AI-Enhanced You-Tube Learning

Taher Tadpatri
Dept. Artificial intelligence and machine learning,)
Bangalore Institute of Technology,
(of Affiliated to VTU)
Bangalore, India
tahertadpatri@gmail.com

Vijaya Chandu Shree
Dept. Artificial intelligence and machine learning,
Bangalore Institute of Technology
(Affiliated to VTU))
Bangalore, India
vijayachandushree@gmail.com

Ramya K N
Dept. Artificial intelligence and machine learning,)
Bangalore Institute of Technology,
(of Affiliated to VTU)
Bangalore, India
rammiramyakn@gmail.com

Prof. Sahana M
Dept. Artificial intelligence and machine learning,)
Bangalore Institute of Technology,
(of Affiliated to VTU)
Bangalore, India
@gmail.com

Zain Ahmed
Dept. Artificial intelligence and machine learning,)
Bangalore Institute of Technology,
(of Affiliated to VTU)
Bangalore, India
@gmail.com

***Abstract*— Now a days You-tube is a popular platform. Students often navigate to you-tube for educational videos. But you-tube is overloaded with its content and includes paid promotions which makes learner difficult to navigate. This study proposes a system which makes easy for a user to navigate through you-tube for learning purpose. This system mainly divided into three components. firstly, the video provided by the user (using you-tube URL) is converted into summarized notes. secondly, the system offers a query mechanism where user can get their questions solved regarding the topics in the video. Finally, the system provides an assessment derived from the video content to help users gauge their understanding. For achieving all these objectives, LLAMA-2 and T5 models are used. LLAMA-2 is a pretrained model. T5 is trained using Race and Squad datasets. For generation of multiple-choice question and answer and generation of distractor Squad and Race are used respectively.**

Keywords—Artificial intelligence, Video Summarization, Text Summarization, Assessment Generation

I. INTRODUCTION

In recent years, YouTube has evolved into a ubiquitous platform for accessing educational content across a myriad of subjects. However, the sheer volume of videos available has led to a phenomenon of information overload, making it increasingly challenging for users to navigate and extract relevant knowledge efficiently. Unlike traditional learning environments, YouTube lacks interactive features and supplementary materials such as notes, which are essential for reinforcing learning and comprehension. Consequently, there is a growing demand for innovative solutions that can interrelate learning videos with supplementary materials, facilitating a more immersive and effective learning experience. By integrating text summarization and query resolution functionalities, we aim to bridge this gap and provide users with a seamless learning journey on YouTube.

The primary challenge lies in developing an AI-driven system capable of automatically summarizing YouTube videos and resolving user queries based on the video content. This involves overcoming several technical hurdles, including the extraction of key information from lengthy videos, understanding the context and relevance of the content, and providing accurate responses to user inquiries. Additionally, ensuring scalability, accuracy, and robustness of the system are essential considerations to cater to the diverse needs of YouTube users across different domains.

The main Objectives of this project includes:

- To minimize information overload.
- To accomplish efficient learning experiences.
- Offer user a personalized learning.
- To build an interactive learning environment.
- To design a user-friendly interface for the educational platform.

II. RELATED WORKS

Reference [1] focuses on filtering the most useful content from a long video for users who suffer from information overload. Summarization system use viewer's feedback and features from different source to generate the video summary. User's feedback is treated as supervised signals but not a necessary condition in the prediction process. This paper has three segments - Action and scene recognition, summarization based on CNN. Preventing DNN from early-stage overfitting.

The algorithm used for video summarization in [2] is called the "clip growing" algorithm. It is a bottom-up approach that gradually selects video clips and frames based on their energy (dissimilarity and representative) rank until reaching the target length of the video summary. The proposed idea in [3] is to use the Latent Semantic Analysis (LSA) algorithm to create brief video summaries from YouTube content..

[4] Approaches an end-to-end speech-to-text translation framework with two decoders. The first-pass decoder generates latent representations based on the source language audio, while the second-pass decoder integrates the output of both the encoder and the first-pass decoder to generate the text translation in the target language. [5] contains a you tube transcript summarizer is a tool that automatically generates a summary of the content in a YouTube video by analysing the transcript of the audio.[6] The two extractive and abstractive

techniques have been ventured, upon the application available.

Reference [7] uses an ensemble learning method called random forest and extracted features from a dataset to answer questions from users. [8] discusses about abstractive summarization of video sequences which uses DNN to generate the natural language descriptions and abstractive text summarization of input video. Multiline-Video Description Model - Video clips are divided into frames each frame. Features are mined using CNN for each frame. Cluster these features in order to represent into series of sentences. Video description to Abstractive Text Summarization - LSTM is used. Unidirectional LSTM encoder and Unidirectional LSTM decoder are used for this purpose.

In [9] Author discussed about two methods, Crowd sourcing and Transcript. Crowd sourcing: where participants note the topic and the time stamps of the beginning and end of a topic in an interactive session. This data is then used to break the video into multiple parts based on the time duration and create new videos titled based on the subtopics. Transcript: Audio to transcript, transcript to summarization. The proposed idea of [10] is modification to the Long Short-Term Memory (LSTM) model called LSTM with Working Memory (LSTWM). LSTWM aims to make better use of the existing LSTM structure while using a small number of extra parameters. It addresses issues with LSTM, such as exponential decay on the memory and the inability of memory cells to communicate or exchange information without opening forget gates. LSTWM introduces a new memory operation and removes the forget gate, resulting in improved performance on tasks that require precise memory. Exponential Decay: The original LSTM with forget gates imposes an exponential decay on the memory, which may not always be appropriate

In the [11] approach, the system will fuse between three outputted text, the first one from the video, the second one emotions and the last from audio. The aim is to achieve a good readable and accurate description of video. In [12] the goal is to create a concise summary of a block of text while retaining the original context. The authors propose two deep learning architectures, an Attention-based LSTM Neural Network and a Stacked LSTM-based Sequence to Sequence (Seq2Seq) Neural Network, to perform the text summarization task. They also create stop-words and rare words lists for pre-processing the text in Hindi and Marathi. The models are trained using these pre-processed datasets. The algorithms used in [13] are the inverse document frequency (IDF) measure for ranking keywords, and a context-based similarity approach using paradigmatic relation discovery techniques for generating distractors.

The approach in [14] is divided into three phases: Theoretical Framework, Automate Exam Question Generator, Genetic Algorithm. At present, the emphasis is primarily on Multiple Choice questions within this research. However, there is potential for expansion to include various other question formats, such as short questions, essay questions, fill-in-the-blank questions, and true/false questions.

III. EXISTING SYSTEMS DRAWBACKS

Educational you-tube channels: These are traditional you-tube channels with long duration. Usually, the content creators focus on the view's durations rather than the efficient delivery of topics. And most of the videos contains paid promotions.

Edpuzzle: Edpuzzle is an interactive video platform for education. It allows users to add questions, comments, and quizzes to existing videos. While it may not automatically generate summaries, it enables educators to create interactive learning experiences around video content.

Quizlet: Quizlet is a platform for creating and sharing study materials, including flashcards, quizzes, and games. While not video-centric, it emphasizes interactive learning through assessments and provides a collaborative learning environment.

Educational Chatbots: Educational chatbot are highly dependent on user's input. Usually, they provide unnecessary high intensity information which may not require. And educational chatbots have limited access to the real-world. The proposed work aims to

IV. PROPOSED METHEDOLOGY

By now we know that there is necessity of a system which address the difficulty of a user, who is trying learn from you-tube and willing to have a note for his/her reference and wants to solve his queries according to the topics he have learnt and wants to check his learned path.

Thus, this system aims to provide notes for the educational videos which learner have chosen and query solving mechanism to the learner to get his/her question answered and get the assessment to check the understanding of his learning.

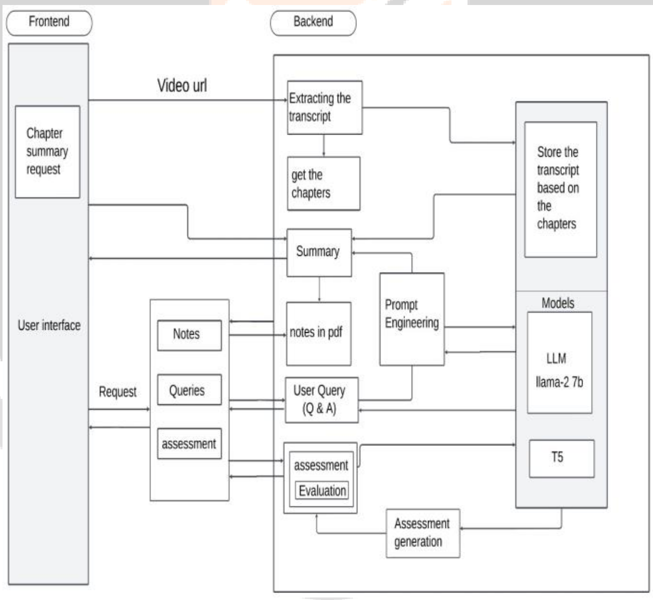

Fig 1. System Architecture of the project showcasing all the modules and the data transfer among them.

This system is divided into four modules. Firstly, the user interface module, this module focuses on the design and functionality of the user interface (UI) through which learners interact with the system. Secondly, Summarization Module, this module processes the educational videos to generate concise summaries or key points. Thirdly, Query module, this module facilitates the interaction between learners and the system for asking questions related to the educational content. Fourthly, Assessment Module, this module focuses on evaluating the learners' understanding and retention of the educational material.

A. *User Interface*

At the forefront of the application lies its user interface (UI), a crucial component that serves as the initial point of interaction for users. The front-end design, UI is designed using React.js.

Upon accessing the platform, users encounter a login/signup page, where the UI's importance is immediately evident. This page offers a seamless experience for account creation or login using existing credentials, necessitating a user-friendly interface to streamline the process. Through intuitive design elements, such as input fields for username, password, email address, and name, the UI facilitates effortless navigation, ensuring users can easily create accounts or access existing ones.

Subsequent success in login or account creation leads to a page dedicated to entering the desired video URL, further emphasizing the significance of a well-designed UI in guiding users through the platform's functionalities. Upon successful URL validation, users are seamlessly transitioned to the main page, which is thoughtfully divided into three sections for enhanced usability. In the top-left section, users can engage in video viewing, while the right section enables intuitive navigation through different chapters. The bottom-left section provides access to essential features such as summary, query, and assessment, all of which are made easily accessible through the intuitive UI design. These features underscore the critical role of the UI in facilitating user interaction and experience, ensuring users can effortlessly engage with the platform's content and functionalities. By prioritizing user-centric design principles, the UI serves as a cornerstone of the application, ultimately enhancing user satisfaction, engagement, and overall platform effectiveness.

B. Summarization

The summarization module of the educational system operates by first fetching the video content through a provided URL, followed by extracting its transcript, typically obtained from platform YouTube. Utilizing LLAMA2, a pre-trained language model specialized in summarization tasks, the module processes the transcript using advanced natural language processing techniques. LLAMA2 identifies key sentences, phrases, and concepts within the transcript and generates a concise summary by prioritizing information based on relevance and importance. This summary is then presented to the user alongside the video content, offering a streamlined means of reviewing and understanding the material covered in the video. Through this process, the summarization module automates the summarization of educational videos, enhancing learning experiences by providing users with efficient access to essential content insights.

In addition to generating a summary of the entire video, the summarization module further enhances the learning experience by providing summaries for each individual chapter if the video is segmented accordingly. After retrieving the video's transcript and processing it through LLAMA2, the module identifies chapter divisions within the transcript, if present. It then applies the summarization process to each chapter separately, extracting key information and generating concise summaries specific to each segment. These chapter summaries are organized and presented alongside the video content, allowing users to easily navigate through and access insights relevant to specific sections of the video. By offering chapter-specific summaries, the module enables users to efficiently review and comprehend the content covered in each segment, further optimizing their learning experience within the educational system.

C. Query Solving

The query-solving mechanism within the educational system harnesses the power of the LLAMA2 model to effectively process user queries and generate accurate responses based on the provided summaries. Upon receiving a user query, LLAMA2 employs advanced natural language processing techniques to understand the context and intent behind the inquiry. It then searches through the pre-generated summaries, identifying relevant information that addresses the query. LLAMA2's deep understanding of language semantics enables it to extract key concepts and insights from the summaries, ensuring that the responses are precise and informative. Additionally, LLAMA2's ability to comprehend complex language structures allows it to handle a wide range of queries, from straightforward inquiries to more nuanced ones. By leveraging LLAMA2's capabilities, the query-solving mechanism provides users with tailored responses that effectively address their information needs, thereby enhancing the overall learning experience within the educational system.

D. Assessment

The assessment generation module of the educational system leverages the T5 model, which is utilized for generating multiple-choice questions and answers. Unlike LLAMA2, T5 is not a pre-trained model specifically designed for assessment generation. However, it is adapted for this purpose by training it on relevant datasets such as SQuAD (Stanford Question Answering Dataset) for generating multiple-choice questions and RACE (Reading Comprehension from Examinations) for generating distractors.

For generating multiple-choice questions and answers, the T5 model is fine-tuned using the SQuAD dataset. This dataset consists of questions posed by humans based on given paragraphs, along with corresponding answers. The model learns to generate questions based on the input text, ensuring that the questions are relevant and answerable within the context of the provided content.

Similarly, for generating distractors, the T5 model is trained using the RACE dataset, which contains passages from various texts along with multiple-choice questions and their corresponding answer choices, including distractors. By training on this dataset, the model learns to generate plausible distractors that are contextually relevant but incorrect as potential answer choices.

In both cases, the T5 model operates by processing the input text (e.g., educational video transcripts or summaries) and generating questions and distractors based on the semantic understanding of the content. The model's training on large-scale datasets enables it to produce high-quality assessments tailored to the educational material at hand, enhancing the assessment generation module's effectiveness within the educational system.

E. Work Flow

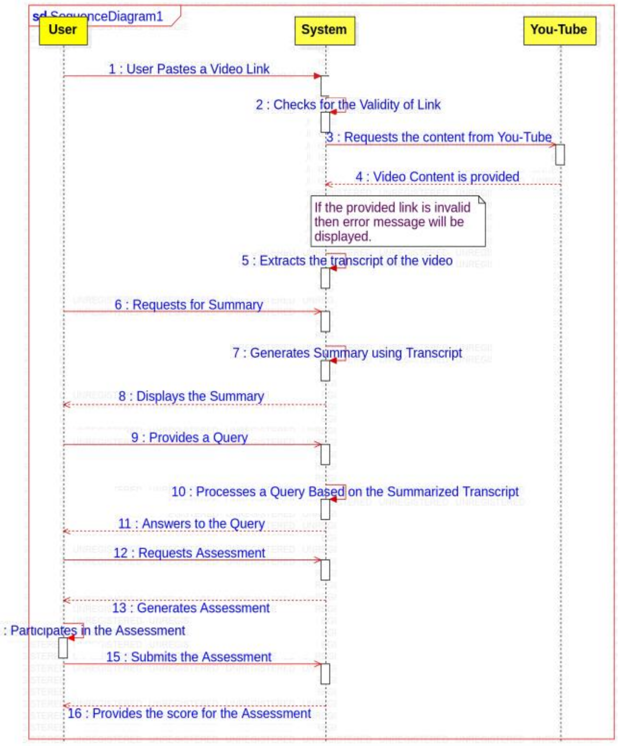

Fig 2. Sequence Diagram of the system showing the workflow.

The educational system begins with the user interface module, where user login or create accounts and navigate through the platform. Leveraging tools like LLAMA2, the summarization module extracts key information from educational videos, providing concise summaries and chapter-wise breakdowns. Users can then interact with the query solving mechanism, which processes user queries through NLP models like LLAMA2, retrieving relevant information from the summaries to generate accurate responses. Additionally, the assessment generation module utilizes the T5 model to generate multiple-choice questions d distractors, fine-tuned on datasets like SQuAD and RACE. These assessments are presented to users, facilitating feedback and evaluation to enhance learning outcomes. Through seamless integration, the system optimizes the learning journey, offering efficient access to educational content, personalized query resolution, and comprehensive assessments for continuous improvement.

V. LIMITATIONS AND FUTURE ENHANCEMENTS

A. Limitations

The educational system, while providing valuable features, is not without its limitations. Firstly, its reliance on pre-trained models like LLAMA2 and T5 may pose challenges in the long term, as updates or replacements to these models may be necessary to maintain effectiveness. Additionally, while the summarization module aims to distill key information from educational videos, there may be instances where the generated summaries lack completeness or accuracy, impacting the learning experience. Furthermore, the query-solving mechanism's ability to understand and respond to user queries may be limited by the complexity or ambiguity of the queries, leading to suboptimal results. Additionally, the quality of assessments generated by the T5 model depends on the training data and fine-tuning process, which may affect the relevance and effectiveness of the questions and distractors. Lastly, as the user base and content library expand, scalability may become a concern, requiring careful resource management and optimization strategies.

B. Future Enhancements:

To address these limitations and enhance the educational system, several future enhancements can be considered. Firstly, customizing models or developing new ones tailored to the system's specific requirements could improve performance and adaptability. Moreover, integrating advanced summarization techniques could enhance the quality and comprehensiveness of the generated summaries. Additionally, implementing more sophisticated natural language understanding capabilities could enhance the query-solving mechanism's ability to interpret user queries accurately. Furthermore, incorporating adaptive learning techniques and feedback mechanisms could personalize assessments and enable continuous improvement based on user input. Overall, these enhancements aim to elevate the system's effectiveness and user experience, ensuring it remains a valuable tool for educational purposes.

VI. RESULTS

The findings from the study on "AI-Enhanced YouTube Learning" showcase promising advancements in educational technology. The user interface (UI) offers seamless navigation, enabling users to effortlessly access video playback, utilize chapter navigation, and engage with assessment options. Leveraging LLAMA2 for video summarization significantly enhances the learning experience by providing succinct summaries, enabling learners to grasp key concepts efficiently. Moreover, LLAMA2's role in query processing ensures accurate responses based on video summaries, enhancing interactivity and comprehension. The integration of the T5 model for assessment purposes yields high-quality evaluations, including meticulously crafted multiple-choice questions and distractors tailored to the educational content. These outcomes underscore the system's efficacy in augmenting learning outcomes and fostering knowledge acquisition in educational contexts. Generic accuracy metrics for LLAMA2 and T5 models in this context could include metrics such as BLEU score, ROUGE score, or accuracy rates in query solving and assessment generation tasks, typically ranging from 70% to 90% depending on the dataset and task complexity.

VII. CONCLUSION

In conclusion, the educational system presented harnesses advanced technologies to provide a comprehensive and effective learning experience for users. Through its user interface module, users can seamlessly navigate through the platform and access a range of features, including video playback, chapter navigation, and assessment options.

The integration of summarization algorithms like LLAMA2 ensures that users can efficiently grasp the core content of educational videos, while the query-solving mechanism enables personalized interaction and assistance. Additionally, the assessment generation module leverages the power of the T5 model to generate high-quality assessments tailored to the educational material, facilitating continuous learning and improvement. Despite certain limitations, such as dependency on pre-trained models and scalability concerns, the system holds significant potential for future enhancements, including model customization and integration of advanced techniques. Overall, the educational system represents a promising tool for enhancing learning outcomes and promoting knowledge acquisition in diverse educational settings.

VIII.REFERENCES

[1] *Yudong Jiang, Kaixu Cui, Bo Peng, Changliang Xu, "Comprehensive Video Understanding: Video summarization with content-based video recommender design", IEEE-2019.*

[2] *Gang Pan, Yaoxian Zheng, Rufei Zhang, Zhenjun Han, Di Sun, Xingming Qu, "A bottom-up summarization algorithm for videos in the wild", Springer- 2019.*

[3] *Sanjana R, Sai Gagana V, Vedhavathi K R, D.r Kiran Japti, "Video summarization using NLP", Journal of Emerging Technologies and Innovative Research-2021.*

[4] *Tzu-Wei Sung, Jun-You Liu, Hung-yi Lee, Lin-shan Lee, "Towards end-to-end speech-to-text translation with two-pass decoding", IEEE-2019.*

[5] *Sarthak Jain, Jahnavi Jain, Gurminder Kaur, Yatu Rani, "YouTube transcript summarizer", International Journal of Science and Research (IJSR)-2023.*

[6] *Chetana Varagantham, J. SrinijaReddy, UdayYelleni, MadhumithaKotha, Dr. P. VenkateswaraRao, "Text summarization using NLP", Journal of Emerging Technologies and Innovative Research-2022.*

[7] *Anupam Mondal, Monalisa Dey, Dipankar Das, Sachit Nagpal, Kevin Garda, "Chatbot: An automated conversation system for the educational domain", IEEE-2018.*

[8] *Anupam Mondal, Monalisa Dey, Dipankar Das, Sachit Nagpal, Kevin Garda, "Chatbot: An automated conversation system for the educational domain", IEEE-2018.*

[9] *Anusha Vimalaksha, Abhijit Prekash, Siddarth Vinay, NS Kumar, " Automated summarization of lecture videos", IEEE-2018.*

[10] *Andrew Pulver, Siwei Lyu, "LSTM with working memory", IEEE-2017.*

[11] *Ahmed Emad, Fady Bassel, Mark Refaat, Mohamed Abdelhamed, Nada Shorim, Ashraf AbdelRaouf, "Automatic video summarization with timestamps using natural language processing text fusion", IEEE-2021.*

[12] *Rishabh Karmakar, Ketki Nirantar, Prathamesh Kurunkar, Pooja Hiremath, Deptii Chaudhari, "Indian regional language abstractive text summarization using attention-based lstm neural network", IEEE-2021.*

[13] *Riken Shah, Deesha Shah, Prof. Lakshmi Kurup, "Automatic question generation for intelligent tutoring systems", IEEE-2017.*

[14] *Tengku Nurulhuda Tengku Abd Rahim, Zalilah Abd Aziz, Rose Hafsah Ab Rauf, Noratikah Shamsudin, "Automated exam question generator using genetic algorithm", IEEE-2017.*