



Shri Vile Parle Kelavani Mandal's

DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING

(Autonomous College Affiliated to the University of Mumbai)

NAAC Accredited with "A" Grade (CGPA : 3.18)



SLIDESTER

Submitted in partial fulfillment of the requirements
of the course Innovative Product Development
under

T. Y. B. Tech. Artificial Intelligence and Machine Learning

By

Hatim Mullajiwala 60017220038

Antara Kadam 60018220099

Haadi Rakhanghi 60017220037

Guide

Dr. Nilesh Rathod

(Assistant Professor)

Mrs. Komal Patil

(Assistant Professor)

University of Mumbai

2022-2023

CERTIFICATE

This is to certify that the project entitled “**SlideSter**” is a bonafide work of **Haadi Rakhang**
(60017220037) submitted to the **Department of Artificial Intelligence and Machine Learning**
in partial fulfillment of the requirement for the course of Innovative Product Development.

(Dr. Nilesh Rathod)
Internal Guide

(Dr. Aruna Gawade)
Head of the Department

(Dr. Hari Vasudevan)
Principal

CERTIFICATE

This is to certify that the project entitled “**SlideSter**” is a bonafide work of **Antara Kadam (60018220099)** submitted to the **Department of Artificial Intelligence and Machine Learning** in partial fulfillment of the requirement for the course of Innovative Product Development.

(Dr. Nilesh Rathod)
Internal Guide

(Dr. Aruna Gawade)
Head of the Department

(Dr. Hari Vasudevan)
Principal

CERTIFICATE

This is to certify that the project entitled “**SlideSter**” is a bonafide work of **Hatim Mullajiwala (60017220038)** submitted to the **Department of Artificial Intelligence and Machine Learning** in partial fulfillment of the requirement for the course of Innovative Product Development.

(Dr. Nilesh Rathod)
Internal Guide

(Dr. Aruna Gawade)
Head of the Department

(Dr. Hari Vasudevan)
Principal

Project Report Approval for Innovative Product Development.

This project report entitled *of SLIDESTER* by **Haadi Rakhangi (60017220037)**, **Hatim Mullajiwala (60017220038)**, **Antara Kadam (60018220099)** is approved for the course of Innovative Product Development.

Examiners

1. _____

2. _____

3. _____

Date:

Place: Mumbai

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Hatim Mullajiwala (60017220038)

Antara Kadam (60018220099)

Haadi Rakhanghi (60017220037)

Date:

Abstract

Slidester, positioned at the forefront of presentation innovation, integrates advanced Natural Language Processing (NLP), speech recognition, and generative AI to redefine the presentation creation process. Through the utilization of Dynamic Slide Suggestions, Slidester facilitates a structured initiation by recommending slides based on input topics, thereby streamlining the commencement of presentations. The platform dynamically generates content-rich slides, elevating visual appeal and augmenting informational depth. The Image Generation feature further automates the process by retrieving relevant visuals based on the content of each slide. A notable attribute of Slidester is its capacity to transcend the constraints of traditional Generative AI models, which often remain static with information available only up to their training data. In contrast, Slidester's generative AI model can access the latest information from the internet, ensuring that presentations are enriched with the most up-to-date and pertinent data. The Internet-Powered Content feature provides access to real-time data, thereby guaranteeing that presentations remain current. Slidester embraces Document Integration, enabling the seamless inclusion of external data. The platform also offers downloadable formats and creative templates, enhancing versatility and establishing Slidester as an indispensable tool for impactful, automated, and visually compelling presentations. The overarching goal of the project is to enhance productivity by automating the presentation creation process and eliminating time-consuming tasks. Slidester caters to a diverse audience, including students, professionals, and individuals seeking to create impactful presentations efficiently. Leveraging NLP techniques, speech recognition, and image generation, Slidester aspires to redefine presentation creation as an intuitive, enjoyable, and efficient experience. The platform provides a powerful tool that simplifies and expedites the process of creating presentations. Its seamless integration of NLP, speech recognition, image generation, and real-time internet data access ensures a user-friendly platform, empowering individuals to effortlessly curate well-structured and visually appealing presentations. Slidester represents a significant advancement in the realm of presentation creation, promising to be the future of hassle-free and dynamic presentations.

Table of Contents

| Chapter | Contents | Page No. |
|----------------|---------------------------------|-----------------|
| I | Introduction | I |
| | 1.1 Aim of the Project | |
| | 1.2 Motivation | |
| | 1.3 Objective | |
| II | Literature Survey | II |
| III | Problem Description | III |
| | 3.1 Problem Statement | |
| | 3.2 Scope of the Project | |
| IV | Proposed Design | IV |
| V | Results and Discussions | VI |
| VI | Conclusion | VII |
| VII | References | IX |

List of Figures

| Fig. No. | Figure Caption | Page No. |
|----------|--|----------|
| 4.1 | Architecture Diagram | 27 |
| 5.1 | Home Page | 29 |
| 5.2 | Login Page | 30 |
| 5.3 | Register Page | 30 |
| 5.4 | Profile Page | 31 |
| 5.5 | Subject Of The Ppt | 32 |
| 5.6 | Select And Confirm Sub-Titles And No Of Points | 32 |
| 5.7 | PPT Generated | 33 |
| 5.8 | Edit The Points Or Titles | 33 |
| 5.9 | Generated PPT | 33 |
| 5.10 | Image Generated On the Slide | 34 |
| 5.11 | Assistant Api Integration | 34 |

List of Abbreviations

| Sr. No. | Abbreviation | Expanded form |
|---------|--------------|--|
| 1 | PPT | Power Point Presentation |
| 2 | LCM - LoRA | Latent Consistency Model – Low Rank Adaptation |
| 3 | NLP | Natural Language Processing |

1. Introduction

In the dynamic realm of presentations, Slidester emerges as a trailblazer, orchestrating a paradigm shift in the crafting and communication of narratives. As the intersection of technology and communication becomes increasingly crucial, Slidester stands as a transformative force, reshaping the very essence of the presentation creation process. This innovative platform seamlessly integrates advanced technologies, including Natural Language Processing (NLP), speech recognition, and generative AI, aiming to redefine the user experience and elevate the art of presentation.

At the core of Slidester's revolutionary approach lies its pioneering feature: Dynamic Slide Suggestions. This groundbreaking innovation redefines the initiation phase, offering users a structured starting point by recommending slides based on input topics. This not only streamlines the often daunting task of commencing a presentation but also underscores Slidester's commitment to understanding and anticipating user needs. It transcends the role of a mere presentation tool, evolving into a creative companion that guides users through the inception of their presentation journey.

Nevertheless, Slidester's prowess extends beyond initiation. The platform dynamically generates content-rich slides, enhancing visual appeal while deepening the informational reservoir of each presentation. The infusion of Image Generation further automates the process, ensuring that relevant visuals seamlessly align with the content of each slide. This harmonious fusion of technology and creativity positions Slidester as a transformative force, where visual and textual elements coalesce effortlessly.

What propels Slidester into a league of its own is its liberation from temporal constraints. While traditional Generative AI models often remain "stuck in time," Slidester's generative AI model transcends these limitations by accessing the latest information from the internet. This ensures that presentations are not only visually engaging but also enriched with the most up-to-date and relevant data, reflecting the ever-evolving nature of the subject matter.

Slidester's commitment to real-time relevance is further exemplified through Internet-Powered Content, granting users access to live data. This ensures that presentations remain current, reflecting

the latest trends, statistics, and developments. The platform seamlessly embraces Document Integration, allowing for the inclusion of external data sources and reinforcing its adaptability to diverse information landscapes.

As Slidester strides towards redefining the fabric of presentation creation, it does so with the user's experience at its forefront. The provision of downloadable formats and creative templates adds a layer of versatility, making Slidester an indispensable tool for individuals across diverse domains. From students navigating the academic realm to professionals seeking impactful business presentations, Slidester is not merely a tool; it is a transformative force promising an intuitive, enjoyable, and efficient experience. It marks a significant leap forward, heralding the future of hassle-free and dynamic presentations in a technology-driven era.

1.1 Aim of the Project

The Slidester project's primary objective is to revolutionize the landscape of presentation creation, surpassing traditional methodologies and ushering in a new era of efficiency, creativity, and user-centric design. At its essence, Slidester aims to redefine how narratives are crafted and communicated by seamlessly integrating advanced technologies, including Natural Language Processing (NLP), speech recognition, and generative AI.

Central to this innovative initiative is the development and implementation of Dynamic Slide Suggestions. This groundbreaking feature aims to streamline the initiation phase of presentation creation by furnishing users with a structured starting point based on input topics. In doing so, Slidester not only simplifies the often daunting task of starting a presentation but also underscores a commitment to understanding and anticipating user needs, positioning itself as a creative companion in the presentation journey.

Moving beyond initiation, Slidester endeavors to dynamically generate content-rich slides that enhance visual appeal and deepen the informational reservoir of each presentation. The incorporation of Image Generation further automates the process, ensuring seamless alignment between relevant visuals and slide content. This harmonious integration of technology and creativity distinguishes Slidester as a transformative force, where visual and textual elements coalesce effortlessly.

An integral aspect of the project's aim is to free Slidester from temporal constraints. Unlike traditional Generative AI models, Slidester's model accesses the latest information from the internet, ensuring that presentations are not only visually engaging but also enriched with the most up-to-date and relevant data, reflecting the ever-evolving nature of the subject matter.

Slidester also strives for real-time relevance through Internet-Powered Content, granting users access to live data. This commitment ensures that presentations remain current, reflecting the latest trends, statistics, and developments. The platform's embrace of Document Integration further reinforces its adaptability to diverse information landscapes, acknowledging the importance of incorporating external data sources.

As Slidester advances toward redefining the fabric of presentation creation, the user's experience

remains at the forefront. The provision of downloadable formats and creative templates adds a layer of versatility, making Slidester an indispensable tool for individuals across various domains. Whether it's students navigating the academic realm or professionals delivering impactful business presentations, Slidester aspires to be more than a tool—a transformative force promising an intuitive, enjoyable, and efficient experience, marking a significant leap forward in the technology-driven era of dynamic presentations.

1.2 Motivation

The inspiration for Slidester stems from our aspiration to create something revolutionary and distinctive. When tasked with selecting a project topic, our group sought an endeavor that was innovative and unexplored. Given our profound interest in AI and machine learning, we eagerly embraced the opportunity to contribute to a project that could truly make a difference.

In the initial stages of ideation, we scoured online resources for inspiration, yet none resonated with the freshness we desired. Our collective enthusiasm for AI led us to a realization - the projects we encountered didn't align with solving real-world problems. It was during the preparation of a presentation about our chosen idea that the breakthrough occurred. The cumbersome and time-intensive nature of creating presentations became glaringly evident. We felt that valuable time was being squandered on this process, time that could be better spent generating innovative ideas or delving into subjects we were passionate about.

Enter "Slidester." The concept crystallized in our minds, driven by the vision of developing a system that alleviates the tedium and time investment associated with crafting presentations. With Slidester, individuals can articulate their ideas verbally, leaving the system to handle the intricacies - from generating slides to incorporating content and even creating relevant images. The primary objective is to streamline the process of presentation creation, rendering it more accessible and enjoyable for everyone.

At the heart of Slidester lies the motivation to empower individuals to concentrate on the essence of their message, liberated from the burdensome intricacies of presentation design. We aim to furnish them with a tool that not only saves time but also enhances the overall efficiency of the process. Whether one is a student, a professional, or anyone tasked with creating presentations, Slidester aspires to be the preferred solution for delivering hassle-free and dynamic presentations.

1.3 Objective

The overarching objective of Slidester is to redefine the landscape of presentation creation, positioning itself as a transformative force that seamlessly integrates advanced technologies, such as Natural Language Processing (NLP), speech recognition, and generative AI. Building upon this foundation, our primary goal is to enhance the entire presentation creation process, making it intuitive, enjoyable, and efficient for users across diverse domains.

At the core of Slidester's mission is the aim to revolutionize the initiation phase through its groundbreaking feature: Dynamic Slide Suggestions. This innovation provides users with a structured starting point, recommending slides based on input topics and streamlining the often complex task of beginning a presentation. Slidester goes beyond being a mere tool; it aspires to be a creative companion that understands and anticipates user needs, guiding them through the inception of their presentation journey.

Expanding on this, Slidester seeks to dynamically generate content-rich slides, augmenting visual appeal and enriching the informational reservoir of each presentation. The integration of Image Generation further automates this process, ensuring seamless alignment between relevant visuals and the content of each slide. This harmonious fusion of technology and creativity sets Slidester apart as a transformative force where visual and textual elements coalesce effortlessly.

A distinctive objective is Slidester's liberation from temporal constraints. Unlike traditional Generative AI models, Slidester's model accesses the latest information from the internet, ensuring that presentations are not only visually engaging but also enriched with the most up-to-date and relevant data, reflecting the ever-evolving nature of the subject matter.

Slidester commits to real-time relevance through Internet-Powered Content, granting users access to live data. This ensures that presentations remain current, reflecting the latest trends, statistics, and developments. The platform's embrace of Document Integration further reinforces its adaptability to diverse information landscapes, acknowledging the importance of incorporating external data sources.

As Slidester strides towards redefining the fabric of presentation creation, it remains committed to enhancing the user experience. The provision of downloadable formats and creative templates adds a layer of versatility, making Slidester an indispensable tool for individuals across diverse domains. Whether navigating the academic realm or delivering impactful business presentations, Slidester is not just a tool but a transformative force promising an intuitive, enjoyable, and efficient experience, marking a significant leap forward in the technology-driven era of dynamic presentations.

2. Literature Survey

In a paper titled 'Development Of GUI For Text-To-Speech Recognition Using NLP' [1], the authors explore the application of Natural Language Processing (NLP) in the development of a Text-to-Speech synthesizer. The synthesizer, created as a user-friendly application, utilizes NLP for analyzing and processing text, followed by Digital Signal Processing (DSP) technology to convert the processed text into a synthesized speech representation. The resultant tool allows users to input text, which is then transformed into spoken words and can be saved as an mp3 file.

A research paper from [2] focuses on real-time speech-to-text conversion, presenting a system that converts spoken words into text in a manner reflective of the user's pronunciation. The authors designed a real-time speech recognition system tested in a noisy environment. Utilizing a bidirectional nonstationary Kalman filter design enhanced the system's capabilities. The paper emphasizes the real-time aspect of speech-to-text conversion, employing a database created by the authors for flexibility and comparing its accuracy with the HMM-based speech recognition system.

The research paper [3] introduces an audio-visual summarizer designed to summarize presentations. By combining audio evidence of the speaker's delivery with visual images of presentation slides or handwritten materials, the device creates a comprehensive summary logged to a remote database server. The prototype employs a Raspberry Pi, camera, and microphone, utilizing fast RCNN models, Open Source Computer Vision, Google Speech Recognition, and NLP concepts for text detection, extraction, and summarization.

In [4], the paper details a deep learning-based architecture for semantically consistent image generation called recurrent convolutional generative adversarial network (RC-GAN). Trained on the Oxford-102 flowers dataset, the proposed model demonstrates advancements in text-to-image modeling, converting visual notions from words to pixels. The model's performance is evaluated using an inception score and PSNR, showcasing its ability to generate realistic photos of flowers from given captions.

Another paper from [5] explores the development of a Text-to-Speech synthesizer using NLP and Digital Signal Processing (DSP) technology. Similar to the first mentioned paper, this synthesizer transforms text into spoken words, providing a user-friendly application for converting inputted text

into synthesized speech that can be saved as an mp3 file.

A paper from [6] proposes a system to automatically generate presentation slides from textual data. Unlike other tools that focus on formatting, this system utilizes NLP rules to classify data and generate slides, reducing the presenter's time and effort while incorporating only the essential points related to the topic.

The paper from [7] introduces an end-to-end deep learning approach for recognizing English or Mandarin Chinese speech. This approach, leveraging neural networks and HPC techniques, replaces hand-engineered components with neural networks, enabling the system to handle diverse speech scenarios, including noisy environments, accents, and different languages.

Finally, the paper published in 2022 [8] introduces Imagen, a model leveraging large transformer language models for text understanding and diffusion models for high-fidelity image generation. Pretrained on text-only corpora, Imagen excels in encoding text for image synthesis, as evidenced by comparisons with other models in terms of sample quality and image-text alignment. The paper also introduces DrawBench for a comprehensive assessment of text-to-image models.

Table no:1 Literature Survey

| SR NO. | PAPER NAME -TITLE | PUBLICATIONS | AUTHOR/YEAR | FINDINGS |
|---------------|--|---------------------|---|---|
| 1 | Development Of Gui For Text-To-Speech Recognition Using Nlp | IEEE | Partha Mukherjee, Soumen Santra, Subhajit Bhowmick, Ananya Paul, Pubali Chatterjee, Arpan Deyasi | Natural language processing is a widely used technique by which systems can understand the instructions for manipulating text or speech. In the present paper, a Text-to- speech synthesizer is developed that converts text into spoken word, by analysing and processing it using NLP and then using Digital Signal Processing (DSP) technology to convert this processed text into synthesized speech representation of the text. Here we developed a useful text-to- speech synthesizer in the form of a simple application that converts inputted text into synthesized speech and reads out to the user which can then be saved as an mp3 file. |
| 2 | A Real Time Speech To Text Conversion System Using Bidirectional Kalman Filter In Matlab | ICACCI | Neha Sharma Shipra Sardana | A real time speech to text conversion system converts the spoken words into text form exactly in the similar way that the user pronounces. We created a real time speech recognition system that was tested in real time noiseous environment. We used the design of a bidirectional nonstationary Kalman filter to enhance the ability of this Real time speech recognition system. Real time speech to text conversion system introduces conversion of the uttered words instantly after the utterance. speech recognition system. We have used our own created database for its flexibility and TIDIGIT database for its accuracy comparison with the HMM based speech recognition system. MFCC features of speech sample were calculated and words were distinguished according to the feature matching of each sampled word. |

| | | | | |
|---|--|------|---|--|
| 3 | A Cost-Effective Audio-Visual Summarizer for Summarization of Presentations and Seminars | IEEE | Anmol Bhat, Aneesh C Rao, Anirudh Bhaskar, Adithya V, Prof. Pratiba D | <p>This paper delves into the implementation of an audio-visual summarizer that achieves the aforementioned motive. With audio evidence on the speaker's delivery, paired with visual images of PowerPoint slides or handwritten material that is presented in these seminars, this device provides a smart solution of summarizing the entire presentation and logging the summary to a remote database server from where it is accessed through a user-end software application. The prototype comprises a Raspberry Pi coupled with a camera and a microphone. The prototype uses a fast RCNN model for text detection, Open Source Computer Vision for text extraction, Google Speech Recognition and NLP concepts for generating the summarized data.</p> |
| 4 | Text-to-Image Generation Using Deep Learning | MDPI | Sadia Ramzan, Muhammad Munwar Iqbal, Tehmina Kalsum. | <p>A deep learning-based architecture for semantically consistent image generation: recurrent convolutional generative adversarial network (RC-GAN). RC-GAN successfully bridges the advancements in text and picture modelling, converting visual notions from words to pixels. The proposed model was trained on the Oxford-102 flowers dataset, and its performance was evaluated using an inception score and PSNR. The experimental results demonstrate that our model is capable of generating more realistic photos of flowers from given captions, with an inception score of 4.15 and a PSNR value of 30.12 dB, respectively. In the future, we aim to train the proposed model on multiple datasets.</p> |

| | | | | |
|---|--|----------|---|---|
| 5 | Content Based Internet Search Engine For Analysis And Archival Of Mpeg-1 Compressed Newsfeeds | IEEE | Odej K.40 , Gerhard R. Joubert | Natural language processing is a widely used technique by which systems can understand the instructions for manipulating text or speech. In the present paper, a Text-to- speech synthesizer is developed that converts text into spoken word, by analysing and processing it using NLP and then using Digital Signal Processing (DSP) technology to convert this processed text into synthesized speech representation of the text. Here we developed a useful text-to- speech synthesizer in the form of a simple application that converts inputted text into synthesized speech and reads out to the user which can then be saved as an mp3 file. |
| 6 | Enhancing Automatic PPT Generation Technique through NLP for Textual Data | IJARCCCE | Pooja Belote, Sonali Bidwai, Snehal Jadhav, Pradnya Kapadnis4, Nakul Sharma5 | Proposed system come with an idea of automatically generate presentation slides from textual data. The slides mostly incorporate of only the important points related to the topic. There are various tools and systems are available in the market which only address with formatting of the slides but not the content. But it having performance issues regarding of sentence extraction, so this paper proposes an idea of automatically ppt generation from text. This will finally help in reducing a great amount of the presenter's time and efforts. The intended system works on NLP rules to classify data for the desired slides. |

| | | | | |
|---|---|--------------------------------|---|--|
| 7 | End-to-End Speech Recognition in English and Mandarin | IEEE | Dario Amodei, Rishita Anubhai, Eric Battenberg, Carl Case, Jared Casper, Bryan Catanzaro, Jingdong Chen, Mike Chrzanowski, Adam Coates, Greg Diamos, Erich Elsen, Jesse Engel, Linxi Fan, Christopher Fougner, Tony Han, Awni Hannun, Billy Jun, Patrick LeGresley, Libby Lin, Sharan Narang | <p>We show that an end-to-end deep learning approach can be used to recognize either English or Mandarin Chinese speech—two vastly different languages. Because it replaces entire pipelines of hand- engineered components with neural networks, end-to-end learning allows us to handle a diverse variety of speech including noisy environments, accents and different languages. Key to our approach is our application of HPC techniques, resulting in a 7x speedup over our previous system.</p> <p>Hence, using a technique called Batch Dispatch with GPUs in the data center, we show that our system can be inexpensively deployed in an online setting, delivering low latency when serving users at scale.</p> |
| 8 | Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding | Google Research, Brain Team | Chitwan Saharia , William Chan, Saurabh Saxena , Lala Li†, Rapha Gontijo Lopes, Tim Salimans, Jonathan Ho , David J Fleet , Mohammad Norouzi. | <p>Imagen builds on the power of large transformer language models in understanding text and hinges on the strength of diffusion models in high-fidelity image generation. Our key discovery is that generic large language models, pretrained on text-only corpora, are surprisingly effective at encoding text for image synthesis. To assess text-to-image models in greater depth, we introduce DrawBench, we compare Imagen with recent methods including and find that human raters prefer Imagen over other models in side-by-side comparisons, both in terms of sample quality and image-text alignment.</p> |

2.1 Survey Outcome:

The literature survey conducted for the Slidester project has unveiled critical insights and existing works within the domains of natural language processing (NLP), speech recognition, and image generation – pivotal components integral to the success of the project.

One key observation from the survey is the development of text-to-speech synthesizers employing NLP and digital signal processing technology. This methodology serves as a foundational concept for Slidester's voice input feature, allowing users to verbally provide slide titles and subsequently generate dynamic presentations. Furthermore, the exploration of real-time speech-to-text conversion systems in the literature survey resonates with Slidester's incorporation of speech recognition capabilities, facilitating effortless content input via spoken language.

Deep learning-based architectures, particularly recurrent convolutional generative adversarial networks (RC-GAN), have been discussed in the survey in the context of image generation. Such models find applicability in Slidester's image generation feature, contributing to the automatic creation of relevant images for each slide's content, thereby enhancing the overall visual appeal of presentations.

The literature survey also sheds light on methodologies related to automatic presentation slide generation from textual data, aligning seamlessly with Slidester's overarching goal of streamlining the presentation creation process. This approach emphasizes the automation of content generation, offering a potential avenue for Slidester to improve efficiency.

Moreover, the survey delves into end-to-end deep learning approaches for speech recognition, emphasizing the adaptability of these systems across various speech languages and environmental conditions. This particular aspect aligns perfectly with Slidester's vision of integrating speech recognition technology, ensuring a versatile and effective user experience.

Finally, the survey underscores the effectiveness of large transformer language models in comprehending text and their successful application in text-to-image synthesis. This insight is directly relevant to Slidester's utilization of NLP techniques to efficiently generate presentation slides from textual data, showcasing a commitment to leveraging cutting-edge technologies for enhanced performance.

3. Problem Description

3.1 Problem Statement:

The traditional methods of creating presentations pose significant challenges due to their time-consuming and tedious nature. Users often find themselves sifting through vast amounts of information, searching for suitable images, and selecting appropriate templates, which hinders productivity and creativity. This labor-intensive process can lead to frustration and a misallocation of valuable time that could be better utilized for more meaningful tasks. Additionally, manual data entry for each slide can be error-prone, resulting in suboptimal presentation quality.

The lack of efficient tools that streamline the presentation creation process leaves users grappling with the complexity of organizing content, selecting visuals, and maintaining audience engagement. The absence of an automated and user-friendly platform results in a persistent problem for individuals from various domains, including students, educators, professionals, and entrepreneurs.

3.2 Scope of the Project

The scope of the Slidester project is expansive and forward-thinking, with a core focus on revolutionizing the presentation creation process through the seamless integration of advanced technologies. This comprehensive and user-friendly platform aspires to cater to the diverse needs of individuals across various domains, setting itself as a transformative force in the landscape of dynamic presentations.

Central to the project's scope is the groundbreaking feature of Dynamic Slide Suggestions, redefining the initiation phase by recommending slides based on user input topics. This innovative approach streamlines the often complex task of beginning a presentation, setting the stage for a user-centric experience that goes beyond traditional methodologies.

The project extends its scope by dynamically generating content-rich slides that enhance both visual appeal and informational depth. The integration of Image Generation automates the process,

ensuring that relevant visuals seamlessly align with the content of each slide. This harmonious blend of technology and creativity is a cornerstone of Slidester's scope, offering users a transformative experience where visual and textual elements coalesce effortlessly.

Crucially, Slidester liberates itself from temporal constraints, accessing the latest information from the internet to ensure presentations are not only visually engaging but also enriched with up-to-date and relevant data. This forward-looking approach reflects the project's commitment to adapting to the ever-evolving nature of information.

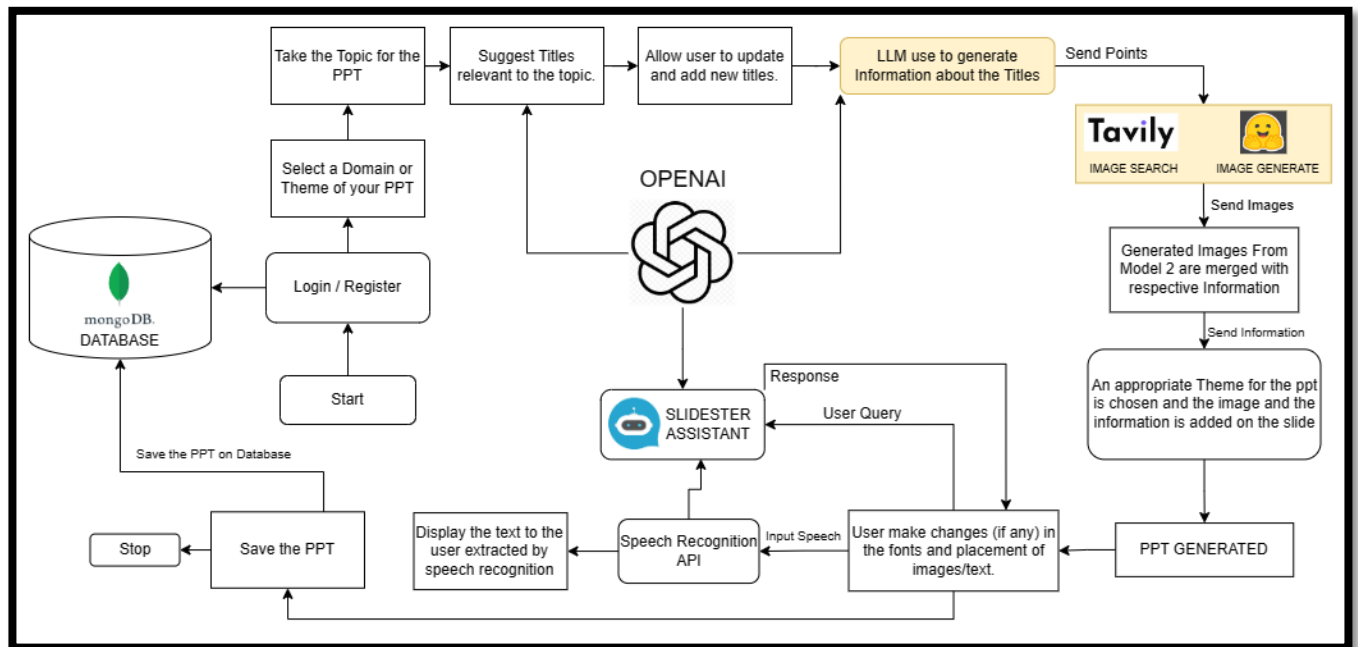
The scope of Slidester encompasses a commitment to real-time relevance through Internet-Powered Content, providing users with access to live data. This ensures that presentations remain current, reflecting the latest trends, statistics, and developments. The project's embrace of Document Integration further reinforces its adaptability to diverse information landscapes, acknowledging the importance of incorporating external data sources.

The project's versatile scope caters to a wide user base, including students, professionals, educators, and entrepreneurs. The provision of downloadable formats and creative templates adds a layer of versatility, making Slidester an indispensable tool for individuals across diverse domains. This multifaceted approach positions Slidester as a groundbreaking solution emphasizing efficiency, creativity, and user-centric design in the realm of presentation creation.

The scope of Slidester extends beyond mere automation; it seeks to redefine the very fabric of presentation creation, promising an intuitive, enjoyable, and efficient experience for users across various domains. Slidester's ambitious scope sets it on a trajectory to be a transformative force in the technology-driven era of dynamic presentations

4. Proposed Design

Users begin by logging in or registering on the website, creating personalized profiles. In the presentation configuration phase, they choose a thematic domain and specify topics. The Language Model suggests slide titles, allowing users to refine the structure. After user edits, a dual-model process enriches titles with detailed points (LLM) and relevant images (Stable Diffusion Model). Users control presentation aspects through an intuitive interface, aided by an interactive chatbot. The final presentation can be exported in pptx or pdf, with all presentations archived for future reference.



4.1 Architecture Diagram

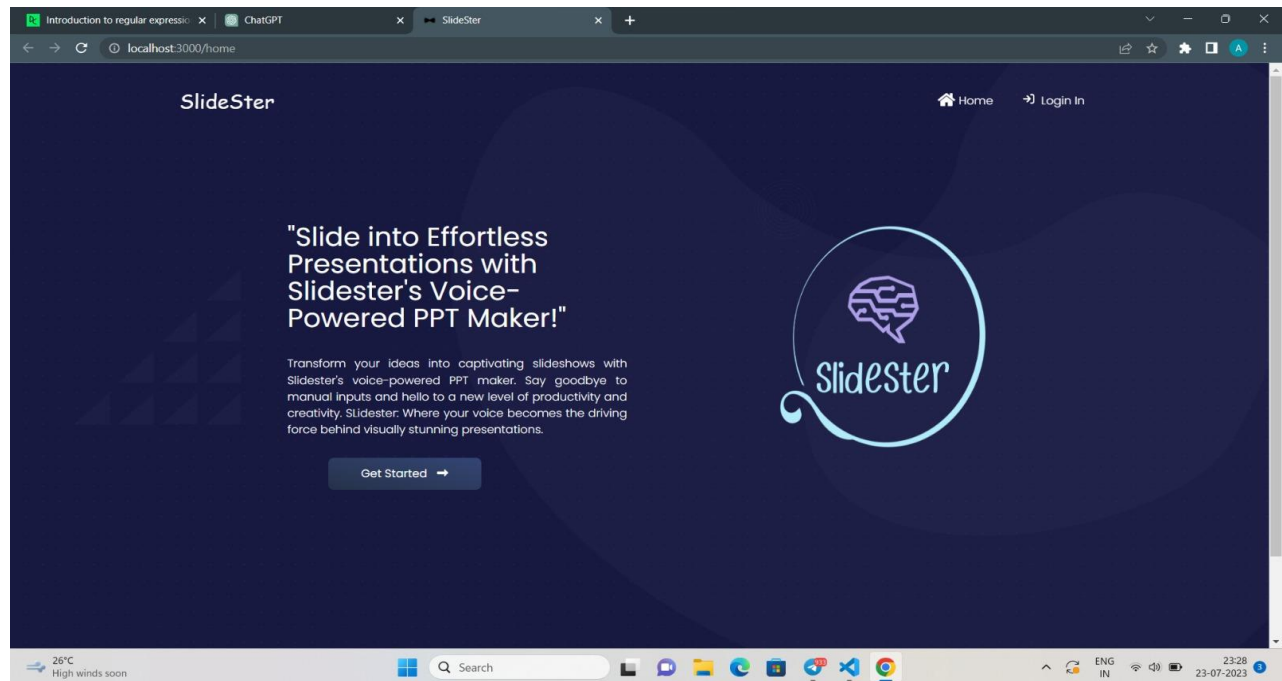
Explanation:

- Users initiate the process by either logging into their existing accounts or, for newcomers, registering on our website to create a personalized user profile.
- Once logged in, users navigate to the presentation configuration phase. Here, they select a thematic domain and specify a topic that aligns with the desired focus of their presentation.
- Our Language Model (LLM) takes center stage, suggesting preliminary slide titles based on the chosen domain and topic. Users are empowered to refine the structure further by specifying the number of points associated with each title.

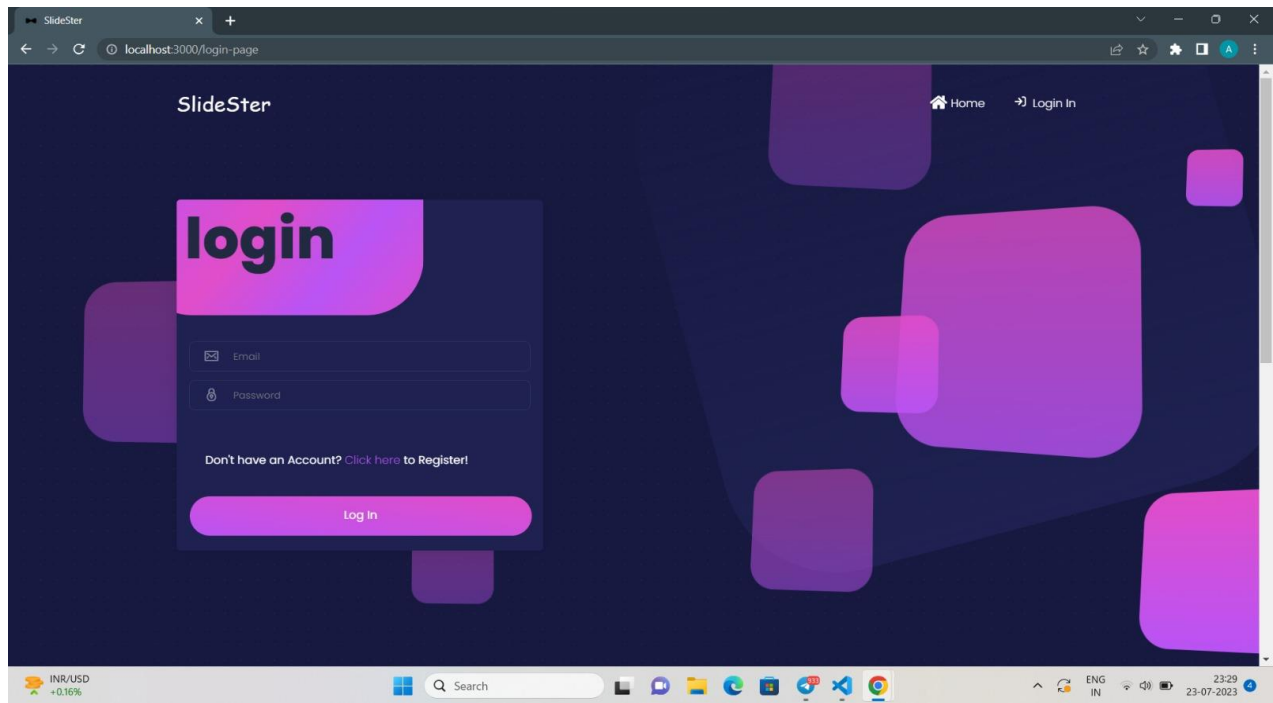
- Users actively engage in the creative process by reviewing and, if necessary, editing the suggested titles. They can introduce entirely new titles or modify existing ones until satisfied with the composition.
- The finalized titles undergo a dual-model process for enrichment. Model 1 (LLM) generates detailed point information for each title, while Model 2 (Stable Diffusion Model) employs the Latent Consistency Model (LCM-LoRA) to produce relevant images. Additionally, the presentation benefits from supplementary images sourced from the web through the Tavily client.
- The textual information and images collected in the previous steps seamlessly merge, resulting in a comprehensive presentation that combines well-crafted content with visually appealing elements.
- Users exercise control over the finer aspects of their presentation, including font styles and content modifications. This is facilitated through an intuitive conversational interface, allowing for a personalized touch.
- An interactive assistant, designed as a chatbot, becomes a valuable ally in the modification process. Users can interact with the assistant, providing necessary information for requested changes, which are then implemented seamlessly into the slides.
- The user has the convenience of exporting the final presentation in a preferred format, be it pptx or pdf, ensuring compatibility with different platforms and applications.
- All finalized presentations are automatically archived within the user's account. This feature serves as a practical repository, enabling users to conveniently revisit and reference their past presentations for future use or review.

5. Results and Discussions

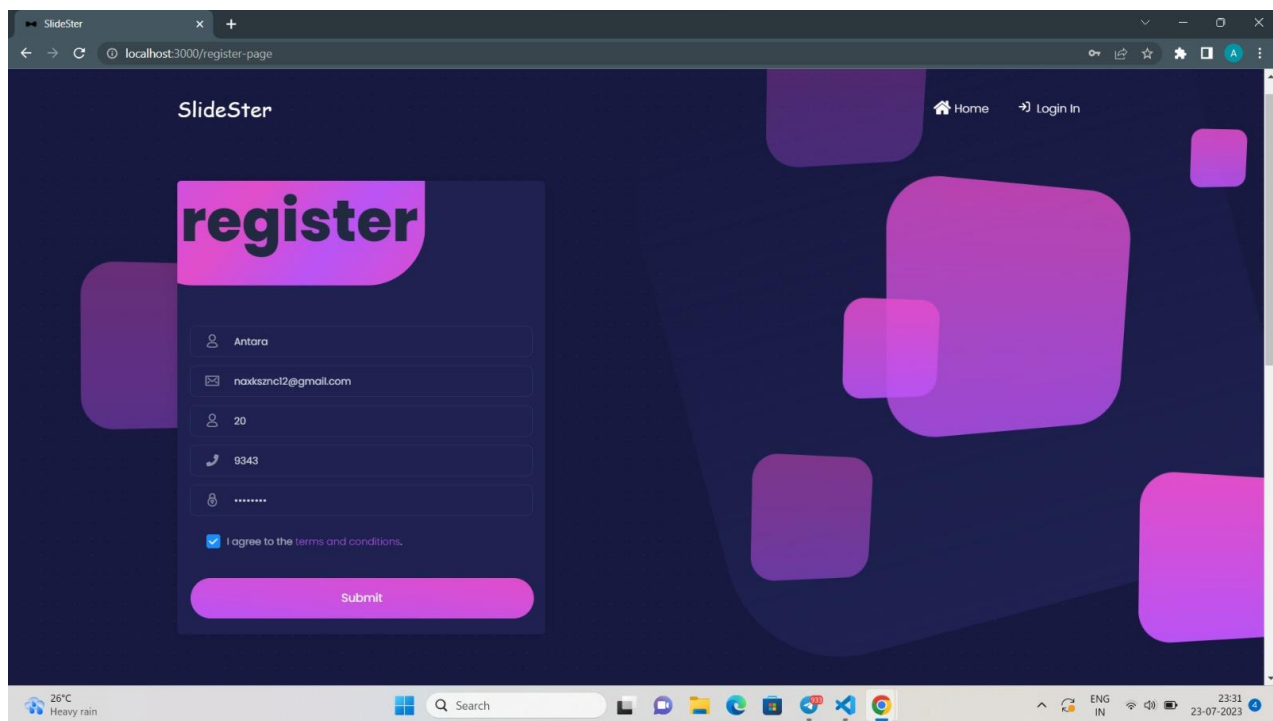
Welcome to our website's login and sign-up page, where user experience meets robust security. Our login page ensures a seamless entry with email and password fields, incorporating validation checks for accurate information. Password strength, email format, and optional two-factor authentication enhance security. New users can effortlessly register on our sign-up page, featuring unique email checks, password confirmation, and a captcha to prevent automated entries. Behind the scenes, we prioritize data security by utilizing MongoDB, a flexible and scalable NoSQL database. Your information is encrypted and stored with the utmost care, ensuring a smooth and protected online experience. Join us and enjoy the convenience of a secure login and sign-up process, backed by reliable data storage practices.



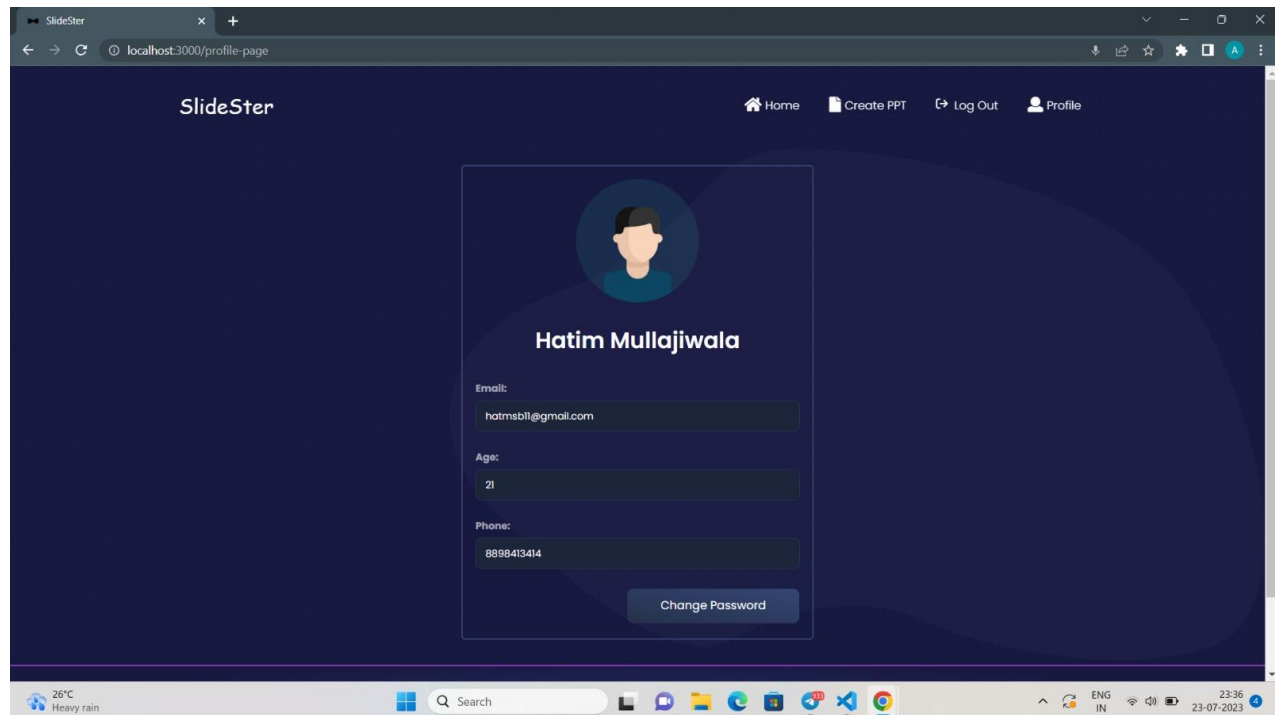
5.1 Home Page



5.2 Login Page

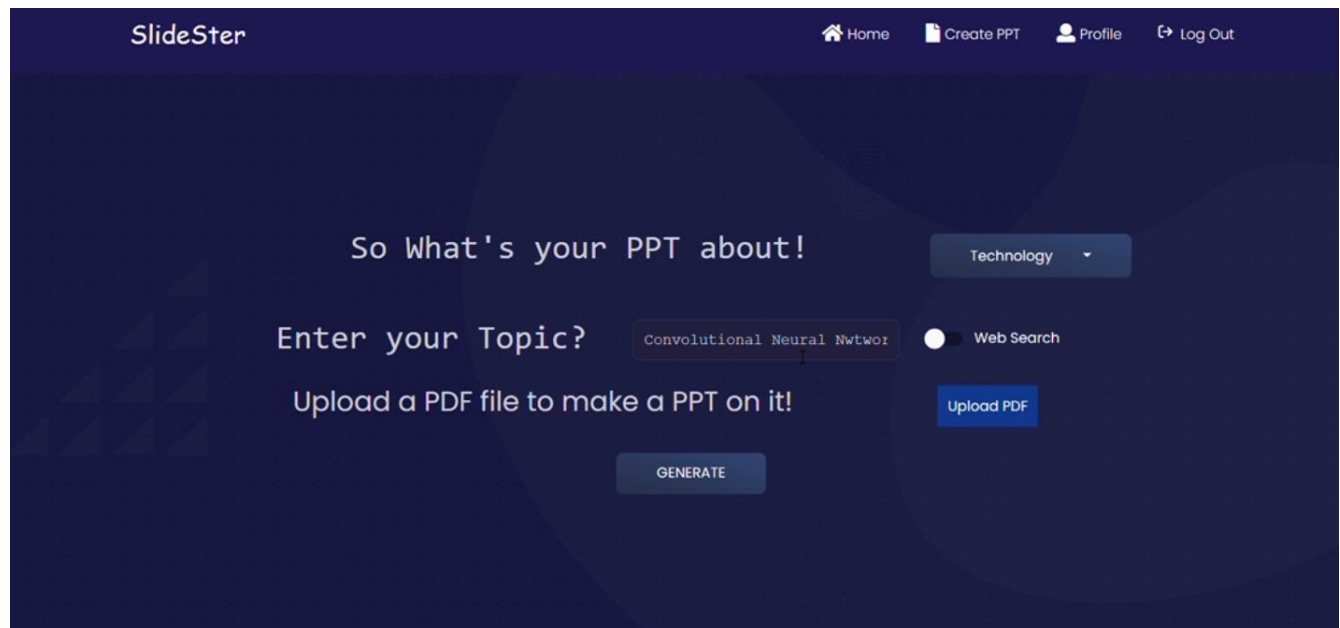


5.3 Register Page

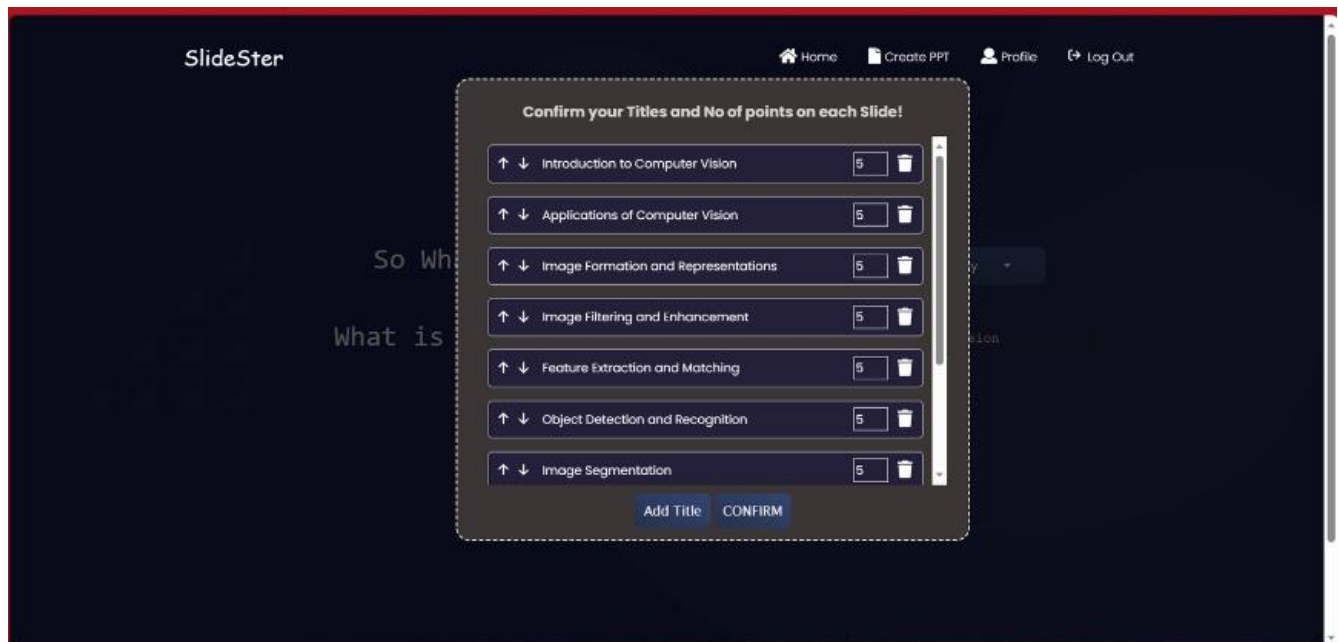


5.4 Profile Page

Our page offers a unique interactive experience, allowing users to tailor their PowerPoint presentations with precision. Users have the option to enable web search, leveraging the vast resources of the internet to enhance their content. Our system, equipped with intelligent algorithms, suggests potential slide titles based on user preferences and search queries. Elevate your presentations with our intuitive interface, harnessing the power of collaborative decision-making and intelligent suggestion algorithms for a more impactful and efficient content creation journey.

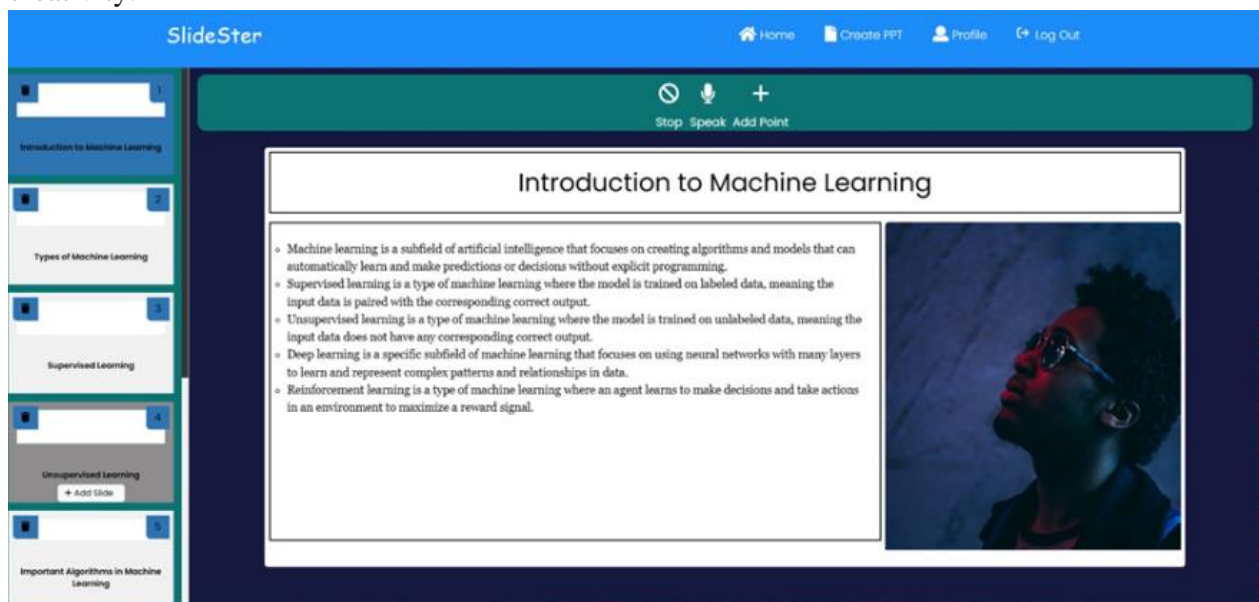


5.5 Subject Of The Ppt

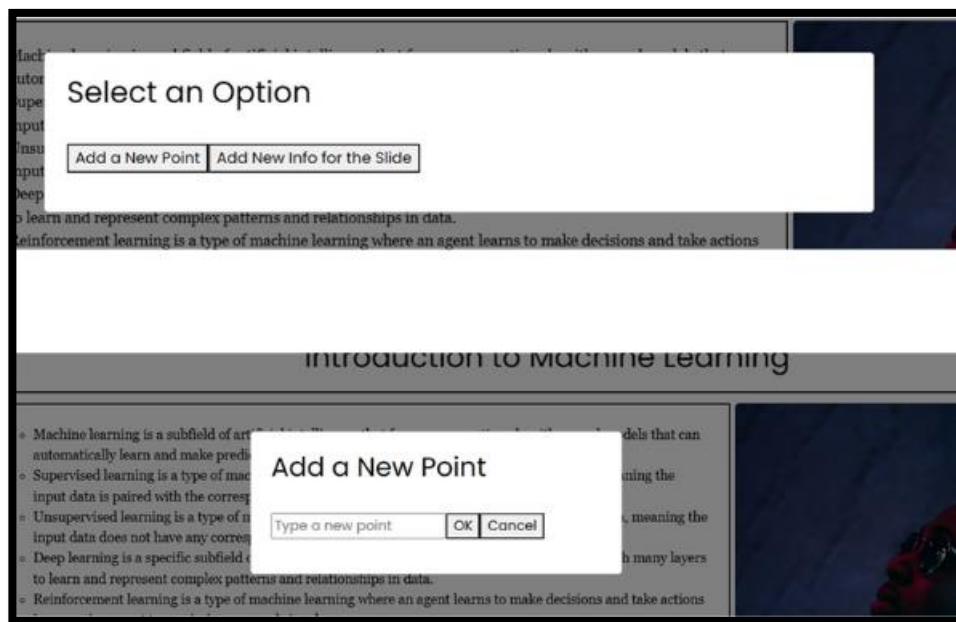


5.6 Select And Confirm Sub-Titles And No Of Points

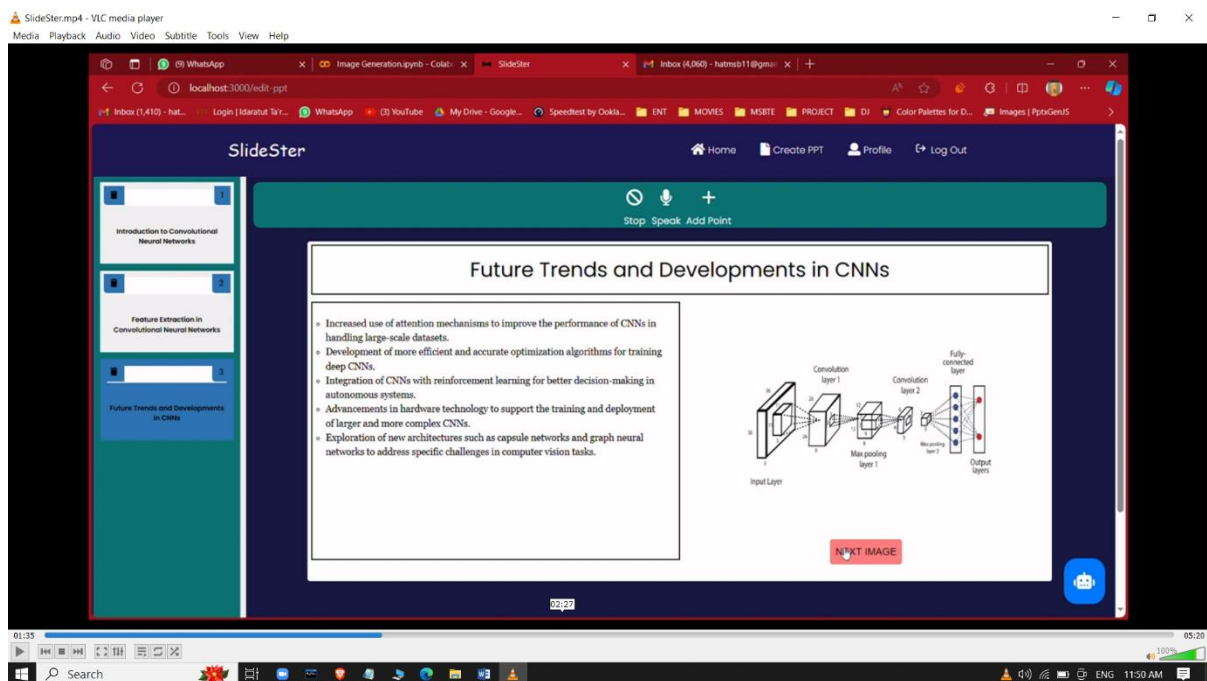
Our innovative system employs a state-of-the-art AI model to automatically generate information and dynamically displays it on a sleek slide interface. Elevating user interaction, a user-friendly chat interface is intelligently integrated, providing a collaborative space for users to effortlessly edit, refine, and add content to their presentations. This two-way interaction ensures a personalized touch, empowering users to tailor presentations to their specific needs. Furthermore, our image generation model enhances visual appeal by automatically incorporating stunning images that complement and enrich your content. Experience the future of presentation creation with our advanced AI-driven platform, combining intelligent automation with user-friendly interfaces for unparalleled efficiency and creativity.



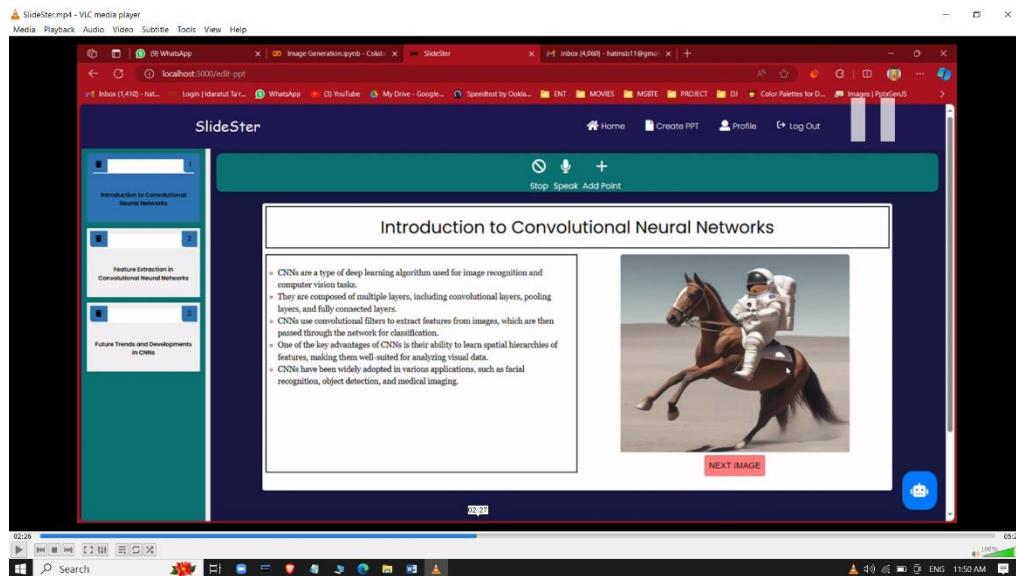
5.7 PPT Generated



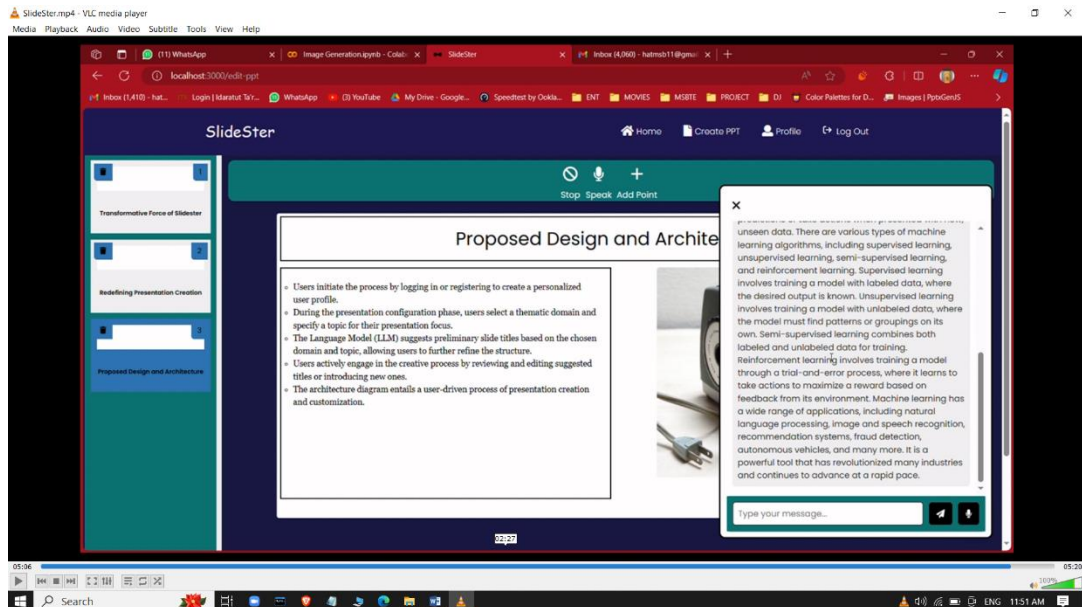
5.8 Edit The Points Or Titles



5.9 Generated PPT



5.10 Image Generated on the Slide



5.11 Assistant Api Integration

6. Conclusion

In conclusion, the Slidester represents a transformative and groundbreaking approach to presentation creation. By harnessing the power of advanced technologies, including NLP algorithms and speech recognition, the project offers users a seamless and intuitive means of converting spoken content into visually stunning presentations. With automated image generation and a user-friendly editing toolbar, Slidester empowers individuals from diverse domains to unleash their creativity and deliver captivating messages without the burden of tedious manual tasks. The project's scope extends to a wide range of users, from students to professionals, providing a versatile and dynamic platform for impactful communication. With its versatile and dynamic platform, the project opens doors for more efficient and enjoyable presentation-making experiences. Slidester paves the way for hassle-free presentation-making, saving time, and elevating the quality of presentations, leaving a lasting impression on audiences worldwide.

7. References

- [1] Sharma, N., & Sardana, S. (2016). A real time speech to text conversion system using bidirectional Kalman filter in Matlab. *Advances in Computing and Communications*. <https://doi.org/10.1109/icacci.2016.7732406>
- [2] Amodei, D., Ananthanarayanan, S., Anubhai, R., Bai, J., Battenberg, E., Case, C. J., Casper, J., Catanzaro, B., Cheng, Q., Chen, G., Chen, J., Chen, J., Chen, Z., Chrzanowski, M., Coates, A., Damos, G., Ding, K., Du, N., Elsen, E., . . . Zhu, Z. (2016). Deep speech 2: end-to-end speech recognition in English and mandarin. *International Conference on Machine Learning*, 173–182. <https://arxiv.org/pdf/1512.02595>
- [3] Saharia, C., Chan, W., Saxena, S., Li, L., Whang, J., Denton, E., Ghasemipour, S. K. S., Ayan, B. K., Mahdavi, S. R., Lopes, R. G., Salimans, T., Ho, J. C., Fleet, D. J., & Norouzi, M. (2022). Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding. *ArXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2205.11487>
- [4] Belote, P., Bidwai, S., Jadhav, S. R., Kapadnis, P., & Sharma, N. (2017b). Enhancing Automatic PPT Generation Technique through NLP for Textual Data. *International Journal of Advanced Research in Computer and Communication Engineering*. <https://doi.org/10.17148/ijarcce.2017.63237>
- [5] Kao, O., & Joubert, G. R. (2000). A content-based Internet search engine for analysis and archival of MPEG-1 compressed newsfeeds. *International Conference on Multimedia and Expo*. <https://doi.org/10.1109/icme.2000.871565>
- [6] P. M., Santra, S., Bhowmick, S., Paul, A., Chatterjee, P., & Deyasi, A. (2018). Development of GUI for text-to-speech recognition using natural language processing. 2018 2nd International Conference on Electronics, Materials Engineering & Nano-Technology (IEMENTech). <https://doi.org/10.1109/iementech.2018.8465238>
- [7] Bhat, A., Rao, A. C., Bhaskar, A., Adithya, V., & Pratiba, D. (2018). A cost-effective audio-visual summarizer for summarization of presentations and seminars. 2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS). <https://doi.org/10.1109/csitss.2018.8768740>
- [8] Ramzan, S., Iqbal, M. M., & Kalsum, T. (2022). Text-to-image generation using Deep Learning. *IEEC 2022*. <https://doi.org/10.3390/engproc2022020016>

- [9] J. D. Tardelli, C. M. Walter, "Speech waveform analysis and recognition process based on non-Euclidean error minimization and matrix array processing techniques". IEEE ICASSP, pp. 1237-1240, 1986.
- [10] Takao Suzuki, Yasuo Shoji, "A new speech processing scheme for ATM switching systems". IEEE, Digital Communications Laboratories, Oki Electric Industry Co. Ltd., Japan, pp. 1515-1519, 1989.
- [11] J. S. Lim "Evaluation of a correlated subtraction method for enhancing speech degraded by additive white noise", IEEE Trans. Acoustics, Speech and Signal Processing, vol. ASSP-26, no. 5, pp.471 -472 1978.
- [12] R. E. Kalman and R. S. Bucy, "New results in linear filtering and prediction theory," Trans. ASME Series D, J. Basic Engineering, pp. 95-108, 1961.
- [13] Gabrea, M.: 'Adaptive Kalman filtering-based speech enhancement algorithm'. IEEE Canadian Conf. on Electrical and Computer Engineering, 2001, vol. 1, pp. 521–526.
- [14] Jeong, S., Hahn, M.: 'Speech quality and recognition rate improvement in car noise environments', Electron. Lett., 2001, 37, (12), pp. 800–802.
- [15] Ma, J., Deng, L.: 'Efficient decoding strategies for conversational speech recognition using a constrained nonlinear state-space model', IEEE Trans. Speech Audio Process., 2003, 11, (6), pp. 590–602.
- [16] Mathe, M., Nandyala, S.P., Kishore Kumar, T.: 'Speech enhancement using Kalman filter for white, random and color noise'. IEEE Int. Conf. on Devices, Circuits and Systems (ICDCS), 2012, pp. 195–198.
- [17] Yeh Huann Goh, Paramesran Raveendran, Yann Ling Goh, "Robust speech recognition system using bidirectional Kalman filter", IET Trans. Pp. 1751-9675, 2015.
- [18] Tony Lacey, "Likelihood interpretation of Kalman filter", tutorial lecture, April 2006.
- [19] Siva Prasad Nandyala, Dr.T.Kishore Kumar," International Journal of Computer Applications" Volume- 12, pp.0975 – 8887, November 2010.
- [20] Paliwal, K.K, Atal, B.S. Efficient vector quantization of LPC Parameters at 24 bits / frame. IEEE Trans. Speech Audio Process, pp.3- 14, 1993.
- [21] M. J. F. Gales, A. Ragni, H. Aldamarki, and C. Gautier. Support vector machines for noise robust ASR. In ASRU, pages 205–2010, 2009.

- [22] A. Graves, S. Fernández, F. Gomez, and J. Schmidhuber. Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks. In ICML, pages 369–376. ACM, 2006.
- [23] A. Graves and N. Jaitly. Towards end-to-end speech recognition with recurrent neural networks. In ICML, 2014.
- [24] A. Graves, A.-r. Mohamed, and G. Hinton. Speech recognition with deep recurrent neural networks. In ICASSP, 2013.
- [25] H. H. Sak, A. Senior, and F. Beaufays. Long short-term memory recurrent neural network architectures for large scale acoustic modeling. In Interspeech, 2014.
- [26] A. Hannun, C. Case, J. Casper, B. Catanzaro, G. Diamos, E. Elsen, R. Prenger, S. Satheesh, S. Sengupta, A. Coates, and A. Y. Ng. Deep speech: Scaling up end-to-end speech recognition. 1412.5567, 2014. <http://arxiv.org/abs/1412.5567>.
- [27] A. Y. Hannun, A. L. Maas, D. Jurafsky, and A. Y. Ng. First-pass large vocabulary continuous speech recognition using bi-directional recurrent DNNs. abs/1408.2873, 2014. <http://arxiv.org/abs/1408.2873>.
- [28] K. Heafield, I. Pouzyrevsky, J. H. Clark, and P. Koehn. Scalable modified Kneser-Ney language model estimation. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, Sofia, Bulgaria, 8 2013.
- [29] G. Hinton, L. Deng, D. Yu, G. Dahl, A. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. Sainath, and B. Kingsbury. Deep neural networks for acoustic modeling in speech recognition. IEEE Signal Processing Magazine, 29(November):82–97, 2012.
- [30] S. Hochreiter and J. Schmidhuber. Long short-term memory. Neural Computation, 9(8):1735—1780, 1997.