

UE20CS390A - CAPSTONE PROJECT REVIEW 3

High Level Design and Proposed Methodology

“News on the Go” - A Video Summarization, Text Summarization and Translation Service



Project ID : 72

Project Guide : Prof. Shilpa S

Project Team :

Rohit V Shastry	PES2UG20CS282
Rushil Ranjan	PES2UG20CS288
Sai Hardik Sriram Talluru	PES2UG20CS296
Niranjan Rao SS	PES2UG20CS226

Outline

- Abstract
- Summary of Literature Survey
- Suggestions from Review 2
- Design Details
- Proposed Methodology
- Architecture
- Design Description
- Technologies Used
- Project Progress
- Capstone Project (Phase 1 & 2) Timeline
- Conclusion
- References

Problem Statement

- Consuming news can be time-consuming and challenging for individuals who are short on time, non-native speakers, or those who prefer to consume news in their native language.
- Current news summarization tools are limited in their capabilities, and most do not provide translation features.
- This project aims to address these limitations by developing a software tool that can automatically summarize news videos into shorter video summaries, transcribe them into text, and generate a summary of the text, while also providing translation capabilities.

Introduction

- "News on the Go" is a platform that offers users access to news videos from various news channels.
- The platform includes several features such as video summarization, text summarization, translation, and recommendation systems to provide users with personalized news content.
- The platform also keeps track of a user's search and viewing history to provide a more tailored news experience, with the ultimate goal of providing users with a convenient and efficient way to access news content from various sources in a personalized and user-friendly way.

Transcription and Tranlation

Paper Details	Objective of paper, Techniques/Methods	Advantages	Limitations
"Robust Speech Recognition via Large-Scale Weak Supervision" by Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. OpenAI - 2022	Proposes a new approach to training speech recognition models that relies on weak supervision and large-scale data to improve model performance and scalability.	The approach relies on self-supervised learning, which allows the model to learn from unlabeled data and improve its performance without relying on large amounts of labeled data.	The proposed approach still requires a significant amount of manual effort in creating the weak supervision data, although it is less than traditional approaches.

Text Summarization

Paper Details	Objective of paper, Techniques/Methods	Advantages	Limitations
"Abstractive Document Summarization with a Graph-Based Attentional Neural Model" by Zhihong Shao, Furu Wei, Sujian Li, and Xiaodong Liu (COLING 2020)	The paper proposes a graph-based attentional neural model for abstractive document summarization, which involves graph convolutional networks, sentence-level attention, and a coverage mechanism.	This paper proposes a graph-based attentional neural model that achieves state-of-the-art results on multiple datasets, demonstrating its effectiveness and generalizability for abstractive document summarization.	The limitation of this paper is that the evaluation is only conducted on small datasets and may not fully reflect the model's performance on larger datasets or real-world applications.
Unsupervised Extractive Text Summarization with Distance-Augmented Sentence Graphs - 44th ACM SIGIR CONFERENCE	Proposes an unsupervised approach for extractive text summarization, called sentence graph-based summarization, that leverages automatically constructed sentence graphs to select salient sentences for summarization.	Both single- and multi-document unsupervised methods either achieves or is comparable to state-of-the art unsupervised methods, and both methods are competitive with strong supervised baselines.	While the proposed approach achieves competitive results, it may not perform as well as supervised methods that have access to large amounts of human-generated summaries.

Video Summarization

Paper Details	Objective of paper, Techniques/Methods	Advantages	Limitations
<p>Leonardo Moraes, Ricardo Marcondes Marcacini, and Rudinei Goularte. 2022. Video Summarization using Text Subjectivity Classification.</p> <p>ACM - NOV, 2022</p>	<p>The objective of this paper is to propose a video summarization method that uses text subjectivity classification to extract subjective and important information from the video.</p>	<p>The proposed approach in this paper can effectively summarize videos by combining text subjectivity classification and video content analysis, and outperforms several baseline methods.</p>	<p>Cannot retrieve objective segments in the video. This would be counter-productive to this project since news videos are all about facts and figures more than opinions.</p>
<p>Unsupervised Video Summarization via Multi-source Features</p> <p>ACM , 2021</p>	<p>Multi-Source Chunk and Stride Fusion (MCSF) model is used. There were three fusion techniques used: 1. Early Fusion. 2. Intermediate Fusion. 3. Late Fusion</p>	<p>Results in more accurate and representative video summaries. Its also Scalable.</p>	<p>Cannot work well with complex videos that have multiple subplots or scenes with different visual or audio characteristics.</p>

Recommendation

Paper Details	Objective of paper, Techniques/Methods	Advantages	Limitations
Jizhou Huang, Haifeng Wang, Wei Zhang, and Ting Liu. 2020. Multi-Task Learning for Entity Recommendation and Document Ranking in Web Search. ACM Trans. Intell. Syst. Technol. 11, 5, Article 54 (July 2020)	The main objective of the paper is to propose a multi-task learning approach that jointly learns entity recommendation and document ranking tasks to improve the accuracy and efficiency of web search.	The use of multi-task learning framework improves the effectiveness and efficiency of entity recommendation and document ranking in web search.	Doesn't address the cold-start personalization problem for new users with no past search history.

Summary

- Supervised methods can be more effective when there is a large amount of high-quality labeled data available, as they can learn directly from this data to generate accurate and high-quality summaries.
- Unsupervised methods can be more practical and accessible when labeled data is scarce or not available, as they can learn from the inherent structures and patterns in the data without relying on external labels. However, unsupervised methods may not always produce summaries that are as accurate or comprehensive as those generated by supervised methods, and may require more effort in fine-tuning and optimization to achieve good performance.
- The use of additional features, such as text subjectivity classification in video summarization, can improve the quality of summaries generated by the model.
- Abstractive summarization methods have the potential to generate more human-like summaries, but are generally more challenging to implement and evaluate than extractive methods.

Suggestions from Review 2

- Go ahead with formulating a High Level Design
- Explore new methods of collecting data.

Design Details

Our project depends on multiple platforms, systems, and processes to function effectively.

- Firstly, it relies on YouTube's API to retrieve news videos from different news channels.
- Secondly, it uses natural language processing (NLP) and machine learning (ML) models to perform video summarization, text summarization, translation, and recommendation systems.
- These models require a powerful computing infrastructure to run efficiently.
- In terms of any necessary changes, it is important to ensure that the app and its models are regularly updated to keep up with changing technologies and user needs.
- The app should also be tested thoroughly before deployment to ensure that it is reliable and secure.
Finally, the app should be deployed on a scalable and reliable infrastructure to ensure that it can handle increased traffic and demand as the user base grows.

Design Details

Innovativeness - The project's innovativeness lies in the integration of multiple cutting-edge technologies, including video summarization, text summarization, translation, and recommendation systems, to provide a personalized and user-friendly news experience.

Security - The project should ensure that all user data is encrypted and stored securely, and implement secure authentication and authorization mechanisms to prevent unauthorized access to the system.

Interoperability - The project has a high degree of interoperability, allowing it to seamlessly interact with a variety of external systems and platforms. This is achieved through the use of standard protocols and technologies which enable the exchange of data between different systems in a standardized and platform-independent manner.

Reusability - The modular design of the project allows for the individual components, such as video and text summarization, to be reused in other applications.

Design Details

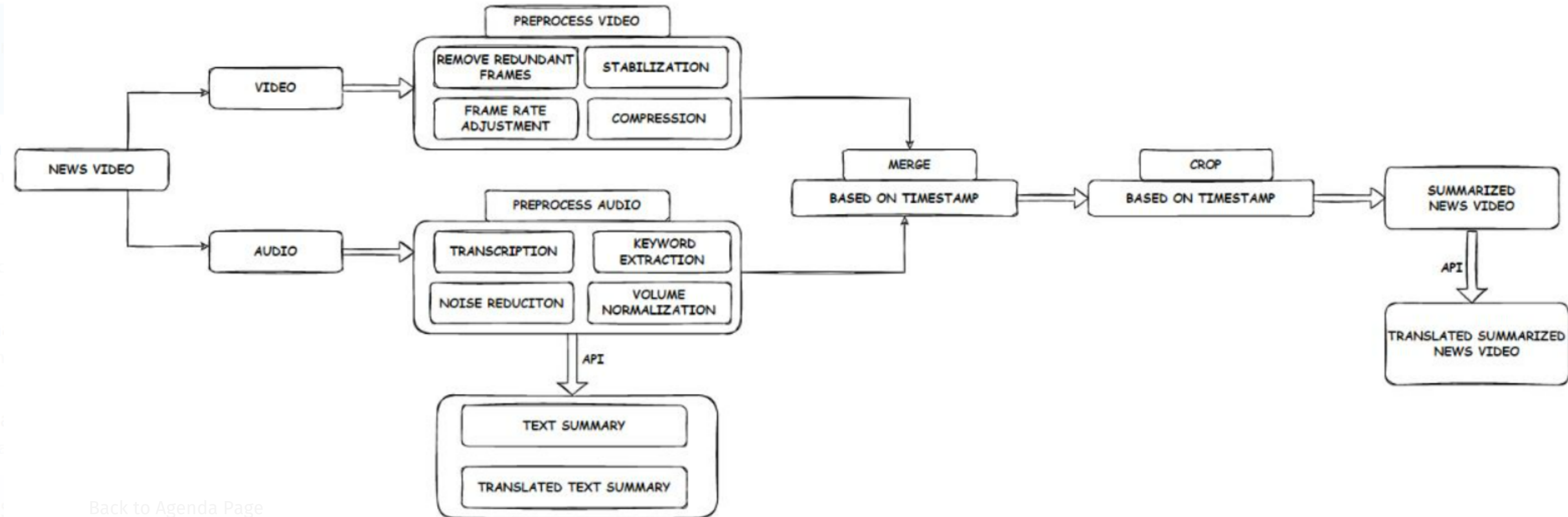
Portability - The News on the Go app is highly portable, accessible from any hand-held device with internet connection and suitable for users on the move.

Reliability - The app will be reliable with robust hosting infrastructure and proper maintenance procedures being implemented frequently.

Legacy to Modernization - As the app is being developed with modern technologies and design patterns, it should not have any significant legacy issues and should be relatively easy to maintain and upgrade in the future.

Architecture

High Level Design View



High Level Design View

The figure provides an overview of the high-level design. The raw, unprocessed news video is first fed into the model and is then split into separate audio and video pipelines.

Video Pipeline - The video pipeline processes the video by compressing it, stabilizing it, adjusting the frame rates, and removing redundant frames.

Audio Pipeline - The audio pipeline normalizes the audio, transcribes it and extracts important keywords and phrases.

- In the video pipeline, the key frames are obtained from the from the timestamps using the important keywords obtained during preprocessing, and the frames are labelled with numbers.
- In the audio pipeline, the news video is converted to an mp3 file and then processed by the summarization module.
- The module generates chapters and highlights containing the summary and time-intervals of the key topics.

High Level Design View

- Some of the time intervals are extracted from the chapters and highlights and merged with the video timestamps.

Text Summarization - The transcripts obtained from the video will be given to the text summarization model. First, the text will be tokenized or broken down into words. The tokenizer then encodes the text as a sequence of numbers that can be input into the transformer.

- The transformer has an encoder and a decoder. The encoder captures the important part of the text and the decoder generates a summary out of this.
- The decoder uses an attention mechanism to focus on the most relevant parts of the text.

Transcription and Translation will be performed with the use of an API.

Architecture

Additional Points

Data Collection - Data (videos) will be obtained with the help of a Web Crawler that will be designed by us.

Functions of the Web Crawler - Runs continuously on a remote server and waits for the News channel to upload a video.

Immediately triggers an API request to the Youtube Data API to retrieve the video and its meta-data and stores them in a cloud based relational database, ready for us to pre-process and use.

Architecture

Logical User Groups

Logical User Groups can be widely classified into two - **Viewers** and **Administrators**.

Viewers -

- **Casual users** - are interested in staying up-to-date with the latest news and prefer a user-friendly interface.
- **News enthusiasts** - want to customize their news feed based on their interests.
- **Researchers** - want to access a wide range of news articles from various sources.
- **Journalists** - want to monitor specific topics or keywords and access news articles in real-time.
- **Business professionals** - need to stay updated with the latest industry news and trends.
- **Students** - want to access news articles related to their academic interests.
- **Travelers** - want to stay updated on current events in the regions they are traveling to.

Architecture

Logical User Groups

Logical User Groups can be widely classified into two - **Viewers** and **Administrators**.

Administrators - Administrators are responsible for maintaining the system and ensuring that it is running smoothly. They will monitor system performance and usage patterns, and handle technical issues as they arise. They may also be responsible for managing user accounts and permissions, and ensuring that the app is in compliance with any relevant regulations or policies.

Application Components

- Video retrieval module: This component retrieves the news videos and stores them in the database.
- Video processing module: This component processes the videos to generate video summaries, transcriptions, and visualizations. It also performs text extraction from the transcriptions.
- Video storage module: This component stores the processed videos, transcriptions, and visualizations in the database.
- User interface module: This component provides the user interface for the users to search for and view the processed news videos, summaries, and visualizations.
- Search engine module: This component provides the search functionality to the users to search for videos using various search criteria such as keywords, channel names, and names of people.

Application Components

- Translation module: This component translates the video summaries, transcriptions, and Text Summaries to various languages based on the user's preferences.
- Recommendation module: This component recommends videos to the users based on their search history, preferences, and user behavior.
- Authentication and authorization module: This component handles the authentication and authorization of users and ensures that only authorized users have access to the application.
- Notification module: This component sends notifications to users about new videos, updates, and recommendations.
- Database module: This component stores and manages all the data required by the application, including user profiles, video metadata, processed videos, summaries, and transcripts.

Data Components

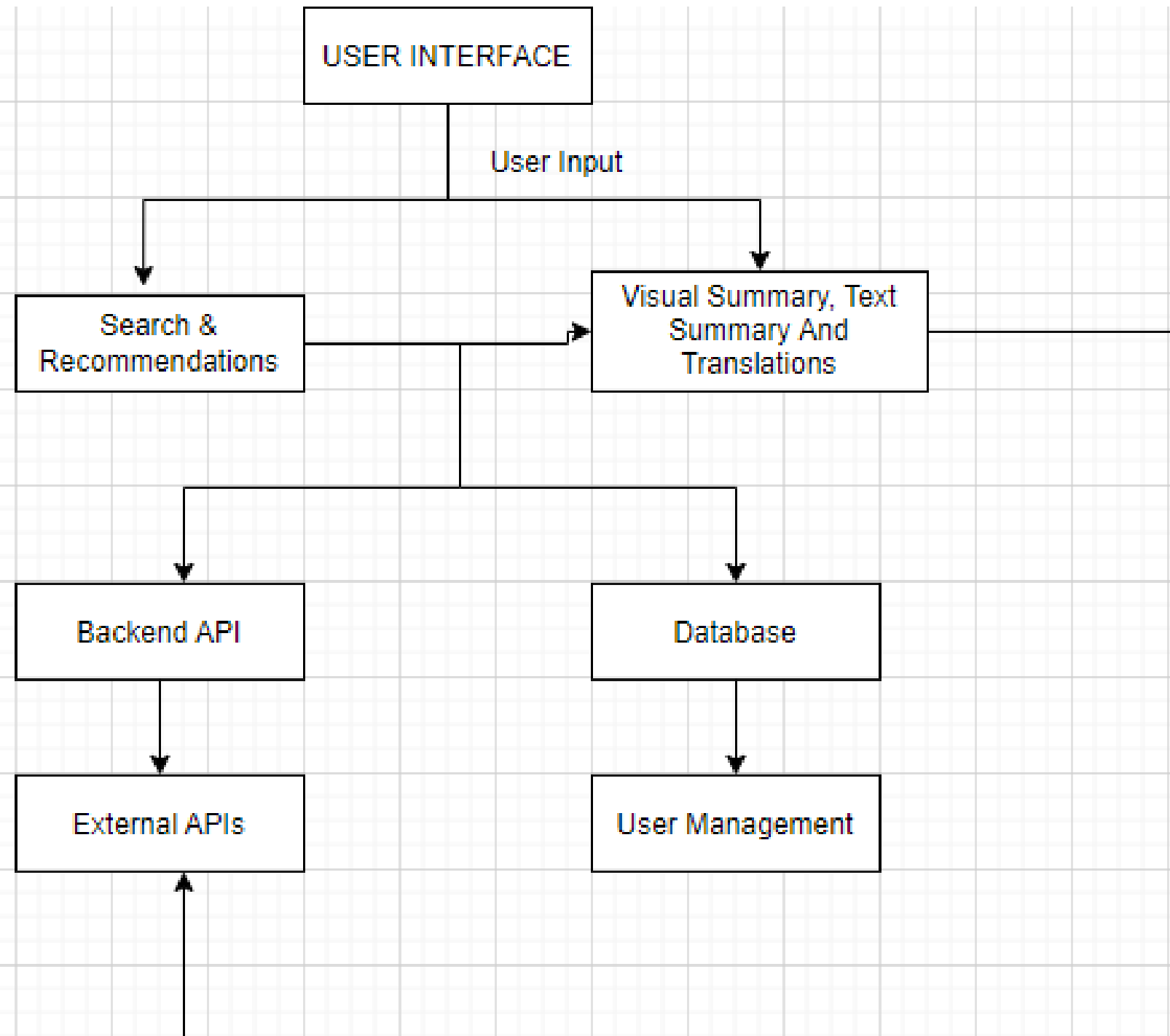
- News Video Data: This includes the news videos that are added to the system, along with their metadata such as the title, description, channel name, upload date, and tags.
- User Data: This includes the data related to the user's account, such as their watch history, search history, and preferences. This data is to provide personalized recommendations to the user.
- Summary Data: This includes the text and visual summaries generated for each news video, along with their translations in different languages.
- Metadata: This includes the metadata associated with each news video, such as the duration, resolution, and format of the video.
- Logs and Analytics Data: This includes the data related to system logs, user analytics, and performance metrics. This data is used to monitor the system's performance, identify issues and improve the overall user experience.

Interfacing Systems

- News video sources: This system will provide the news videos to the application. It can be interfaced with through APIs or web scraping tools.
- Translation service: This system will provide the translation of the summary and visual summary into different languages. It can be interfaced with through APIs.
- Search engine: This system will be used to retrieve videos based on user search queries. It can be interfaced with through APIs.
- Notification service: This system will send push notifications to the user when a new summarized video is available. It can be interfaced with through APIs.
- Analytics service: This system will track user interactions and generate insights for the application. It can be interfaced with through APIs or SDKs.

Architecture

Interaction Between Major Components



Proposed Methodology

The Basic approach is to implement Keyframe Extraction and Shot Detection to retrieve key frames and frames with images and factual data shown. This approach captures the Objective data presented in a video. Using this data, the video will be Summarized.

The model will make use of the transcripts to trim the video based on timestamps and merge them into a visual summary.

Test Summarization will be performed using NLP techniques and BERT which is a pre-trained model. BERT will be used mainly to perform Abstractive Text Summarization which achieves more human-like summary.

The summaries will be translated into multiple languages with the help of an API.

Proposed Methodology

Is there a need for changing the approach ?

- There is a need for changing the approach of this project. The necessary changes will be made to the video summarizing model and the text summarizing model.
- The Video summarizing model will be summarizing based on key frames and the transcripts generated from the video.
- The previous approaches only consider the objective data in a video and neglect subjective data or vice- versa. There is a need to implement a model that takes both types of data into consideration while summarizing a video.
- Therefore, we will implement a model that uses key-frame extraction, object detection and text analysis to summarize a news video hence covering both objective and subjective data in our summaries.

Proposed Methodology

Is there a need for changing the approach ?

- Text summarization will also require the use of an attention mechanism to effectively pick out keywords and phrases. BERT will use this data to create abstractive summaries.
- Translation will be implemented using an API trained on Hours of audio data and text.

Benefits and Drawbacks of using this approach

Benefits -

- It is a well rounded approach to summarizing videos.
- Use of BERT makes the text summarization model highly efficient.

Proposed Methodology

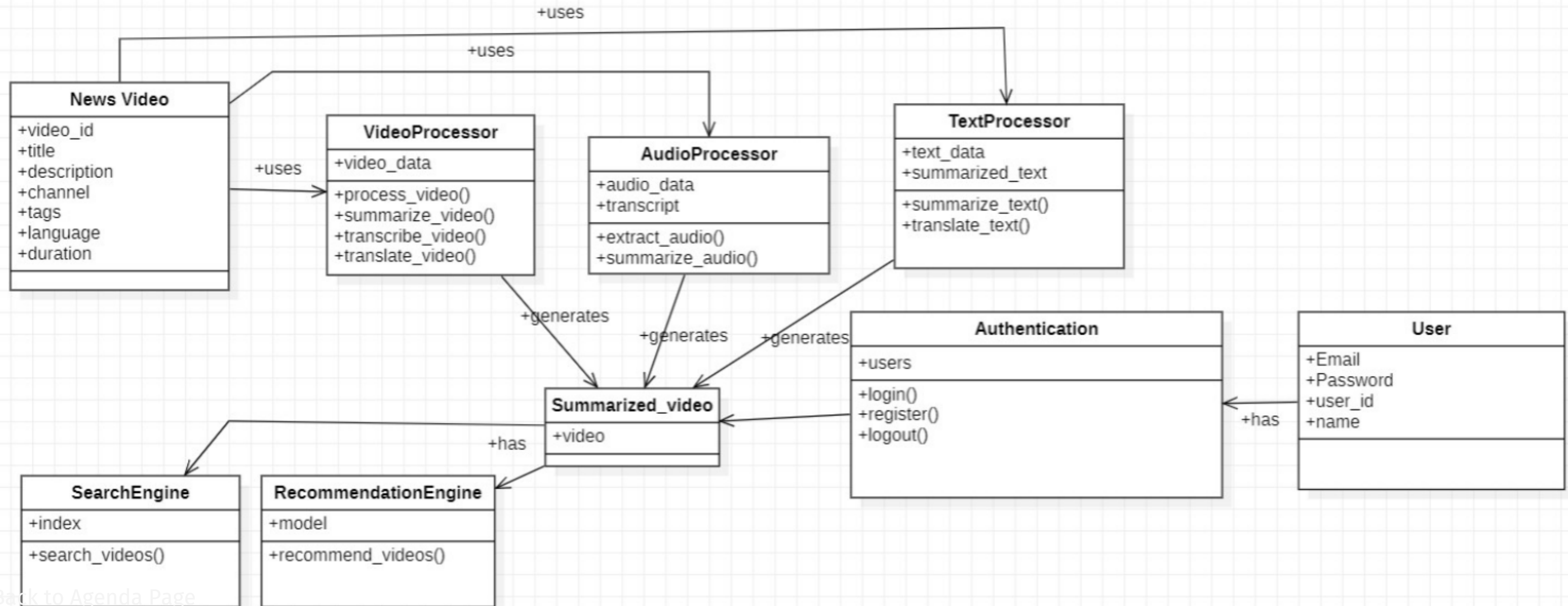
Benefits and Drawbacks of using this approach

Drawbacks -

- Text summarization may oversimplify complex news stories, potentially leading to a loss of important details or key points.
- Transcription and translation accuracy may vary, particularly for languages with complex grammar or nuances.

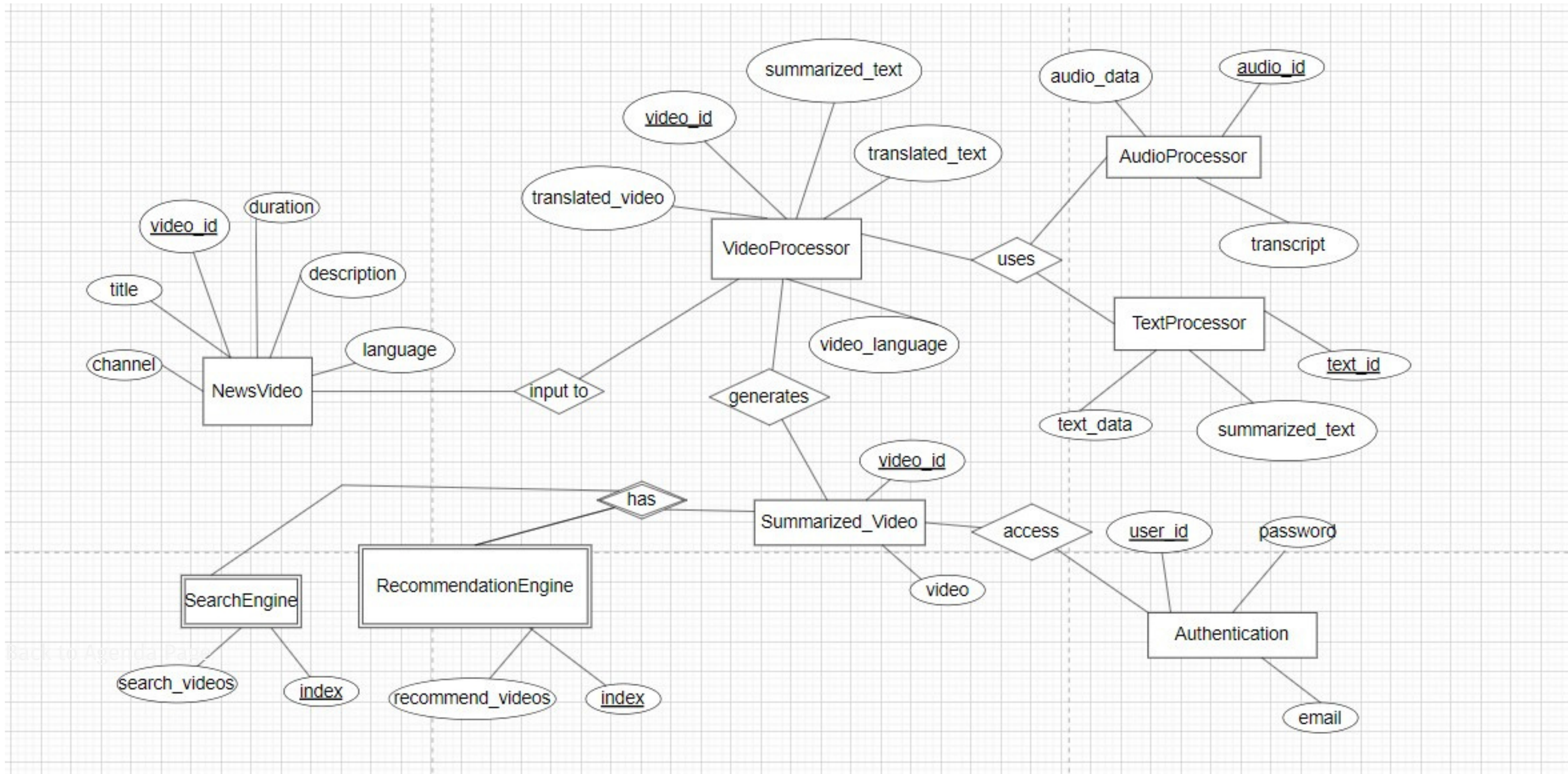
Design Description

Master Class Diagram



Design Description

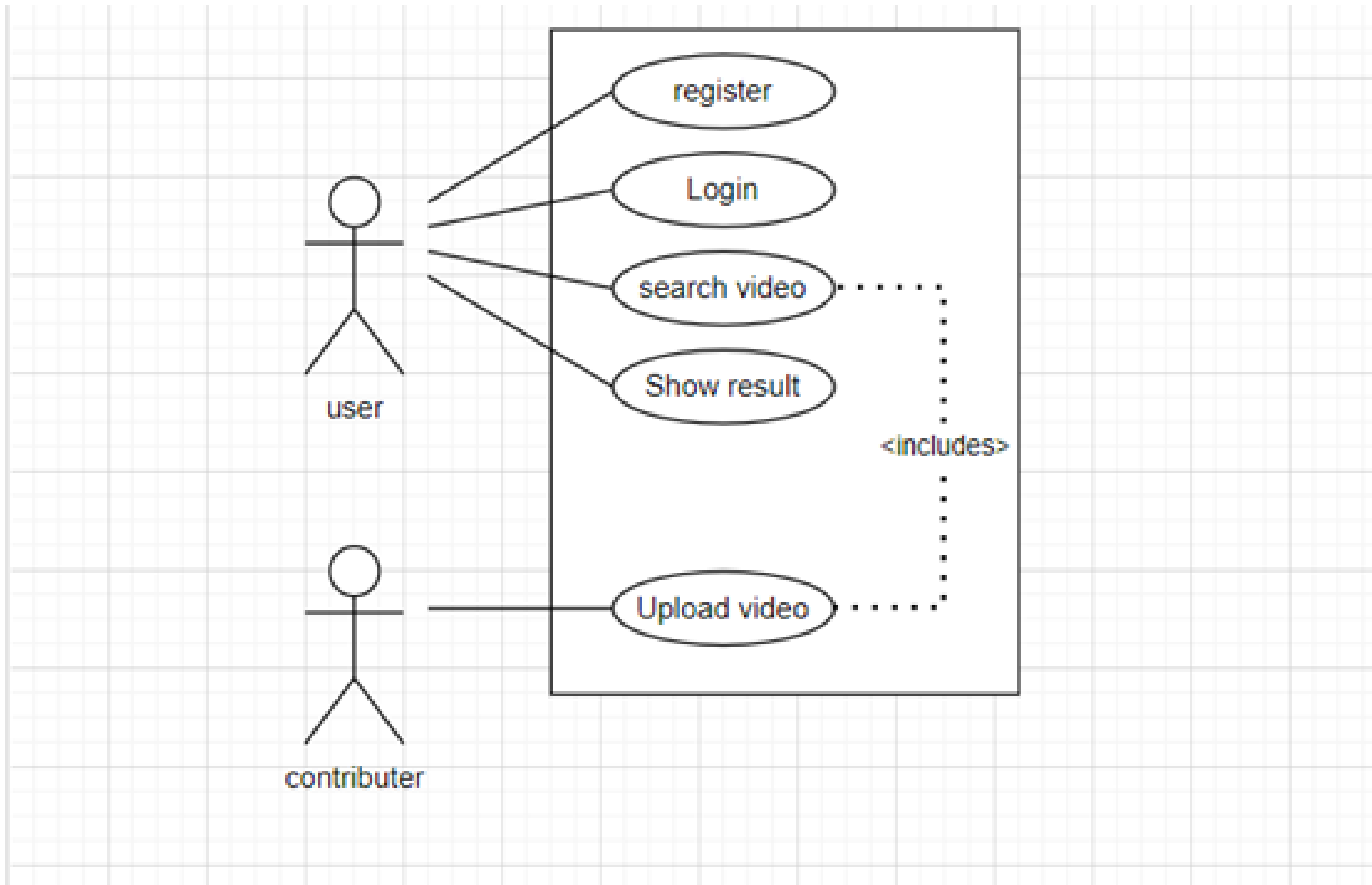
ER Diagram



Design Description

Use Case Diagram

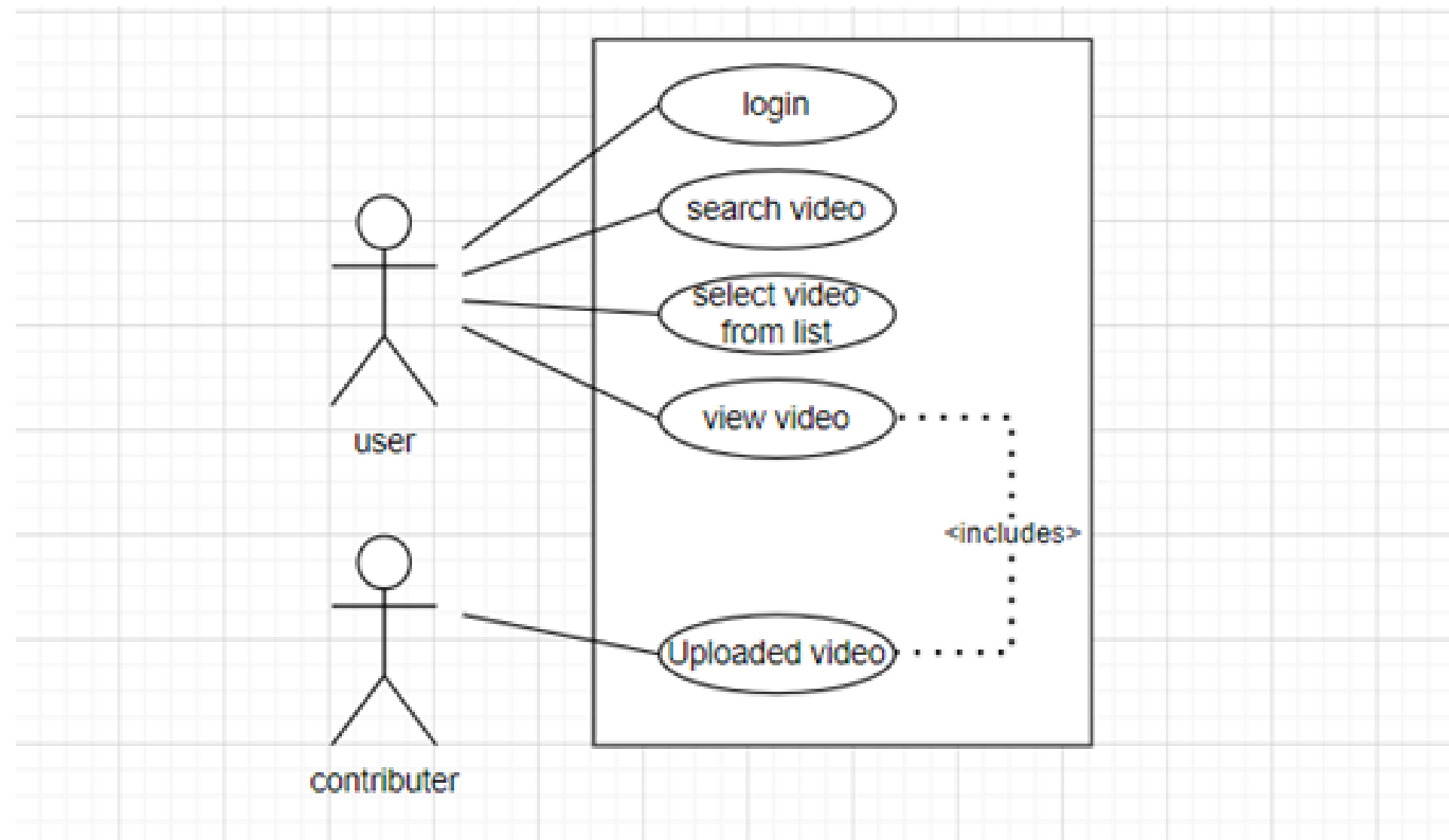
1. Search Use Case



Design Description

Use Case Diagram

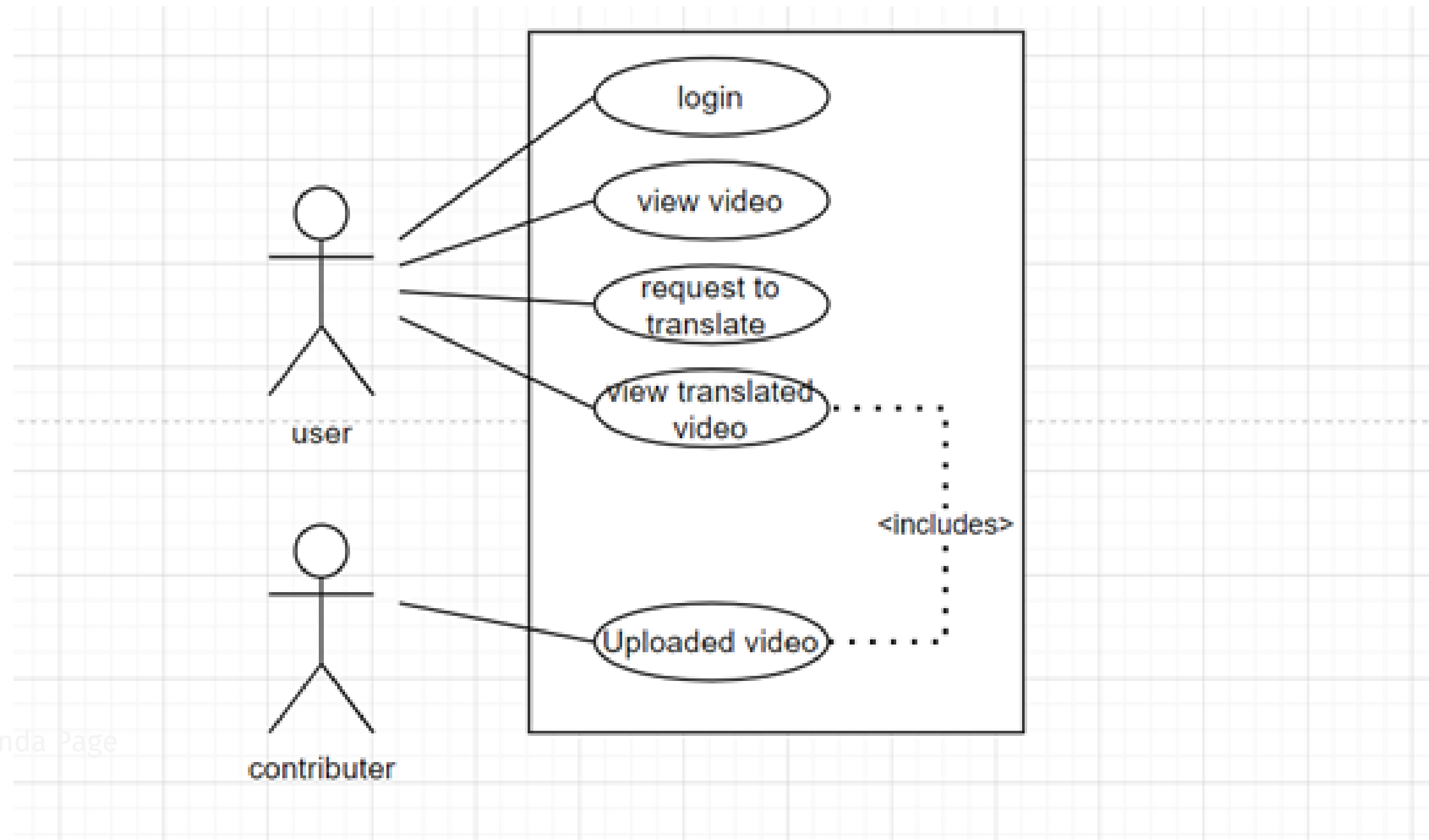
2. View Video



Design Description

Use Case Diagram

3. Translate Video



Design Description

Report Layouts

News Summary Page:

Title: News Summary Page

Introduction: The News Summary Page is the main landing page for the news application, displaying a list of news videos with their respective video summaries, text summaries, and thumbnail images.

Body: The list of news videos should be sorted by the most recent videos first and should include a pagination feature to allow users to view more videos. Each item on the list should display the source of the news video, the date it was published, and the duration of the video. The video summary should provide a brief overview of the news story and its significance, while the text summary should provide a more detailed explanation of the story's content.

[Back to Agenda Page](#)
Conclusion: The News Summary Page provides users with a quick and easy way to browse and select news videos based on their interests and preferences.

Design Description



Report Layouts

Video Player Page

Title: Video Player Page

Introduction: The Video Player Page is displayed when a user clicks on a particular news video from the News Summary Page, and it displays the video player for the selected news video along with its accompanying video summary and text summary.

Body: The user should be able to select the language for translation of the text summary, and the video player should include standard controls for playing, pausing, rewinding, and fast-forwarding the video.

Conclusion: The Video Player Page provides users with a comprehensive view of the selected news video, including video and text summaries, translation options, and standard video player controls.

[Back to Agenda Page](#)

Design Description

Report Layouts

Search Page

Title: Search Page

Introduction: The Search Page allows users to search for news videos by entering keywords, channel names, or people's names.

Body: The search results should display a list of news videos that match the search criteria, along with their respective video summaries, text summaries, and thumbnail images. The list should be sorted by relevance and should include a pagination feature to allow users to view more search results. Each item on the list should display the source of the news video, the date it was published, and the duration of the video.

Conclusion: The Search Page provides users with a powerful tool for finding news videos related to specific topics, people, or channels.

Design Description

Report Layouts

Settings Page

Title: Settings Page

Introduction: The Settings Page allows users to customize their preferences and settings for the news application.

Body: Users should be able to set their preferred language for video and text summaries, as well as their preferred translation languages. The page should also include options for enabling or disabling auto-translation, as well as the ability to set notifications for new news videos or updates to existing news videos. The settings page should be easily accessible from any page in the application.

Conclusion: The Settings Page provides users with the ability to tailor their news viewing experience to their specific needs and preferences.

[Back to Agenda Page](#)

Design Description



Report Layouts

Profile Page

Title: Profile Page

Introduction: The Profile Page displays the user's profile information and allows them to modify their settings and preferences.

Body: Users should be able to view their history of watched news videos and update their profile picture and email address. The profile page should be easily accessible from any page in the application and should include a logout button for users to log out of their account.

Conclusion: The Profile Page provides users with a convenient and personalized hub for managing their news viewing history and preferences.

[Back to Agenda Page](#)

External Interfaces

- External interfaces are the connections between the system and the outside world, including other systems and devices. In the case of this news app, there are several external interfaces to consider.
- YouTube API: This is the interface that the system uses to access new videos from subscribed channels on YouTube.
- Translation API: This is the interface that the system uses to translate the summaries of news videos into different languages.
- Database API: This is the interface that the system uses to store and retrieve data from the database.
- User interface: This is the interface that the users use to interact with the app.
- Push notifications: This is the interface that the system uses to send push notifications to users about new videos, personalized recommendations, or other app-related updates.

Technologies Used

Primary Programming Language - Python

- It is widely used in machine learning and natural language processing. It has a large number of libraries and tools that can be used for data analysis, machine learning, web scraping, etc.

Backend Server - Python Language will be used for the backend.

- Python has a large standard library that includes many modules for app development, database connectivity, and more. Reduces the need for 3rd Party libraries.
- Allows for Rapid Prototyping and Development.
- Scalability: Python is capable of handling large amounts of traffic and data, making it a good choice for developing scalable backend systems.

Technologies Used

App Development Framework - FLUTTER will be used to develop the app.

- It allows us to build high-performance, visually appealing apps for both Android and iOS platforms using a single codebase.

Database Management Systems

- We will be using Google Cloud SQL since it simplifies the implementation of your system, as we do not have to worry about hardware setup, software installation, or system maintenance.
- Google Cloud SQL many advantages, such as scalability, high availability, and ease of maintenance. These services also have built-in security features that can help protect user data.

Technologies Used

Cloud Services - Google Cloud will be used to host the application, store data, which would allow for scalability and cost-effectiveness.

Video Processing and Summarization

- **OpenCV:** OpenCV is an open-source computer vision library that could be used for various video processing tasks, such as object detection and tracking.
- **TensorFlow:** TensorFlow can be used for building and training neural networks. It provides a wide range of tools and functions for deep learning tasks, such as image and speech recognition.

Text Analysis and Summarization

- **Natural Language Processing (NLP) Libraries:** Libraries such as NLTK or spaCy could be used for various text analysis tasks, such as text summarization and entity recognition.

Technologies Used

Search and Recommendation

- Search engines and indexing technologies like Elasticsearch, Solr.
- Machine learning algorithms and frameworks like TensorFlow and scikit-learn for building recommendation engines.
- Natural Language Processing (NLP) libraries like NLTK, Spacy, or Gensim for processing text data and extracting features for search and recommendation purposes.

Project Progress

- Literature Survey Completed.
- Designing of High Level Architecture.
- Documentation and Reporting.

Percentage Completion of the Project - 20%

Capstone (Phase-I & Phase-II) Project Timeline



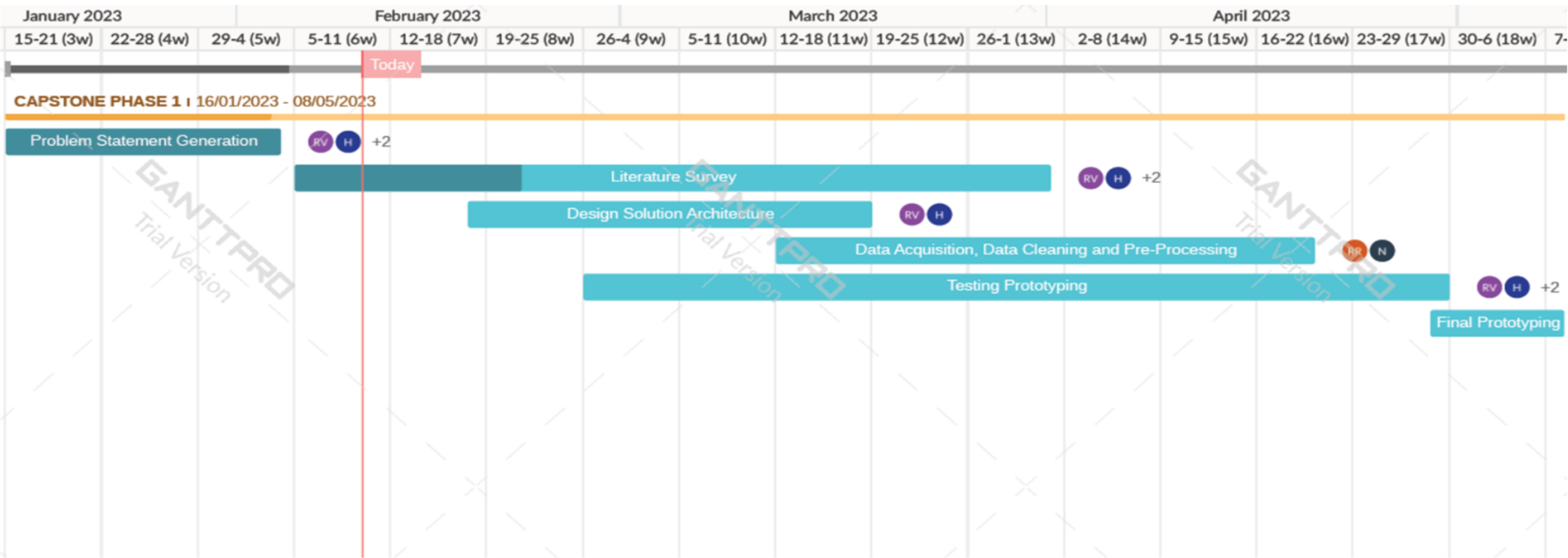
Tasks involved in the 2 Phases of the Project

Task name	Assigned	Start date	End date	Status
		16/01/2023	06/11/2023	
☐ CAPSTONE PHASE 1		16/01/2023	08/05/2023	
Problem Statement Generation	RV H +2	16/01/2023	03/02/2023	● Done
Literature Survey	RV H +2	06/02/2023	31/03/2023	● In progress
Design Solution Architecture	RV H	17/02/2023	17/03/2023	● Open
Data Acquisition, Data Cleaning and Pre-Processing	RR N	13/03/2023	19/04/2023	● Open
Testing Prototyping	RV H +2	27/02/2023	28/04/2023	● Open
Final Prototyping	RV H N	28/04/2023	08/05/2023	● Open
☐ CAPSTONE PHASE 2		31/05/2023	06/11/2023	
Project Development	RV H +2	31/05/2023	08/08/2023	● Open
Finetuning and Refining	RV H +2	28/07/2023	21/08/2023	● Open
Testing	RV H RR	23/08/2023	19/09/2023	● Open
Developing App and UI	RV H +2	18/09/2023	20/10/2023	● Open
Final Documentation	RV H +2	23/10/2023	06/11/2023	● Open

Capstone (Phase-I & Phase-II) Project Timeline



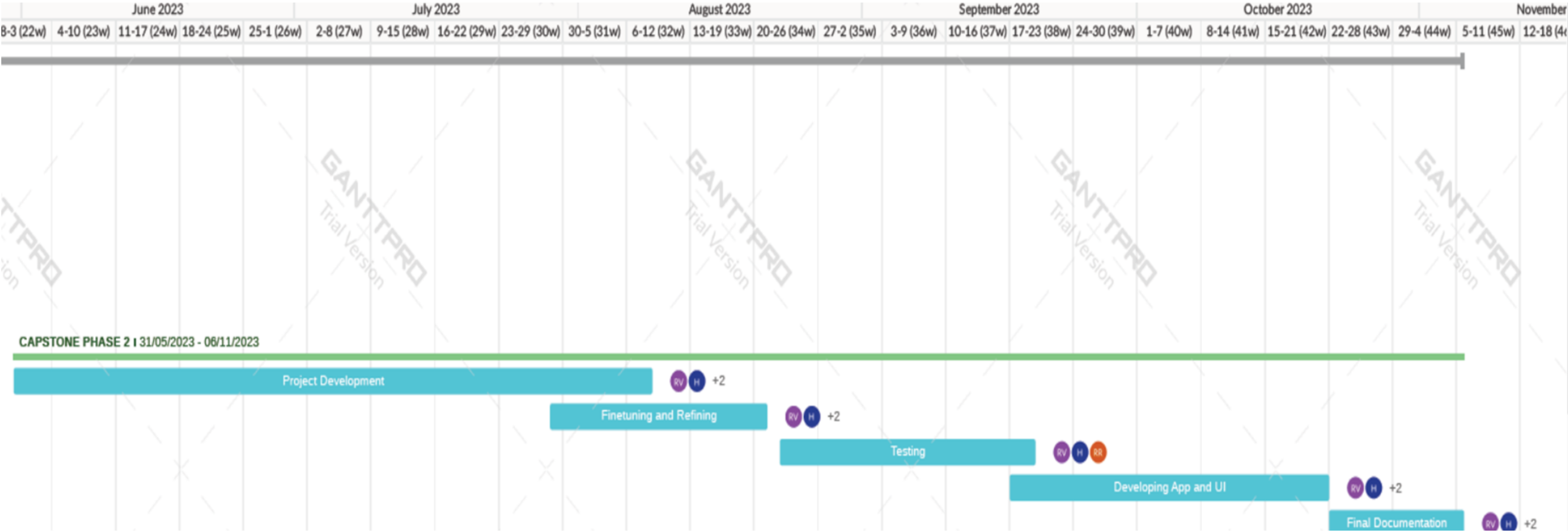
Gantt Chart - Capstone Phase 1



Capstone (Phase-I & Phase-II) Project Timeline



Gantt Chart - Capstone Phase 2



Capstone (Phase-I & Phase-II) Project Timeline



Estimate of Effort

Expand allCollapse allCascade sorting

Filter

Quarters

Export

Task name	Assigned	Start date	End date	Status	+	Quarter 4, 2022			Quarter 1, 2023			Quarter 2, 2023			Quarter 3, 2023			Quarter 4, 2023							
						Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec					
Workload																									
▼ All active projects						<div>Hours</div> <div>Tasks</div>																			
RV ROHIT V SHASTRY						0	0	0	1	3	0	1	1	1	2	2	3	2	1	1					
RR Rushil Ranjan						0	0	0	1	2	1	1	1	1	2	2	3	2	1	1					
H hardiktalluru						0	0	0	1	3	0	1	1	1	2	2	3	2	1	1					
N niranjanraoss1234						0	0	0	1	2	1	2	1	1	2	1	2	2	1	1					

Summary

- The project is a video and text summarization app that utilizes machine learning techniques such as object detection, audio analysis, and text extraction to summarize news videos.
- The app will be constantly updated with new news videos and users can search for videos by keywords, channel names, or people's names.
- The app will also have a recommendation module that suggests videos to users based on their viewing history.
- The high-level architecture of the system includes multiple components such as the front-end application, back-end server, and data storage.
- The implementation of the project will involve the use of various technologies such as Python, TensorFlow.
- The system will require interfacing with external systems such as YouTube and translation APIs.
- The design of the system includes various diagrams such as the high-level architecture diagram, master class diagram, and user interface diagrams

- "Robust Speech Recognition via Large-Scale Weak Supervision" by Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever.
OpenAI - 2022
- "Abstractive Document Summarization with a Graph-Based Attentional Neural Model" by Zhihong Shao, Furu Wei, Sujian Li, and Xiaodong Liu (COLING 2020)
- Jizhou Huang, Haifeng Wang, Wei Zhang, and Ting Liu. 2020. Multi-Task Learning for Entity Recommendation and Document Ranking in Web Search. *ACM Trans. Intell. Syst. Technol.* 11, 5, Article 54 (July 2020)
- Julie Beth Lovins. "Development of a stemming algorithm". In: *Translation and Computational Linguistics* 11.1 (1968), pp. 22–31.
- M. Barbieri, L. Agnihotri, and N. Dimitrova, "Video summarization: methods and landscape," in *Internet Multimedia Management Systems IV*, J. R. Smith, S. Panchanathan, and T. Zhang, Eds., vol. 5242, International Society for Optics and Photonics. SPIE, 2003, pp. 1 – 13.
- V. K. Jeevitha and M. Hemalatha. "Natural Language Description for Videos Using NetVLAD and Attentional LSTM". In: 2020 International Conference for Emerging Technology (INCET). 2020, pp. 1–6. DOI: 10.1109/INCET49848.2020.9154103.

Thank You

