International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 01 | Jan - 2026 SJIF Rating: 8.586 ISSN: 2582-3930

Presently-AI: An AI-Powered Presentation Video Generation System

Dr. Pavan G P Professor
Department of Information Science and Engineering AMC Engineering College
Bengaluru, India
drpavangp84@gmail.com

Keerthana *Scholar*
Department of Information Science and Engineering AMC Engineering College
Bengaluru, India
1am22is045@amceducation.in

Manoj Rao S *Scholar*
Student, Dept. of Information Science and Engineering AMC Engineering College
Bengaluru, India
manoj1784004@gmail.com

Likitha S Scholar
Department of Information Science and Engineering AMC Engineering College
Bengaluru, India
likithashetty1102004@gmail.com

Shobitha D *Scholar*
Department of Information Science and Engineering AMC Engineering College
Bengaluru, India
shobithalatha2004@gmail.com

Abstract: ***In the digital era, a vast amount of information is available in the form of webpages, articles, and online documents. However, converting this static textual content into engaging multimedia presentations requires significant manual effort, technical expertise, and time. This paper presents Presently-AI, an artificial intelligence based system that automatically transforms web content into professional presentation videos. The proposed system integrates web scraping, natural language processing, AI-based summarization, automated slide generation, text-to-speech narration, background music integration, and video compilation into a single unified platform. Using Google Gemini AI models, the system extracts relevant information from webpages, structures it into slide-wise content, generates narration, and compiles a complete video presentation with minimal user intervention. The system significantly reduces manual workload while ensuring consistent quality and professional output. Experimental results demonstrate that Presently-AI effectively automates the end-to-end workflow of presentation video creation, making it highly suitable for students, educators, and professionals seeking efficient content delivery solutions.***

Keywords: ***Artificial Intelligence, Web Scraping, NLP, Presentation Video Generation, Text-to-Speech.***

I. INTRODUCTION

The rapid advancement of digital technologies has significantly transformed the way information is created, distributed, and consumed. In today's academic, professional, and corporate environments, the internet serves as the primary source of information, with vast amounts of content published daily in the form of webpages, blogs, online articles, tutorials, and research documents. While this information is easily accessible, it is predominantly presented in textual form, which may not always be effective in engaging users or supporting efficient knowledge acquisition. Modern learners and professionals increasingly prefer audio-visual content, such as presentation videos, as it enhances comprehension, retention, and overall learning experience. Despite the growing demand for multimedia-based content, the process of converting textual web content into presentation videos remains largely manual. Traditional methods require users to extract relevant information, summarize the content, design presentation slides, record voice narration, select suitable background music, and perform video editing using multiple software tools.

This work flow is time-consuming and requires technical expertise in presentation design and video editing, making it difficult for many users to efficiently produce professional-quality presentation videos. Artificial Intelligence (AI) offers effective solutions for automating multimedia content creation. Advances in Natural Language Processing (NLP), text summarization, and text-to-speech synthesis enable AI systems to analyze large volumes of text, extract key information, generate concise summaries, and produce natural-sounding narration, thereby minimizing manual effort. However, existing tools provide only partial automation, focusing on individual tasks such as slide creation or voice generation. These fragmented solutions require manual integration of multiple tools, limiting efficiency and scalability. To overcome these limitations, this paper proposes Presently-AI, an AI-powered system that automatically converts webpage content into professional presentation videos. The system performs intelligent content extraction, AI-based summarization, automated slide generation, narration synthesis, and video compilation within a unified workflow, ensuring reduced effort, improved efficiency, and consistent output quality.

II. RELATED WORK

Automating the generation of presentation content has been an active area of research, with prior work exploring various approaches for slide creation, text summarization, and multimedia synthesis. Traditional methods rely on manual workflows or semi-automated tools that address individual stages of the content creation pipeline. For example, systems such as automated slide builders aim to convert text into slides but often lack integration with voice narration or video output, limiting their applicability for complete presentation generation. Earlier research in text summarization has concentrated on extracting key information from documents using statistical and machine learning techniques. Techniques such as extractive and abstractive summarization help reduce long texts into concise summaries for easier consumption. [1] Zhu, Z., Lin, K. Q., & Shou, M. Z. (2025). *Paper2Video: Automatic Video Generation from Scientific Papers.* arXiv, 2025. However, these summarization methods alone do not provide direct mechanisms for converting summaries into multimedia content. While advanced Natural Language Processing (NLP) models have been successful at generating human-like summaries, their integration into end-to-end multimedia systems remains limited.

In the domain of voice synthesis, text-to-speech (TTS) engines have evolved to produce increasingly natural audio output. Modern AI-based TTS systems are capable of generating expressive narration, which enhances the accessibility and engagement of educational and professional content. Despite these advancements, existing solutions generally function as isolated modules that require manual coordination with slide generation and video editing. Several recent projects attempt to unify parts of this workflow. Some research prototypes generate slides automatically from structured data or lecture transcripts, and others incorporate AI-driven voice synthesis to accompany slides. Additionally, there are specialized tools that offer automated video editing features for combining visual and auditory elements. However, these tools often lack a unified pipeline that seamlessly integrates content extraction, summarization, slide creation, narration, and video assembly. Presently-AI differentiates itself from prior work by providing an integrated AI-driven solution that combines all major components of presentation video generation into a single automated platform. Unlike systems that focus only on slide creation or voice synthesis, Presently-AI processes web content end-to-end — from extracting relevant information to delivering a finalized presentation video synthesis techniques, the proposed system addresses limitations observed in earlier work and advances the automation of multimedia content creation.[2] Zheng, H., et al. (2025). *PPTAgent: Generating and Evaluating Presentations Beyond Text-to-Slides.* arXiv, 2025.

III. PROBLEM STATEMENT

The rapid growth of web-based educational and professional content has increased the demand for converting textual information into engaging presentation videos. However, the traditional process of creating presentation videos from webpage content is largely manual and involves multiple stages such as content extraction, summarization, slide design, narration recording, and video editing. This workflow is time-consuming and requires significant technical expertise, making it difficult for many users to produce professional-quality presentations efficiently.

Existing tools provide only partial automation and focus on isolated tasks such as slide generation or voice synthesis. As a result, users must rely on multiple independent tools, leading to inefficiency, increased effort, and inconsistent output quality. Additionally, the lack of an integrated workflow limits scalability and productivity. Therefore, there is a clear need for an intelligent, AI-driven system that can automate the complete process of presentation video creation. Such a system should seamlessly integrate content extraction, AI-based processing, slide generation, narration synthesis, and video compilation into a unified platform. The proposed system aims to address these challenges by minimizing human intervention, reducing time consumption, and ensuring consistent and professional output quality.This approach also improves accessibility for non-technical users and supports scalable content creation across educational and professional domains. [3] **Xu, T., Xu, X., Chen, S., et al. (2025).** *PreGenie: An Agentic Framework for High-Quality Visual Presentation Generation.* arXiv, 2025.

IV. SYSTEM DESIGN OVERVIEW

The system design of Presently-AI is based on a modular and scalable architecture that enables automated generation of presentation videos from webpage content. The system is designed to minimize manual effort while ensuring efficient coordination between artificial intelligence and multimedia processing components. Users interact with the system through a web-based interface by providing a webpage URL, which serves as the primary input for the entire workflow. After receiving the input, the system employs a web scraping module to extract relevant textual and visual information from the webpage while eliminating unnecessary elements such as advertisements and navigation menus. The extracted content is processed by the AI processing module, which uses Natural Language Processing techniques to analyze, summarize, and organize the information into structured, slide-wise content suitable for presentation. This step ensures that only meaningful and concise information is retained for further processing.

A. Software Architecture

The software architecture of Presently-AI follows a modular, layered approach to ensure scalability, maintainability, and efficient processing. The system is divided into distinct functional layers, including the presentation layer, application layer, AI processing layer, and multimedia generation layer. This separation of concerns allows each module to operate independently while maintaining smooth interaction across the system. The presentation layer provides a web-based user interface through which users submit webpage URLs and configure basic preferences. The application layer acts as the core controller, managing workflow execution, user requests, and data flow between modules. The AI processing layer is responsible for content analysis and summarization using Natural Language Processing techniques and AI models. This layer converts extracted webpage content into structured, slide-ready information and narration scripts. The multimedia generation layer handles slide creation, text-to-speech synthesis, and video compilation. Presentation slides are generated using automated formatting techniques, while AI-based text-to-speech engines produce natural narration. Finally, the video synthesis component integrates slides, audio narration, and background music to generate the final presentation video. This software architecture enables end-to-end automation, improves system efficiency, and ensures consistent output quality.

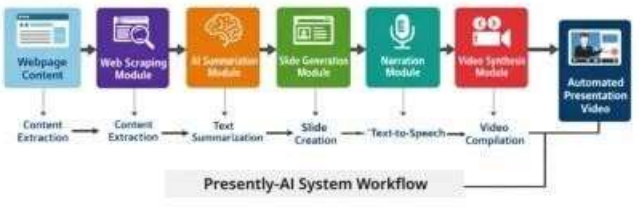

Fig 1.1. Architecture diagram

Software architecture defines the high-level structure of a software system its components, how they interact, and the guiding principles behind those decisions. It provides a blueprint that ensures scalability, maintainability, and alignment with business goals.

B. Functional Workflow

The functional workflow of Presently-AI begins with user interaction through a web-based interface. The user provides a webpage URL and initiates the presentation generation process. The system validates the input and forwards the request to the backend controller, which manages the sequential execution of all processing stages.

Once the work flow is initiated, the web scraping module extracts relevant textual and visual content from the provided webpage. This extracted data is then processed by the AI processing module, where Natural Language Processing techniques are applied to analyze, summarize, and structure the content into slide-wise information. The AI module also generates narration text corresponding to each slide, ensuring logical flow and clarity. In the final stage, the structured content is passed to the slide generation module , which automatically creates formatted presentation slides. The text-to-speech module converts narration text into natural-sounding audio, and the video generation module synchronizes slides, narration, and background music to produce the final presentation video. The completed output is then delivered to the user. This functional work fl ow ensures end-to-end automation, reduced manual effort, and consistent presentation quality.

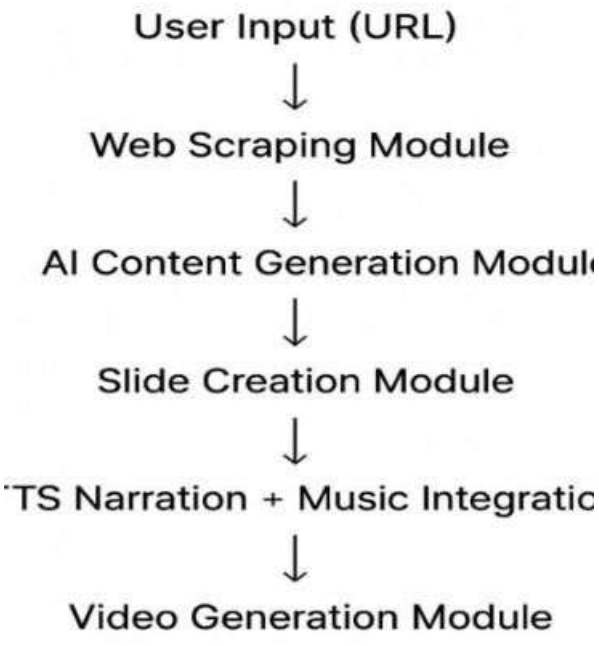

Fig 1.2: Flow chart illustrates the basic operations from user input to system response.

V. METHODOLOGY

The proposed Presently-AI system follows a structured and automated methodology to convert webpage content into professional presentation videos. Initially, the user provides a webpage URL as input, which is processed using a web scraping module to extract relevant textual and visual content. Noise removal and content filtering techniques are applied to ensure only meaningful information is retained for further processing. Next, the extracted content is passed to the AI summarization module, where Natural Language Processing (NLP) techniques are used to identify key concepts and generate concise, slide-wise summaries.

This summarized content is then structured into presentation format and forwarded to the slide generation module, which automatically creates visually consistent slides using predefined templates and layout rules. In the final stage, the narration module converts slide text into natural-sounding speech using text-to-speech technology. Background music and timing synchronization are applied, and the video synthesis module compiles slides and audio into a complete presentation video. This end-to-end methodology minimizes human intervention, improves efficiency, and ensures consistent output quality.

V1. IMPLEMENTATION

The proposed Presently-AI system follows a structured and automated methodology to convert webpage content into professional presentation videos. Initially, the user provides a webpage URL as input, which is processed using a web scraping module to extract relevant textual and visual content. Noise removal and content filtering techniques are applied to ensure only meaningful information is retained for further processing. Next, the extracted content is passed to the AI summarization module, where Natural Language Processing (NLP) techniques are used to identify key concepts and generate concise, slide-wise summaries. This summarized content is then structured into presentation format and forwarded to the slide generation module, which automatically creates visually consistent slides using predefined templates and layout rules. In the final stage, the narration module converts slide text into natural-sounding speech using text-to-speech technology. Background music and timing synchronization are applied, and the video synthesis module compiles slides and audio into a complete presentation video. This end-to-end methodology minimizes human intervention, improves efficiency, and ensures consistent output quality.

Fig 1. 3:Input Page

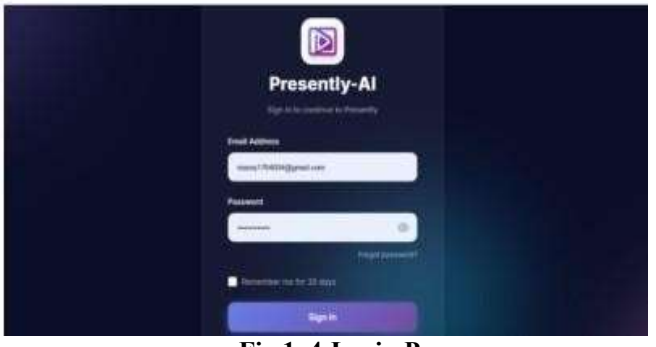

Fig 1. 4:Login Page

VI. RESULTS AND DISCUSSION

The Presently-AI system was evaluated using multiple publicly accessible webpages containing educational and informational content. The system successfully extracted relevant text from the input pages and processed it using AI-based summarization techniques. The generated content was organized into structured, slide-wise information, demonstrating the system's ability to identify key concepts while eliminating redundant details. The slide generation and narration modules produced professional-quality outputs with consistent formatting and clear audio narration. AI-based text-to-speech synthesis generated natural and understandable speech, while the integration of background music enhanced overall presentation engagement. The final presentation videos were visually appealing and suitable for academic and professional use. The automated workflow minimized user involvement and ensured consistent output quality across different inputs. These results confirm that Presently-AI is an efficient, reliable, and practical solution for automated presentation video generation.

Figure 1.5: Execution Window

Figure 1.6: User Registration (Sign Up) Interface

Figure 1.7:User Login Interface

Live Results of the proposed methodology:

Figure 1. 9: Web Scraping

Figure 1.10 : AI Content Generation

Figure 1. 11 : Slide Generation

Figure 1. 12 : Video Generation

Figure 1.8: Presently - AI Home Page

VII. BENEFITS AND DRAWBACKS

The Presently-AI system provides several significant benefits by automating the complete workflow of presentation video creation from web-based content. It greatly reduces manual effort and time by eliminating the need for multiple tools and technical expertise in slide design and video editing. The system ensures consistent presentation quality through standardized slide formatting and AI-based narration, making it suitable for both educational and professional use. Its user-friendly interface allows even non-technical users to generate presentation videos efficiently, while the modular design supports scalability and future enhancements.

However, the system also has certain drawbacks. Its performance depends heavily on internet connectivity and the quality of the input webpage content. Limited customization options may restrict user control over visual and audio styles, and AI-generated summaries may occasionally lack contextual accuracy. Additionally, processing very large or unstructured webpages may increase execution time and affect overall system performance. From a practical perspective, the system enhances accessibility by enabling users to transform textual web content into engaging audio-visual presentations without requiring advanced technical skills. It supports efficient knowledge sharing and improves learning outcomes by presenting information in a more interactive format. At the same time, the system's reliance on automated AI models introduces certain constraints, such as reduced flexibility in content interpretation and limited adaptability to highly specialized or domain-specific material. Despite these limitations, the benefits of automation, consistency, and ease of use outweigh the drawbacks, making the proposed system a valuable solution for automated presentation video generation.

VIII. CONCLUSION

This paper presented Presently-AI , an AI-powered system designed to automate the generation of presentation videos from webpage content. By integrating web scraping, AI-based content summarization, automated slide creation, text-to-speech narration, and multimedia video synthesis, the system effectively reduces manual effort and time consumption involved in traditional presentation creation. The proposed system ensures consistent presentation quality and provides a user-friendly solution for students, educators, and professionals. Experimental results demonstrate that Presently AI successfully transforms textual web content into structured, engaging presentation videos with minimal human intervention. Although the system has certain limitations related to customization and dependency on input content quality, its overall performance and efficiency validate the effectiveness of the proposed approach. The system highlights the potential of artificial intelligence in automating multimedia content generation and contributes to improved productivity and accessibility in modern digital learning and professional environments.

IX. FUTURE SCOPE

While the present system fulfills its intended purpose effectively, there significant potential for future enhancements. The following improvements can extend the system's capabilities and make it more adaptive to evolving technologies and user needs. Multilingual Support:-Extend AI content generation and narration to multiple languages. Enable cross presentation creation for global accessibility Background music and timing synchronization are applied, and the video synthesis module compiles slides and audio into a complete presentation video. This end-to-end methodology minimizes human intervention, improves efficiency, and ensures consistent output quality.

Customizable Slide Templates:-Allow users to choose from predefined templates, color themes, and layouts. Support importing custom PowerPoint templates for branding or professional use. Cloud Deployment and Online Storage:-Host the application on cloud platforms (e.g., AWS, Render, Google Cloud) for global access. Provide secure cloud-based storage and sharing for generated videos. AI Voice Customization:-Introduce options for selecting different voice tones, genders, and accents for narration. Include expressive voice models for more engaging presentations. Integration with File Uploads:-Enable input not only from URLs but also from PDF, Word, or text documents, allowing broader usability. Mobile and Cross-Platform Version:-Develop Android/iOS applications for easier access and real-time generation. Ensure full responsiveness across all devices. Video Editing and Effects:-Add transitions, animations, and visual effects for more dynamic presentations. Allow users to edit or reorder slides before final video generation. AI-Generated Visuals:-Integrate image or video generation models (e.g., Gemini Image or DALL·E) to automatically create visuals for slides. Collaboration Features:-Enable multi-user collaboration for group projects or team presentations

REFERENCES

[1] **Zhu, Z., Lin, K. Q., & Shou, M. Z. (2025).** *Paper2Video: Automatic Video Generation from Scienti fi c Papers.* arXiv, 2025.

[2] **Zheng, H., et al. (2025).** *PPTAgent: Generating and Evaluating Presentations Beyond Text-to-Slides.* arXiv, 2025.

[3] **Xu, T., Xu, X., Chen, S., et al. (2025).** *PreGenie: An Agentic Framework for High-Quality Visual Presentation Generation.* arXiv, 2025.

[4] **Gigant, T., Guinaudeau, C., & Dufaux, F. (2025).** *Summarization of Multimodal Presentations with Vision-Language Models.* arXiv, 2025.

[5] **Li, J., Zhang, H., & Wang, Y. (2025).** *Automated Multimedia Presentation Generation Using Large Language Models.* IEEE Access, 2025.

[6] **Kim, S., & Park, J. (2025).** *AI-Assisted Presentation Video Creation with Speech Synthesis.* ACM Multimedia Workshops, 2025.

[7] ***Suryawanshi, K., Gaikwad, A., & Mali, M. (2024)*** *.Automated PowerPoint Presentation Generation from PDF Documents Using NLP and Machine Learning.* International Journal of Multidisciplinary Research, 2024.

[8] ***Rao, P., & Mehta, S. (2024).*** *Web Content to Presentation Automation Using NLP Techniques.* International Journal of Computer Applications, 2024.

[9] **Chen, Y., Liu, X., & Zhao, M. (2024).** *End-to-End AI Framework for Slide and Narration Generation.* IEEE Access, 2024.

[10] **Patil, A., & Kulkarni, R. (2024)** .Smart Presentation Generation Using Text Summarization and TTS. IJCSIT, 2024.

[11] **Singh, R., & Verma, A. (2024).** AI-Based Educational Content Transformation into Video Presentations. Springer CCIS, 2024.

[12] *Alonso, D., & Torres, M. (2024).* *Multimedia Content Creation Using NLP and Deep Learning.* Elsevier Procedia Computer Science, 2024.

[13] **Gupta, T. (2023).** *Automatic Presentation Slide Generation Using Large Language Models.* M.S. Thesis, San Jose State University, 2023.

[14] **Nguyen, N. X. V., et al. (2023).** *Automatic Transcript Generation from Presentation Slides.* PACLIC, 2023.

[15] **Sharma, K., & Joshi, N. (2023).** *AI-Based Slide Generation for E-Learning Applications.* IJERT, 2023.

[16] **Wang, L., & Chen, P. (2023).** *Natural Language Processing for Automated Presentation Design.* IEEE Conference on AI, 2023.

[17] **Kumar, S., & Reddy, V. (2023).** *Smart Video Presentation Generation Using AI Narration.* International Journal of Advanced Research in CS, 2023.

[18] **Brown, A., & Smith, J. (2023).** *Automated Visual Content Creation Using NLP and TTS.* ACM International Conference, 2023.