**AI Powered Voice-Guided Presentation: Revolutionizing PPT Explanations**

***A Report Submitted in Partial Fulfillment of the Requirements***

***for the***

***Final Year B.Tech Project***

**By**

Mohit Ranjan 2012109  
Rituraj Chanda 2012157

Shibam Debnath 2012181

Under the Guidance of

**Dr. Partha Pakray**

Associate Professor

**Department of Computer Science and Engineering**

**National Institute of Technology Silchar**

**Cachar, Assam, INDIA - 788010**

**June - July 2023**

**DECLARATION**

**“*AI Powered Voice-Guided Presentation: Revolutionizing PPT Explanations*”**

We declare that the presented work represents largely our own ideas and work in our own words. Where other ideas or words have been included, we have adequately cited them listed in the reference materials. We have adhered to all principles of academic honesty and integrity.

**Name Scholar ID**

Rituraj Chanda 2012157

Shibam Debnath 2012181  
Mohit Ranjan 2012109

Department of Computer Science and Engineering

**National Institute of Technology Silchar, Assam**

**ACKNOWLEDGEMENT**

We would like to extend our heartfelt gratitude to our mentor, **Dr. Partha Pakray**, for his invaluable guidance and support throughout the duration of our project. Dr. Partha Pakray's expertise, dedication, and mentorship have played a pivotal role in the Progress of our project. We also wish to express our sincere appreciation to his teaching assistant, **Advaitha Vetagiri**, for his assistance, patience, and willingness to help us navigate the intricacies of our project. Their contributions and support have been indispensable, and we are truly thankful for their efforts. Dr. Partha Pakray's profound knowledge, constructive feedback, and encouragement have not only enriched our project but have also contributed significantly to our academic and professional growth. We are fortunate to have had the opportunity to work under his mentorship. To Advaitha Vetagiri, we extend our gratitude for their guidance and responsiveness to our queries. Their assistance has been invaluable, and we are appreciative of their contributions to our project. We are deeply grateful to both Dr. Partha Pakray and Advaitha Vetagiri for their unwavering support and mentorship. Their belief in our capabilities has motivated and inspired us to strive for excellence. Once again, thank you to Dr. Partha Pakray and Advaitha Vetagiri for their exceptional mentorship and support. We greatly appreciate all that they have done to help us succeed in this project.

**Name Scholar Id**

Rituraj Chanda 2012157

Shibam Debnath 2012181

Mohit Ranjan 2012109

**Abstract**

The advent of Artificial Intelligence (AI) has precipitated a paradigm shift across various sectors, ushering in a new era of innovation and efficiency. Within this transformative landscape, our project emerges as a pioneering endeavor to leverage the full potential of AI to elevate the educational experience. At its core, our initiative introduces a groundbreaking AI-powered PowerPoint (PPT) presenter, designed not only to adeptly parse the content of presentations but also to dynamically generate a script for an AI model to articulate. This synthesis of advanced technologies, including text summarization and script generation, seeks to overcome the limitations inherent in traditional PPT presentations, particularly in the context of providing tailored and personalized learning experiences. By incorporating text summarization techniques, our AI presenter distills complex information into concise, comprehensible segments, ensuring that learners receive a nuanced understanding of the subject matter. The script generation component further enhances the adaptability of our system, enabling it to cater to diverse learning styles and preferences. This integration empowers our AI model not only to generate scripts but also to articulate them in a natural and engaging manner. The significance of this feature becomes apparent when considering its potential to make educational content accessible to visually impaired learners. In doing so, we aim to break down barriers and foster inclusivity in education, ensuring that all students, regardless of physical abilities, can partake in the learning process. Motivating our project is a profound awareness of the inadequacies inherent in current educational models, particularly in meeting the diverse needs of learners, especially those in economically challenged and remote areas. Recognizing the disparities in access and affordability, our vision is to create an AI-based teacher available around the clock. Through the seamless integration of AI technologies, our project envisions a future where education is not only democratized but also personalized, ensuring that every learner can unlock their full potential.

.

**List of Figures**

1. AI based Teacher . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . … .. . . . …. … .. . .. . . . . . .. .. … 01

2. LLAMA-2-fine tuned chat model . .. … . .. . . . . .. . . . . . . . . .. .. . . . . . . . . . . . . . . . . … .. .. .02

3. Text Generation Flow…. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .. … … .. 03

4. Text Summarization . . . . . . . . . . . .. . . . . . . . . . . . . . . . . .…. … .. . . . . . . . . .… .. .… .. .. … .. .04

5. Text to Speech model . . . . . . . . . . .. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . … .. . 05

6. workflow of our methodology . . . . . .. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .. … . . . . 07

8. Google’s text-to-speech based on wavenet outperforms other synthetic voices . . . . . . . … . 13

9. Flask server codebase . . . . . . . . . . . . .. . . . . . . . . . . . . . . .. . . . . . . . . . . . . . . . . . . .. … . . 15

11 Image of Flask server running. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 1612 Output of text-to-speech model . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .. . . . . .. . . 16

**List of Tables**

1. Llama 2 outperforms other open source language models. . . . . . . . . . . . . . . . . . . . . . . . . …11

2. Comparison of GPT with llama-2 model on different parameters . . . . . . . . . . . . . . . . . . . . . . . . 12

**CONTENTS**

**Declaration**

**Acknowledgement**

**Abstract**

**List of Figures**

**List of Tables**

1. **Introduction** . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .. 1
2. **Literature Survey** . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .. 2
3. **Motivation**. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . ... . . . . . . . . . .. .. 5
4. **Proposed Methodology**  . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .. .. 6

4.1 Content generation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . ... . . . . . . . . . .. 6

4.2 Text to speech . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .. . ... . . . . . . . . . .. 8

4.3 Web Integration. . . . . . .. . . . . . . . . . . . . . . . . . . . . . . . . . . . .. . ... . . . . . . . . . .. 8

**5. Result and Discussion**  . . . . . . . . . . . . . . . . . . . . . . . . . .. ... . . . . . . . . .. . . . . . . . .. 18

**6. Conclusion and Future Work**  . . . . . . . . . . . . . . . . . . .. ... . . . . . . . ... . . . . . . .. . .. 27

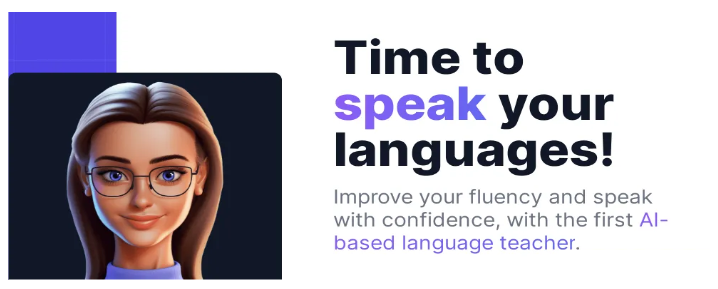
**7. References**  . . . . . . . . . . . . . . . . . . . . . . . . . .. . . . . . . . .. ... . . . . . . . ... . . . . . . . . .. 28

**1. INTRODUCTION**

In today's education world, everyone uses PowerPoint presentations to share information. But the usual way of doing it doesn't always meet the different needs of students. Our project wants to change this by using AI to make educational content clearer and more accessible. We start by looking at why regular PowerPoint presentations may not be good enough, talking about the problems they have and why we need learning that is more personal and doesn't cost a lot. We explain our plan for fixing these issues in the introduction, where we talk about our main goals and how we're going to tackle the problems we found.

Regular PowerPoint presentations are everywhere in education, but they don't always work well for everyone. Some students learn better in different ways, and the usual presentations don't take that into account. Our project wants to use AI to make educational content better for everyone. Let's look at why regular PowerPoint presentations have problems. First, they don't always fit the way students learn. Some students need more personalized learning, something that speaks to them directly. Second, the usual way of doing things can cost a lot. This is a big issue, especially for students who may not have a lot of money.

Now, let's talk about what our project is all about. We want to change how educational content is presented using AI. We start by setting the stage, explaining why regular presentations have problems and why our project is needed. Then, we dive into how we're going to do it – our plan and goals. Our project wants to use AI to make educational content better for everyone. We see that regular PowerPoint presentations may not be the best fit for all students. Some learn in different ways, and the usual presentations don't cater to these diverse needs. Our goal is to fix this by making learning more personal and not costing a lot. The usual way of doing things can be expensive, and this makes education less accessible, especially for those with limited resources. So, our project focuses on finding a solution that doesn't burn a hole in students' pockets. This is a crucial aspect because education should be affordable for everyone.



**Fig. 1**: AI based Teacher

The remainder of the paper is organized in the following manner. Section II presents Literature Survey. In section III, we demonstrate the Proposed Methodology. In section IV of the paper we explained Performance. In section V we demonstrate experiment results and discussion In section VI includes conclusion and future work. Insection VII contains the references section

**2. LITERATURE SURVEY**

# The integration of Artificial Intelligence (AI) into educational technologies has witnessed substantial growth in recent years, with a focus on enhancing learning experiences and addressing the diverse needs of students. In this literature survey, we explore the key areas relevant to our project, including AI in education, text summarization, script generation, and Text-to-Speech (TTS) technology.

# **2.1 Llama 2: Open Foundation and Fine-Tuned Chat Models:**

# The paper [1] introduces the development and release of Llama 2, a collection of pretrained and fine-tuned large language models tailored for dialogue applications. Through comprehensive benchmarks, Llama 2 is demonstrated to outperform existing open-source chat models, showcasing superior performance metrics and response coherence. The paper [1] delves into the fine-tuning methodology, emphasizing safety improvements to mitigate biases and enhance model robustness. Llama 2's open foundation is highlighted, inviting community collaboration and contributions. The authors stress the importance of responsible AI development and provide detailed insights into the technical aspects, including model architecture, training data, and methodologies, offering a valuable resource for researchers and practitioners in the field of natural language processing.

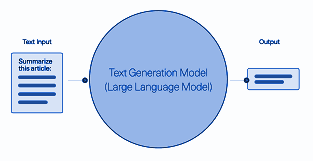


**Fig.2** : LLAMA-2-fine tuned chat model

**2.2 Text Generation**:

Text generation is a dynamic field witnessing substantial contributions that leverage innovative approaches to enhance the quality and interactivity of generated content. One notable paper, "*Interactive Text Generation,*" [8] pioneers the introduction of an Interactive Text Generation task. This task allows training generation models interactively through user simulators, facilitating model guidance without direct user involvement. Employing Imitation Learning, the paper [8] demonstrates the superiority of interactively trained models over non-interactive counterparts, even when subjected to the same user input or edit budgets. This contribution enhances the understanding and applicability of interactive approaches in the realm of text generation.

In tandem, "*A Systematic Literature Review on Text Generation Using Deep Neural Network Models*" [6] offers a comprehensive overview of text generation research employing deep neural network models. Conducting a systematic literature review from 2015 to 2021, the paper[6] identifies key aspects such as language translation, dataset balancing, next word prediction, and various other domains. By cataloging 90 primary studies, the review not only presents existing research but also identifies research gaps, providing valuable insights for future exploration. The paper's [6] contribution extends to offering guidelines for practitioners and proposing future directions to steer advancements in text generation using deep neural network models.Additionally, the "*Long Text Generation Challenge*" [7] introduces a unique shared task focused on generating consistent, human-like long texts. Proposing an innovative statistical metric, GAPELMAPER, to measure the structuredness of generated text, the paper [7] presents a holistic evaluation protocol that incorporates human assessment. This challenge encourages researchers to explore diverse sampling approaches, prompting strategies, and architectures, thereby addressing the limitations associated with generating lengthy texts. Collectively, these contributions not only advance the understanding of interactive training methodologies but also provide a comprehensive overview of existing research trends and foster new avenues for overcoming challenges in long text generation.



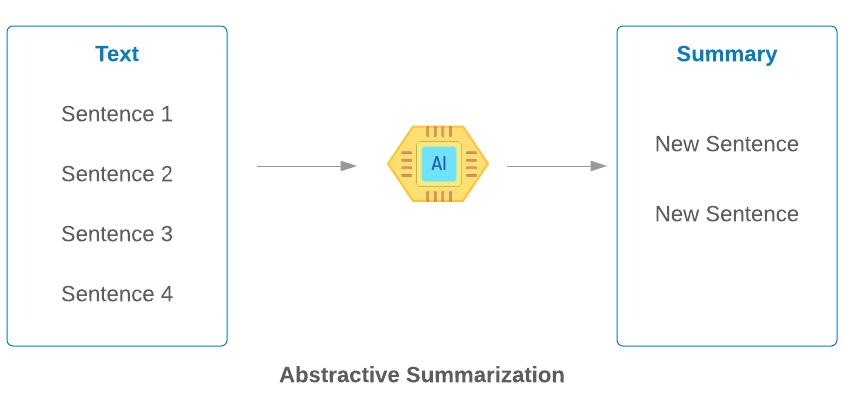
**Fig. 3**: Text Generation Flow

**2.3 Text-Summarization:**

In the dynamic field of natural language processing, text summarization plays a pivotal role in distilling essential information from extensive documents. There are many research works that have made significant contributions in the advancement of the understanding and application of text summarization. Some of them can be, "*Summarization Of Text Using Natural Language Processing,*" [4] which serves as a comprehensive guide to recent advancements in text summarization. It [4] meticulously categorizes summarization techniques, ranging from extraction-based to abstraction-based and hybrid approaches. The paper [4] delves into the evaluation metrics commonly used to gauge summarizer effectiveness and surveys state-of-the-art models, including the application of deep learning-based approaches such as transformers. Its significance lies in providing researchers and practitioners with an invaluable resource for understanding the nuanced landscape of text summarization.

"*Inshort Text Summarization of News Article*" [3] introduces a novel methodology for abstractive text summarization, particularly tailored for news articles. Leveraging recurrent neural networks and sentence scoring approaches, the paper stands out by demonstrating superior summarization quality and efficacy in comparison to existing methods. This contribution signifies a paradigm shift towards more advanced abstractive techniques, offering a promising direction for improving the efficiency and accuracy of text summarization, especially within the realm of news articles.

"*AI Text Summarization System Utilizing GPT*" [2] where the integration of GPT, a powerful language model, is employed alongside state-of-the-art natural language processing techniques. The modular architecture of the system allows for scalability and flexibility, showcasing effectiveness in generating concise and meaningful summaries. Evaluation metrics, such as Rouge and F1 scores, attest to the system's capability in capturing key information. Its [2] integration with Android platforms enhances accessibility, making it suitable for diverse applications. This paper [2] contributes to the practical side of text summarization, presenting a promising approach for automatically summarizing extensive text volumes while providing users with time-saving and meaningful summaries. Together, these three papers collectively advance the understanding and application of text summarization across diverse methodologies and innovative applications.



**Fig. 4**: Text Summarization

**2.4 Text-to-Speech Technology:**

# There are now two types of TTS[15] conversion systems on the market, rule-based and machine learning-based. Rule-based TTS[16] systems generate voice from text based on a predetermined set of rules and algorithms. To generate speech that is convincingly human-like, these systems often consult a massive database of phonetic and prosodic information. High-quality speech is possible with rule-based TTS systems, but their development and upkeep can be time-consuming and costly due to the complexity of the requisite database. On the other hand, TTS systems based on machine learning employ statistical models to convert text into speech. The systems may be trained on massive datasets to produce high-quality speech, and they commonly use deep neural networks (DNNs) to represent the connections between text and speech. While they have their benefits, TTS systems that rely on machine learning are still susceptible to overfitting and generalization issues, and they are only as good as the training data they are given.Recent research has analyzed the current state of TTS conversion, and the limits of current TTS systems have been discussed[17]. The limitations of rule-based TTS systems, such as their need for a vast database of phonetic and prosodic information and its difficulty in modeling complex linguistic processes, were noted by Liu et al. (2018) [11] in their review of TTS[18] systems. They also talked on the challenges of modeling long-term dependencies in speech and the need for high-quality training data machine learning-based TTS systems. In a similar vein, Tacchini et al. (2019) [12] reviewed deep learning-based TTS systems and noted the shortcomings of current TTS systems, such as the need for massive volumes of annotated speech data and the challenge of accurately modeling prosodic variation and expressiveness. They also talked about the issues associated with training deep neural networks for TTS conversion, such as the necessity for a lot of computational resources and the difficulties of avoiding overfitting and generalization problems.

****

**Fig. 5**: Text to Speech model

**3. MOTIVATION**

The project on AI Powered Voice-Guided Presentation: Revolutionizing PPT Explanations is based upon multiple goals. Starting from Helping everyone learn better, supporting teachers for better education , Efficiently and saving resources etc. The key motivations behind the project is explained below:-

**3.1. Helping Everyone Learn Better:**

We want to make a tool so that everyone , including those who can’t see well, can learn easily . It's about making sure nobody feels left out when it comes to learning new things. Sometimes presentations are hard for people who can’t see well . Our project want’s to make sure learning is easy for everyone, no matter what challenges they might have.

**3.2. Supporting Teachers for better Education:**

Teachers already have a lot on their plates. With our system, we want to take care of some routine tasks, allowing teachers to focus on creating engaging and personalized learning experiences for their students. Imagine a tool that ensures every student, no matter who the teacher is, receives a consistently high-quality learning experience. This is about creating a standard of excellence in education.

**3.3. Efficiency and Saving Resources:**

For schools and organizations, resources are often limited. Our project aims to use resources efficiently, making sure that every dollar spent on education delivers the maximum impact.Technology should be about the future. Our project is not just a quick fix; it's a long-term approach to education that considers scalability and adaptability for the future.

**3.4. Bringing Education to Remote Places:**

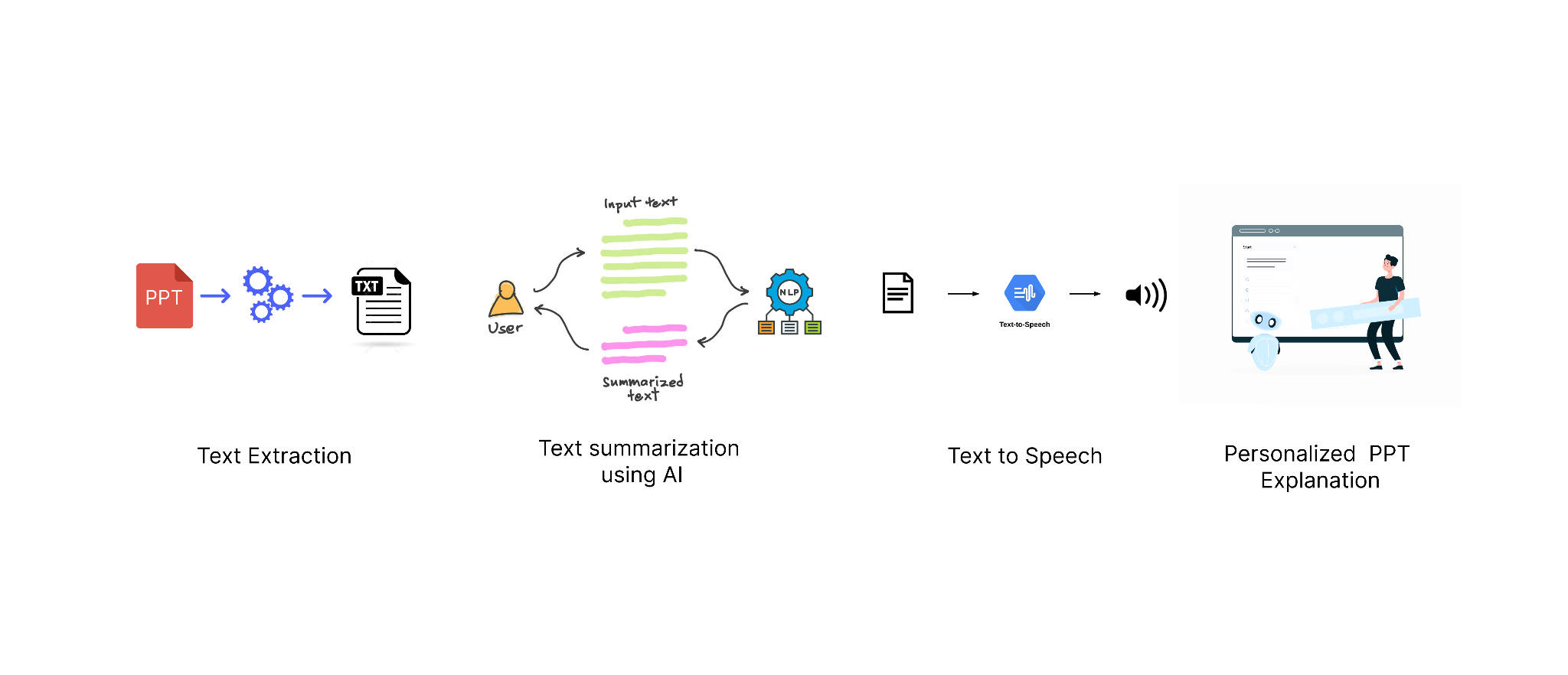
Our project is about making sure that students in very remote areas have the same chance to learn and explore as students in big cities. It's about giving everyone, no matter where they are, an equal shot at a great education.Sometimes, students in faraway places feel left out because they don't have the same resources. Our project wants to change that. It wants to bring the excitement of learning to every corner, so no student feels left behind.

In conclusion, a research study on AI Powered Voice-Guided Presentation Revolutionizing PPT Explanations is motivated by a wide range of real-world applications, from Education and assistive technology to consumer electronics and urban planning. In addition we use artificial intelligence, for optimum use of resources and provide better education facilities in rural areas

**4. METHODOLOGY**

The goal of this research is to revolutionize PowerPoint presentations by incorporating AI powered voice guidance. The methodology involves a multi-step process, combining natural language processing, text extraction, natural language understanding, text generation, summarization, and speech synthesis to enhance the explanation of each slide's content.

The complete methodology we followed to progress in this project can be divided into three main segments, text-extraction, text-generation, text-summarization, text-to-speech synthesis etc. The detailed explanation of each process is given below

****

**Fig. 6**: workflow of our methodology

**4.1. Content Generation**

In this part we generate the content said by the avatar to explain the given PPT to the user. The main components under this section are text-extraction, text-generation, content-segmentation and text-summarisation. The detailed explanation of each component is given below

**4.1.1. Text Extraction:**

The process of extracting contents from a PowerPoint (PPT) file involves using the `*python-pptx*` library in Python. The script begins by creating a `Presentation` object to represent the PPT file. It then iterates through each slide in the presentation. For each slide, it determines the heading by checking if the slide has a title; if not, it defaults to a generic heading format. The script then traverses each shape within the slide, accumulating text content from shapes that have text attributes. This text content, along with the heading, is organized into a dictionary, where the heading serves as the key. Finally, the script returns this dictionary containing the extracted slide contents. The approach is modular, allowing for flexibility in handling different PPT structures, and the use of conditional statements ensures robust extraction even when specific elements, like titles, may be absent.If in case of pdf we can use pyp library and langchain. Using PyPDFLoader function from langchain we can extract the contents of the pdf. And then splitted the whole content into several segments because large language models works under restricted amount of input datasize.

**4.1.2. Text Generation:**

Meta developed and publicly released the Llama 2 family of large language models (LLMs), a collection of pretrained and fine-tuned generative text models ranging in scale from 7 billion to 70 billion parameters. We used fine-tuned LLMs, called Llama-2-Chat, are optimized for dialogue use cases. Llama-2-Chat models outperform open-source chat models on most benchmarks we tested, and in our human evaluations for helpfulness and safety, are on par with some popular closed-source models like ChatGPT and PaLM.Llama 2 is an auto-regressive language model that uses an optimized transformer architecture. The tuned versions use supervised fine-tuning (SFT) and reinforcement learning with human feedback (RLHF) to align to human preferences for helpfulness and safety.

For each slide, a text generation model from Hugging Face, specifically "*meta-llama/Llama-2-7b-chat-hf*" is employed to generate explanations for individual points in the context of the slide heading. This step ensures coherent and relevant content generation.also we performed hyperparameter tuning of the model on our dataset. So that it gives better and relevant results for our given prompts. Then, we concatenated the explanations from each point and summarized the whole concatenated text.

**4.1.3. Content Segmentation:**

If the generated text after concatenation is larger than the maximum limit of input size of our LLM model then we need to divide the text into multiple segments. To perform this task we employed the **“***RecursiveCharacterTextSplitter***”** modulefrom langchain framework. Which splits the contents into a custom size given by us.

**4.1.4. Text Summarisation:**

For text summarization purposes we are using a pre-trained hugging face text-summarization model named **“***facebook/bart-large-cnn***”** and fine tuning it over our dataset.BART is a transformer encoder-encoder (seq2seq) model with a bidirectional (BERT-like) encoder and an autoregressive (GPT-like) decoder. BART is pre-trained by (1) corrupting text with an arbitrary noising function, and (2) learning a model to reconstruct the original text.BART is particularly effective when fine-tuned for text generation (e.g. summarization, translation) but also works well for comprehension tasks (e.g. text classification, question answering). This particular checkpoint has been fine-tuned on CNN Daily Mail, a large collection of text-summary pairs.

In this way we get the whole summarized text for each slide and also the context-relevance of the generated content with the given topic is retained by again fine tuning the well-trained models over our dataset. Since, the models used in this process are trained over a large dataset.So, it gives a good result for the text summarization part.

**4.2. Text to speech**

In our ongoing efforts to make information accessible to everyone, we've introduced Google Text-to-Speech (GTTS)[19] to our platform. This tool helps convert written text into spoken words, making content more accessible, especially for those who prefer listening or have visual impairments. Within the domain of Text-to-Speech (TTS) technology, the incorporation of the Google Text-to-Speech API[19] is a crucial element in our project, delivering an unmatched auditory experience. Harnessing the capabilities of Google's advanced AI technologies, this API provides a comprehensive suite of features that enhance the quality of synthesized speech to a level approaching human-like precision. The API[20] is distinguished by its high-fidelity speech, a key attribute influenced by DeepMind's proficiency in speech synthesis. This characteristic imparts a natural intonation to the voices, thereby enriching the user experience and rendering educational content more captivating.

**Why GTTS Matters:** GTTS has a few important features that make it valuable for our mission:

**Natural Sounding Speech** **:** GTTS doesn't just read text aloud; it makes it sound natural, almost like a person speaking. This makes the information more engaging and easier to follow.

**Language Variety:** It supports many languages and accents, so our users can experience content in the language they're most comfortable with. This inclusivity is crucial for our diverse community.

**User-Friendly web application :** Customization:Users can tweak certain aspects of the speech, like the speed or pitch, making it a more personalized and comfortable experience.

**How We Use GTTS:** Our platform seamlessly integrates GTTS to convert written content into spoken words. Whether users are on a computer, tablet, or phone, they can access information more conveniently through voice.

**4.3. : Web Integration with Flask server and React :**

In the realm of educational technology, the successful integration of web applications can significantly enhance the accessibility and user experience of AI-powered tools. Our project seamlessly marries Flask, a Python web framework, and React, a popular JavaScript library for building user interfaces, to create a robust and user-friendly web application. This section delves into the intricacies of our web integration, outlining the architecture, functionality, and the seamless interaction between the Flask server and React components.

**Architecture Overview:**   
The web application architecture is designed with a client-server model, with Flask serving as the backend server and React handling the frontend interactions. This approach ensures a clean separation of concerns, allowing for scalability, maintainability, and efficient data flow. The Flask server provides RESTful API endpoints, enabling communication between the backend and frontend components.

**Flask Server:**   
Our Flask server acts as the backbone of the web application, handling key functionalities such as authentication, data processing, and communication with external APIs. Different endpoints are meticulously crafted to cater to specific functionalities, ensuring a modular and organized codebase. For instance, authentication endpoints manage user login and registration, while others handle PPT file uploads, text summarization, script generation, and Text-to-Speech functionalities. The Flask server acts as the bridge between the user interface and the AI-powered backend.

**React Frontend:**   
The React frontend complements the Flask backend, providing an interactive and intuitive user interface. Leveraging React's component-based architecture, we design modular components for seamless navigation and an engaging learning experience. The user interface incorporates elements for uploading PPT files, configuring user settings, and interacting with the AI-generated educational content. React's state management ensures dynamic updates, providing real-time feedback to users.

**Endpoint-based Functionality:**   
The Flask server exposes various endpoints, each catering to a specific functionality of the AI-powered educational platform. For instance, the PPT upload endpoint handles the reception and processing of uploaded PPT files. The text summarization and script generation endpoints leverage external APIs to distill information from the PPT content. The Text-to-Speech endpoint converts the generated script into audible content. React components, in turn, interact with these endpoints, facilitating a cohesive flow of data and actions between the frontend and backend.

**Responsive Design:**   
Our web application incorporates responsive design principles, ensuring a seamless experience across devices. React components adapt to varying screen sizes, facilitating accessibility for users on desktops, laptops, tablets, and mobile devices. The responsive design contributes to the accessibility and inclusivity goals of our project.

In conclusion, the web integration of our AI-powered educational platform using Flask and React represents a harmonious synergy between backend functionality and frontend user experience. This architecture enables a dynamic, real-time, and secure learning environment, fulfilling the project's objectives of accessibility, personalization, and user engagement.

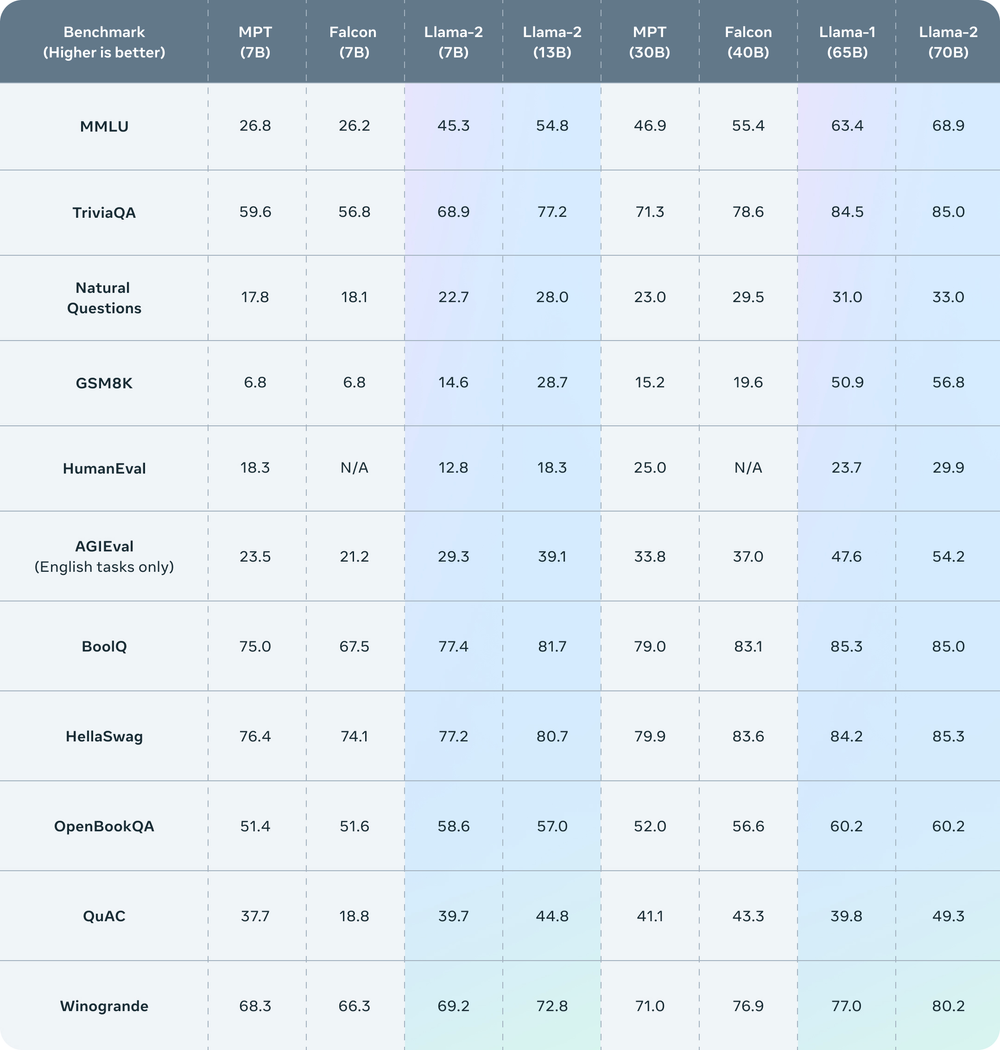
**5. RESULT AND ANALYSIS:**

Presenting the tangible outcomes of our project, this section discusses the results obtained from testing and user feedback. We analyze the impact on personalized learning experiences, accessibility for visually impaired learners, and the overall effectiveness of our AI-powered educational platform. The discussion delves into the implications of the results and potential areas for improvement. We have divided our result and analysis part into three different sections namely

1. Content Generation
2. Text to Speech
3. Web Integration

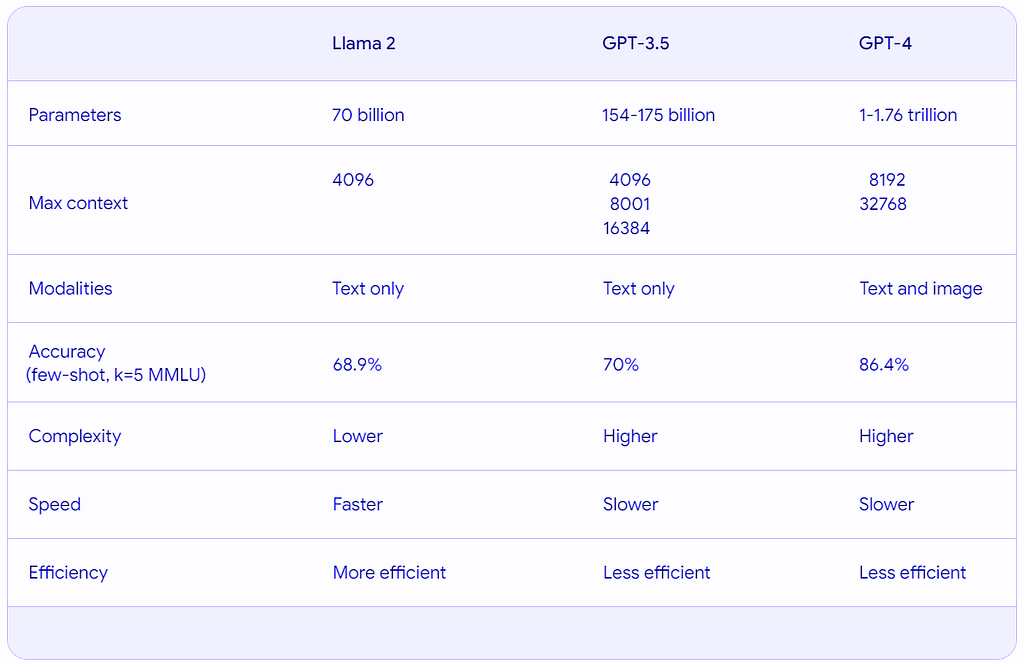
**5.1. Content Generation :**

In this project we have used a pre-trained large language model (LLM) viz. Llama-2 for text generation purpose and outperforms other LLM models the comparative study of other open source models with llama-2 is given below.



**Table.1**: Llama 2 outperforms other open source language models on many external benchmarks, including reasoning, coding, proficiency, and knowledge tests.

The Llama-2 model also gives comparable results with an art-of-state closed source model like GPT-3.5 and it is cost effective whereas GPT-4 gives better results than this. The comparative study of llama-2 with GPT models is given below.



**Table. 2**: Comparison of GPT with llama-2 model on different parameters

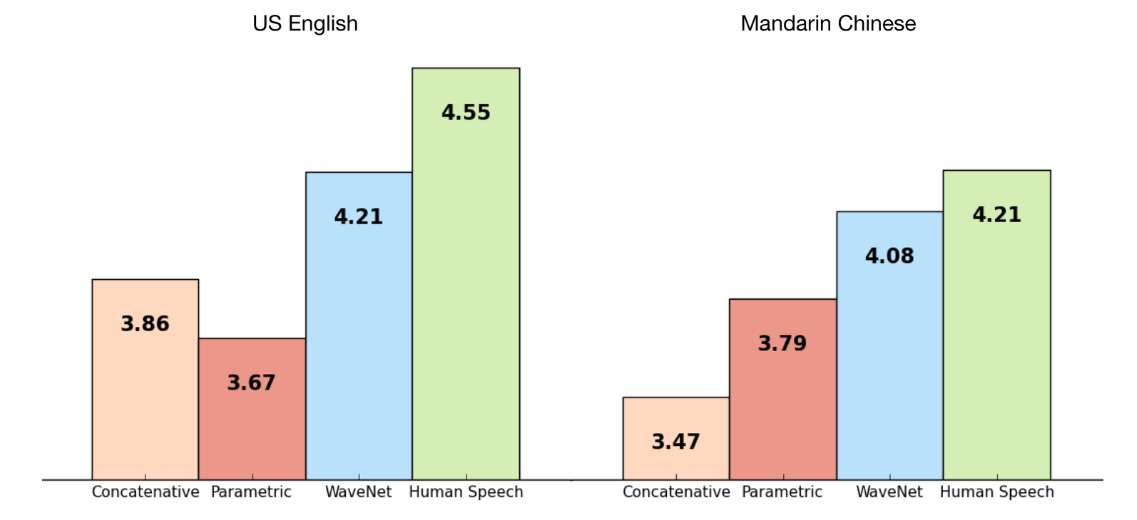
Also, we fine tuned the “*meta-llama/Llama-2-7b-chat-hf*” model with our given dataset so that the context relevance of our generated dataset is maintained with the actual dataset.Hence, we can say that our model will generate good output.

After generating the detailed explanation for each point in the slide we summarized the generated explanation using the “*facebook/bart-large-cnn*” model which we have fine tuned on the generated text for better performance.The results of using the pre-trained Hugging Face text-summarization model, "facebook/bart-large-cnn," and fine-tuning on our dataset are promising. This BART-based model effectively generates comprehensive summaries for each slide, capturing context and relevance. Its versatility extends to tasks like text classification and question answering. The model's effectiveness is attributed to its pre-training on the CNN Daily Mail dataset, ensuring adaptability to various content types. Overall, our approach produces concise, coherent, and contextually relevant text summaries.

**5.2. Google Text-to-Speech Integration :**

In the realm of Text-to-Speech (TTS) technology, the integration of Google Text-to-Speech API stands as a pivotal component in our project, providing an unparalleled auditory experience. Leveraging the power of Google's cutting-edge AI technologies, this API offers a robust set of features that elevate the quality of synthesized speech to near-human levels. The API boasts high-fidelity speech, a critical attribute derived from DeepMind's speech synthesis expertise. This results in voices that exhibit natural intonation, enhancing the overall user experience and making the educational content more engaging.

A WaveNet generates speech that sounds more natural than other text-to-speech systems. It synthesizes speech with more human-like emphasis and inflection on syllables, phonemes, and words. Here is a comparison of different text to speech models available in the market with their accuracy.

****

**Fig.7**: Google’s text-to-speech based on wavenet outperforms other synthetic voices

The y-axis values represent the Mean Opinion Score (MOS) for each voice. Test subjects ranked each voice on a scale of 1-5 according to how much it sounded like natural speech. For more information on MOS scores and WaveNet technology, see the [9] DeepMind WaveNet page.

The decision to integrate Google Text-to-Speech into our project was driven by its comprehensive feature set, wide language support, and the ability to create a unique brand voice. The API's reliability, scalability, and constant updates align with our commitment to providing a state-of-the-art educational experience. In conclusion, the Google Text-to-Speech API emerges as a preferred choice, elevating the auditory component of our AI-powered educational platform to new heights.

**5.3. Web Integration with Flask Server and React :**

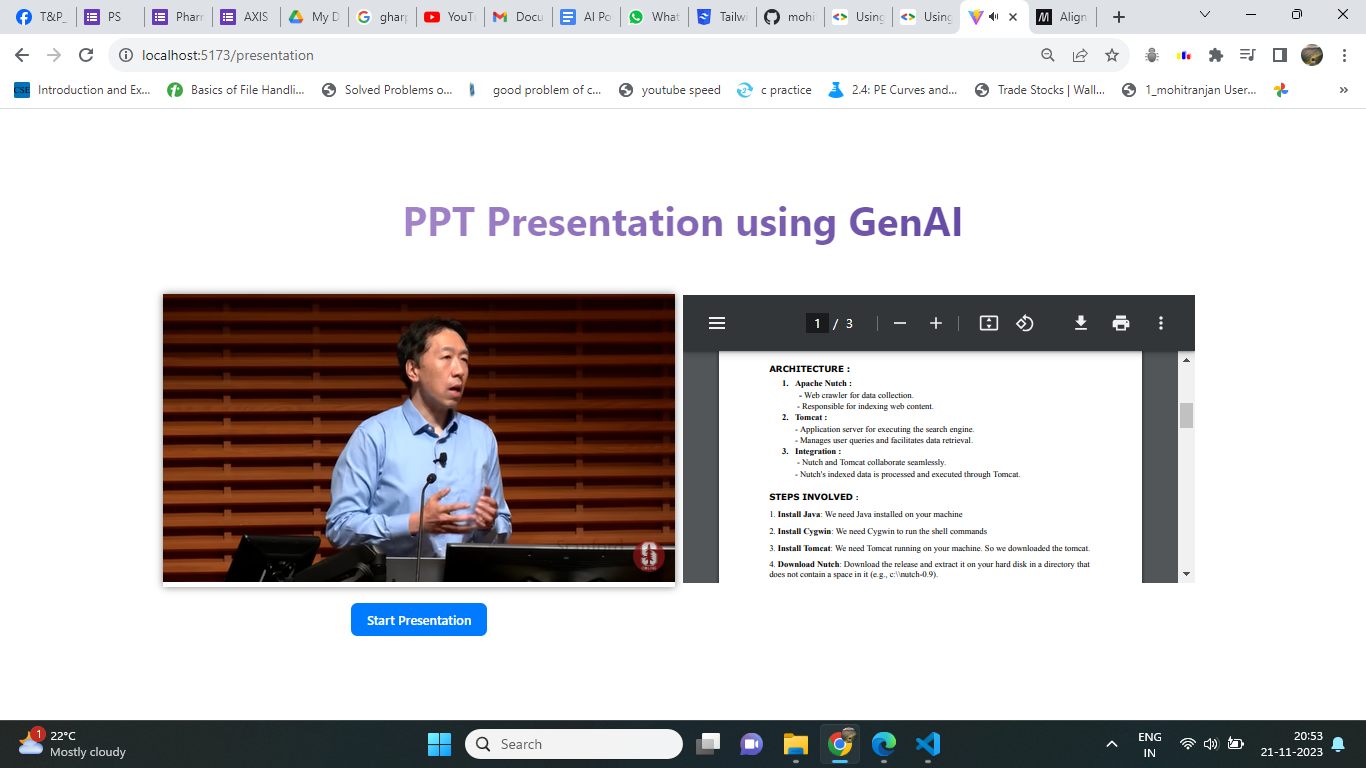
The web application architecture is designed with a client-server model, with Flask serving as the backend server and React handling the frontend interactions. This approach ensures a clean separation of concerns, allowing for scalability, maintainability, and efficient data flow. The Flask server provides RESTful API endpoints, enabling communication between the backend and frontend components.

**5.3.1. React Frontend:**

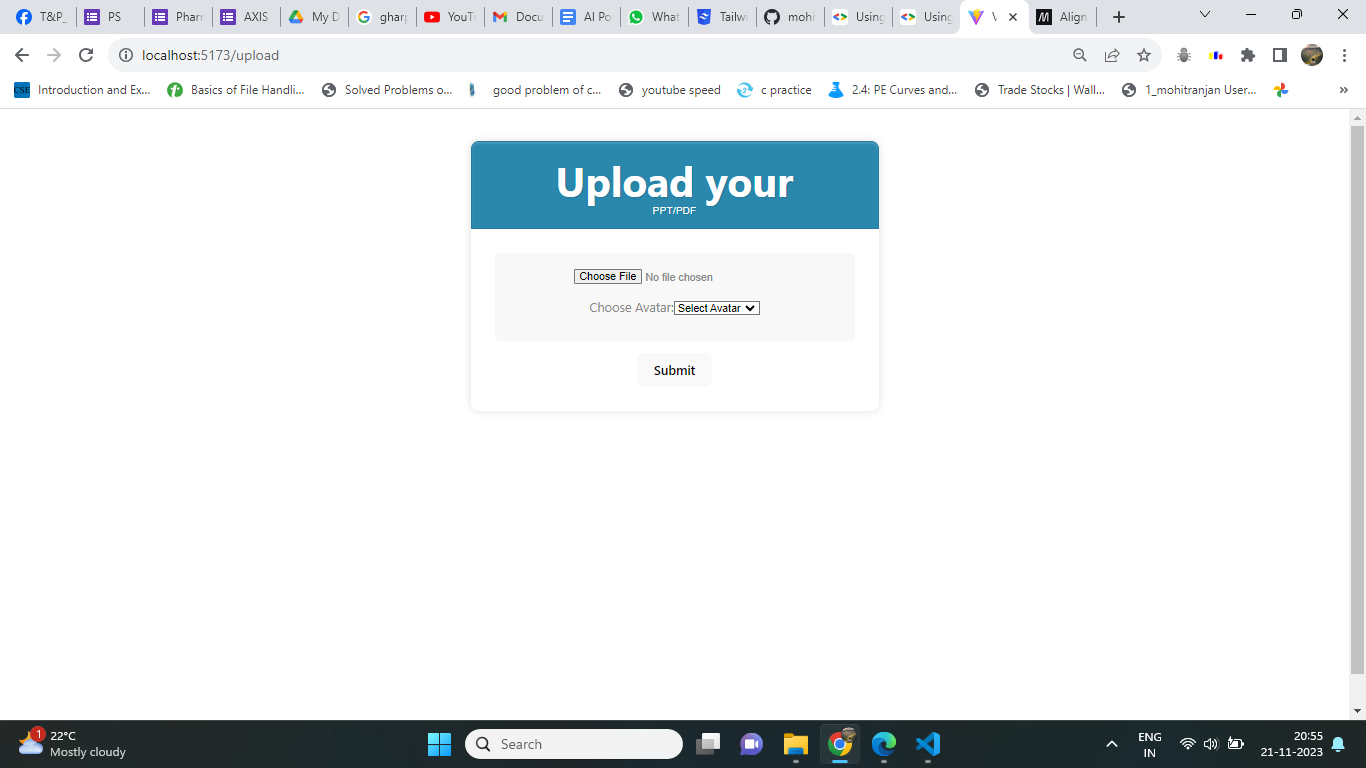
The React frontend provides an interactive and intuitive user interface. Leveraging React's component-based architecture, we design modular components for seamless navigation and an engaging learning experience. The user interface using React framework incorporates following routes -

/home : Landing page which describes the project  
/upload : Route to upload ppt by user and users can choose voice models   
/presentation : Presentation page where user can see the presentation by AI bot

**5.3.2. SAMPLES :**



**Fig. 8** : presentation route



**Fig. 9:** upload route to upload ppt

**5.3.3. Flask Server:**

Our Flask server acts as the backbone of the web application, handling key functionalities such as authentication, data processing, and communication with external APIs. Different endpoints are meticulously crafted to cater to specific functionalities, ensuring a modular and organized codebase. For instance,

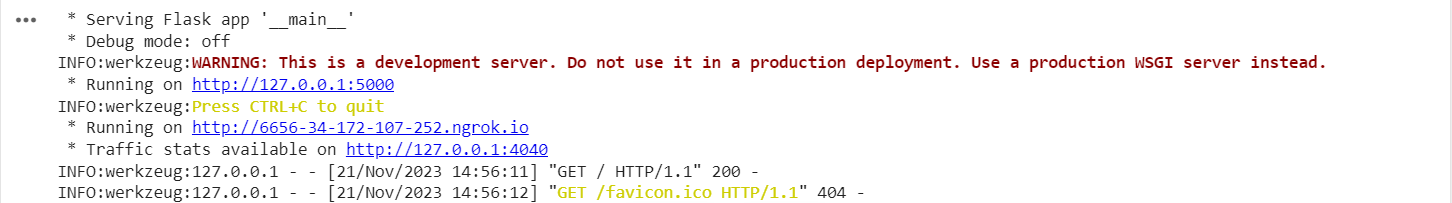
/generate\_audio : generates the text to speech which is a post request.  
/generate\_content : generates the content from the ppt which is also a post request.  
/ : get endpoint to check the flask server is working or not

The Flask server acts as the bridge between the user interface and the AI-powered backend.

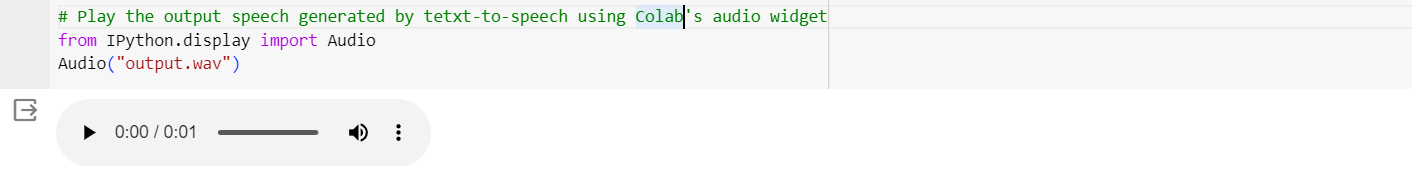
**5.3.4. SAMPLES :**

****

**Fig 10.**: Flask server codebase

****

**Fig 11.**: Image of Flask server running

****

**Fig.12**: Output of text-to-speech model

**6. CONCLUSION AND FUTURE SCOPE**

In conclusion, our innovative AI-powered educational platform, integrating advanced technologies such as text summarization, script generation, and Google Text-to-Speech API, presents a transformative solution to the limitations of traditional PowerPoint-based learning. By prioritizing personalization, accessibility, and user engagement, we have created a dynamic tool that bridges educational gaps, offering a cost-effective, 24/7 accessible teacher-like experience. The successful web integration with Flask and React ensures a seamless and interactive interface. The choice of Google Text-to-Speech API enriches the auditory component with high-fidelity, near-human voices, enhancing the overall educational journey. Our project marks a significant stride towards inclusive and adaptive learning environments.

AI Powered Voice-Guided Presentation: Revolutionizing PPT Explanations can benefit various applications in fields like online education , news anchoring using AI robots in custom voice, reduce teachers workload etc. In the future we can think about how we can integrate lip syncing where we are able to deliver a ppt explanation in any custom voice with video in which that particular person's video will be played and it will seem like the person's real video. And in future we will work on how we can increase preciseness and accuracy of our ppt explanation by trying different models and doing hypertuning of existing model

**7. REFERENCES**

[1] Touvron, Hugo, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov et al. "Llama 2: Open foundation and fine-tuned chat models." *arXiv preprint arXiv:2307.09288* (2023).

[2] Pratiksha, U., Jadhav. (2023). AI Text Summarization System. International Journal For Science Technology And Engineering, 11(5), 916-919. doi: 10.22214/ijraset.2023.51481

[3] J., Deny. (2023). Inshort Text Summarization of News Article. 1104-1108. doi: 10.1109/ICICCS56967.2023.10142549

[4] Madhur, Yadav. (2023). Summarization Of Text Using Natural Language Processing. International journal of scientific and research publications, 13(4) doi: 10.29322/ijsrp.13.04.2023.p13611

[5] Peggy, Tang., Junbin, Gao., Lei, Zhang., Zhiyong, Wang. (2023). Efficient and Interpretable Compressive Text Summarisation with Unsupervised Dual-Agent Reinforcement Learning. arXiv.org, abs/2306.03415 doi: 10.48550/arXiv.2306.03415

[6] (2022). A Systematic Literature Review on Text Generation Using Deep Neural Network Models. IEEE Access, 10, 53490-53503. doi: 10.1109/access.2022.3174108

[7] Nikolay, Mikhaylovskiy. (2023). Long Text Generation Challenge. arXiv.org, abs/2306.02334 doi: 10.48550/arXiv.2306.02334

[8] Felix, Faltings., M, Galley., Baolin, Peng., Kianté, Brantley., Weixin, Cai., Yizhe, Zhang., Jianfeng, Gao., Bill, Dolan. (2023). Interactive Text Generation. arXiv.org, abs/2303.00908 doi: 10.48550/arXiv.2303.00908

[9] Sheng, Wang., Yanping, Wang. (2022). Researches advanced in data-to-text generation based on neural networks. 12348, 123480W-123480W. doi: 10.1117/12.2641838

[10] (2022). Research on Text Generation of Medical Intelligent Question and Answer Based on Bi-LSTM and Neural Network Technology. doi: 10.1109/icis54925.2022.9882349

[11]Allen, Jonathan, M. Sharon Hunnicutt, Dennis H. Klatt, Robert C. Armstrong, and David B. Pisoni. *From text to speech: The MITalk system*. Cambridge University Press, 1987.

[12] D. Sasirekha and E. Chandra, "Text to speech: a simple tutorial," International Journal of Soft Computing and Engineering (IJSCE), vol. 2, no. 1, pp. 275-278, 2012.

[13] M. Jeong, H. Kim, S. J. Cheon, B. J. Choi, and N. S. Kim, "Diff-tts: A denoising diffusion model for text-to-speech," arXiv preprint arXiv:2104.01409, 2021.

[14]Ren, Yi, Yangjun Ruan, Xu Tan, Tao Qin, Sheng Zhao, Zhou Zhao, and Tie-Yan Liu. "Fastspeech: Fast, robust and controllable text to speech." *Advances in neural information processing systems* 32 (2019).

[15]Łańcucki, Adrian. "Fastpitch: Parallel text-to-speech with pitch prediction." In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6588-6592. IEEE, 2021.

[16] Y. Renet al., "Fastspeech: Fast, robust and controllable text to speech," Advances in neural information processing systems, vol. 32, 2019.

[17]Y.-C. Huang and L.-C. Liao, "A Study of Text-to-Speech (TTS) in Children's English Learning," Teaching English with Technology, vol. 15, no. 1, pp. 14-30, 2015.

[18] M. Cohn and G. Zellou, "Perception of concatenative vs. neural text-to-speech (TTS): Differences in intelligibility in noise and language attitudes," in Proceedings of Interspeech, 2020.

[19]“Standard and WaveNet voices | Cloud Text-to-Speech Documentation,” Google Cloud. <https://cloud.google.com/text-to-speech/docs/wavenet>

[20]“Introducing Cloud Text-to-Speech powered by DeepMind WaveNet technology,” Google Cloud Blog. <https://cloud.google.com/blog/products/ai-machine-learning/introducing-cloud-text-to-speech-powered-by-deepmind-wavenet-technology>