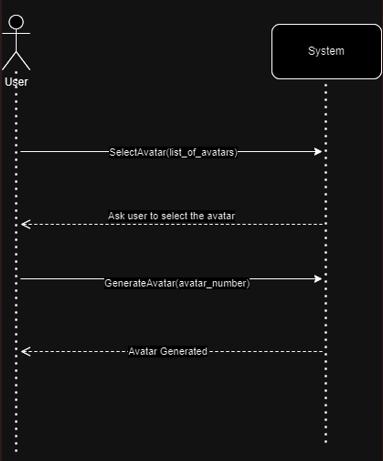
****

# 10. System Sequence Diagram:

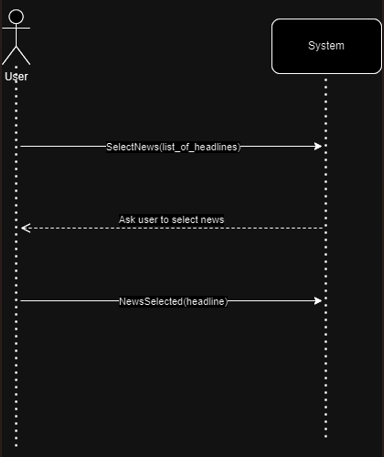
## 1.Registration

****

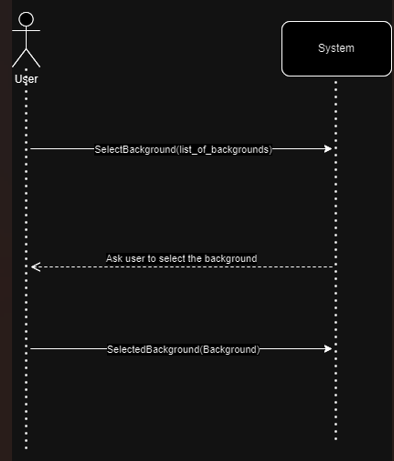
## 2.Avatar Selection

****

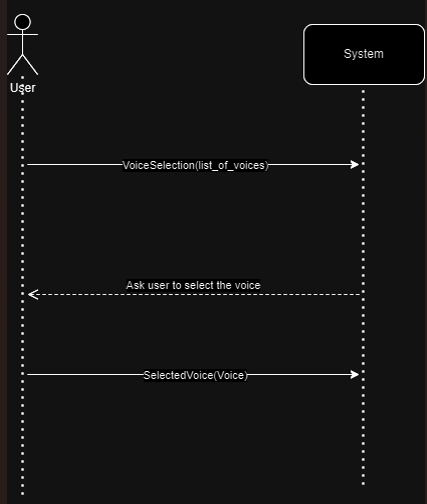
## 3.News Selection

****

## 5.Background Selection

****

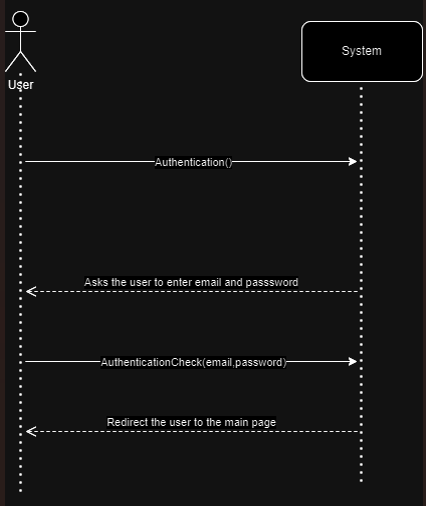
## 6.Voice Selection

****

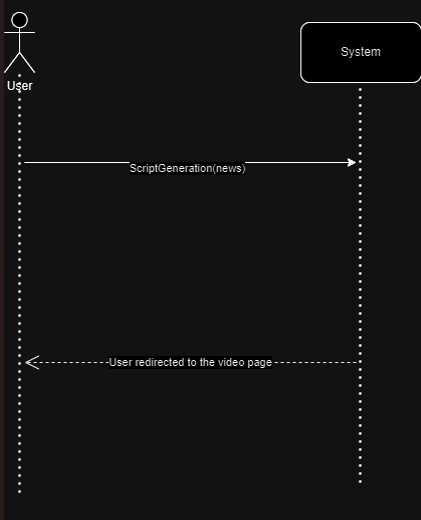
## 7.Voice Generation

****

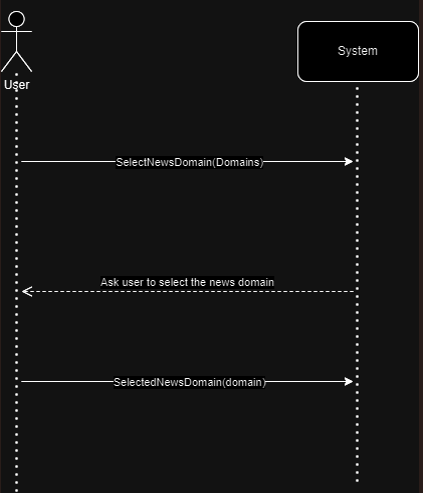
## 8.Authentication

****

## 9.Script Generation

****

## 10.News Domain Selection

****

# 11. Class Diagram:

# 12. JSON Schema:

{

"$schema": "http://json-schema.org/draft-07/schema#",

"definitions": {

"User": {

"type": "object",

"properties": {

"userID": { "type": "string" },

"email": { "type": "string" },

"password": { "type": "string" }

},

"required": ["userID", "email", "password"]

},

"NewsArticle": {

"type": "object",

"properties": {

"articleID": { "type": "string" },

"title": { "type": "string" },

"content": { "type": "string" },

"source": { "type": "string" },

"domain": { "type": "string" },

"link": { "type": "string" }

},

"required": ["articleID", "title", "content", "source", "domain",”link”]

},

},

"type": "object",

"properties": {

"user": { "$ref": "#/definitions/User" },

"newsArticle": { "$ref": "#/definitions/NewsArticle" },

},

"required": ["user", "newsArticle"]

}

# 13. Activity Diagram:

# 

# 14. User Manual:

**Welcome to Talk Tales!**

Talk Tales is a web-based application that allows you to create a virtual 2D environment where based on user preferences selected news is delivered in an innovative way using the power of AI. With just a few simple steps, you can easily get delivered news specifically related to your topics of interest.

**Get Started:**

* Visit the Talk Takes website and create an account if you do not have one already and then successfully log in.
* Once you are logged in, you will be redirected to a new page where you will see a lot of options to choose from.
* Firstly, the user is asked to select news of their liking. Users can find a search bar where they input any phrase which will help them find news related to that phrase.
* Links will be fetched related to the phrase searched for previously. Select a link of your choice and script will be generated based on it.
* Now you will be redirected to a new page where you can see the script generated. If you do not like the script you can easily edit be clicking on the edit button.
* After news selection, different sets of avatars are shown from which users can choose the one they like who will be delivering the news in the virtual 2D environment.
* Then you can also select the language you want your content to be delivered in.
* Then the user is asked to select a background of their liking. Simple, the user chooses one and hence all the steps are completed.
* Website will be redirected to a new page where a short 1-minute-long video will be generated which the user can play to watch the news that our AI will deliver to them.

**Tips for getting the Best Experience:**

* Use a stable and fast internet with at least a download speed of 8Mbps.
* Choose a background which is clear and does not make things look messy while news is being delivered.
* Choose appropriate avatars with appropriate voice so that the avatar while delivering news looks realistic.

## Frequently Asked Question (FAQS):

Q. Can the user have themselves as the avatar?

Ans: No, the user can have themselves as avatars. The user can only choose from the list of available avatars on the website.

Q. Can you download the generated video?

Ans: No, unfortunately you can not download the video till now.

Q. Can we upload any sort of background for the video?

Ans: No, the user can not upload any sort of background for their video. They can only choose from the list of available backgrounds on the website.

Q. Can I edit my video after it has been generated?

Ans: No, unfortunately once the video is created you can not edit it but you can create a new one which may be well more suited to your liking.

Q. Can I use my video clip for commercial purposes?

Ans: Yes, you can use your video clip for commercial purposes. every generated video is saved in our database so you can provide a link which redirect people to our website containing your video.