iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** V **Month of publication:** May 2026

DOI: https://doi.org/10.22214/ijraset.2026.82212

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

EDUVID: An AI-Powered Educational Video Generation System

S Veerendra Swamy[1], S M Saranya[2], Sushma Kulkarni[3], Vaishnavi Prabhakar[4], G. Chandrakala[5]
Department of Artificial Intelligence and Machine Learning Ballari Institute of Technology and Management, Ballari, India

Abstract: ***Traditional textbook learning often makes it hard for students to understand long and complicated ideas. Online videos are not always in line with the exact content of the textbook or syllabus. Many learners spend extra time looking for good explanations, which can cause confusion and make learning less efficient. This paper introduces EDUVID, an AI-powered system that creates educational videos from selected textbook pages. The system uses automated techniques to make simple and engaging videos. Users can upload a textbook, choose specific pages, and generate videos that match their needs. EDUVID uses OCR to extract text, NLP to summarize content, TTS for audio narration, and AI to create slides for visual explanations. A rule-based chatbot helps users upload files and select pages, making the platform easy to use for those who aren't tech-savvy. Experimental results show that EDUVID greatly cuts down the time needed to process content—from several minutes of manual work to just a few seconds—while improving learning efficiency and reducing the effort required from users through automated video creation.***

Index Terms: AI in Education, Educational Video Generator, OCR, NLP Summarization, Text-to-Speech, E-learning

I. INTRODUCTION

The growing use of digital learning tools highlights the need for educational support systems that can deliver content quickly and clearly in real-world learning environments. Traditional methods such as reading long textbooks or searching for online videos require significant time and effort, which becomes difficult for students under academic pressure. Many existing online learning resources are generic and do not align with the exact syllabus or textbook chapters being studied, often leading to confusion and incomplete understanding.

In academic situations, the available time to understand complex topics is limited, and learners may struggle to process lengthy text-based material. At the same time, fully manual content creation approaches for educational videos are time- consuming, require technical skills, and are difficult to scale for every subject or chapter. Therefore, there is a strong need for a lightweight and automated system that can generate personalized educational videos, operate with minimal user effort, and support multiple content transformation techniques, including text extraction and summarization.

This paper introduces *EDUVID*, an AI-powered educational video generation system that combines textbook processing with automated multimedia creation to provide quick and effective learning support. The system supports textbook up- load, selective page extraction, AI-based summarization, and automatic video generation with narration and visuals. A built- in rule-based chatbot guides users during file upload and page selection, while the generated videos provide a clear and engaging way to understand academic content without relying on external video platforms.

The main contributions of this work are as follows:

1) An end-to-end AI-driven educational video generation framework that converts selected textbook pages into narrated video content.
2) A selective page processing mechanism that enables personalized and focused learning for specific topics.
3) A lightweight technical design using OCR, NLP-based summarization, and Text-to-Speech without reliance on manual content creation.
4) A systematic experimental evaluation including unit, module, integration, and system testing, along with performance comparison with manual learning methods in terms of processing time and usability.

The remainder of this paper is organized as follows. Section II reviews existing research on AI-based educational content generation. Section III describes the proposed system method- ology and architecture. Section IV presents implementation details. Section V discusses experimental results and performance analysis. Section VI outlines future work, and Section VII concludes the paper.

II. LITERATURE REVIEW

Educational technology solutions have been widely explored through e-learning platforms, AI-based tutoring systems, and multimedia learning applications. Several learning systems utilize digital resources such as recorded lectures, interactive modules, animations, and online content repositories to sup- port students during their learning process. Amato et al. [2] proposed an AI-powered learning framework that personal- izes content delivery based on student progress and learn- ing behavior. Although such platforms improve accessibility, they mostly depend on pre-existing digital content and do not support direct processing of user-provided textbooks or documents. AI-driven content generation tools have also been intro- duced to automate learning material creation. Nyame and Bengesi [1] explored generative AI models for video creation using text prompts and visual synthesis techniques. While these approaches reduce manual effort in content creation, they require high computational resources and do not focus on converting structured academic content such as textbook pages into educational videos. The lack of integration between document extraction and video generation limits their direct use in academic learning scenarios. Conversational AI and NLP-based systems are increasingly used to assist educators in content preparation and assessment. Koraishi [3] demonstrated the use of large language models to generate educational materials, exercises, and feedback for language learning. Although such systems improve content generation efficiency, they are primarily text-based and do not provide multimedia transformation such as automated video generation or text-to-speech narration from textbook content. Despite the availability of these technologies, common limitations include dependence on pre-built digital resources, lack of direct textbook processing, and absence of end-to-end automation from document upload to video creation. Many existing systems do not allow users to select specific pages from textbooks, summarize complex content, and automat- ically generate learning videos in a single workflow. This makes personalized, syllabus-specific video creation difficult for students and educators. The proposed system, *EDUVID*, addresses these limitations by integrating a lightweight web-based platform with AI- driven text extraction, summarization, and automated video generation modules. Unlike existing content platforms, the proposed system enables users to upload textbooks, select specific pages, and generate personalized educational videos using OCR, NLP, and Text-to-Speech technologies. Addi- tionally, EDUVID supports user guidance through a rule- based chatbot and allows offline access to generated videos, improving accessibility and usability in real-world learning scenarios.

III. METHODOLOGY

The EDUVID framework is designed as an end-to-end, automated educational content transformation system that in- tegrates textbook processing with AI-based multimedia gener- ation. The methodology includes system architecture design, functional module integration, content extraction and summa- rization mechanisms, and a structured development workflow used to implement and evaluate the system.

A. System Architecture

The system follows a modular architecture that connects the user interface, backend processing server, OCR module, NLP- based summarization engine, Text-to-Speech service, visual generation module, and video rendering component. The fron- tend allows users to upload textbooks and select specific pages, while the backend manages content extraction, processing, and video generation tasks. The OCR module extracts readable text from selected PDF pages, and the NLP module summa- rizes complex content into simplified explanations. The Text- to-Speech module converts the summarized text into audio narration, and the visual generation module creates simple slide-based visuals that are combined to produce the final educational video.

When the user initiates video generation, the selected text- book pages are processed through OCR and summarization modules. The summarized script is then converted into speech and combined with AI-generated visuals to form an educa- tional video. A rule-based chatbot assists users during document upload and page selection, ensuring smooth interaction with the system.

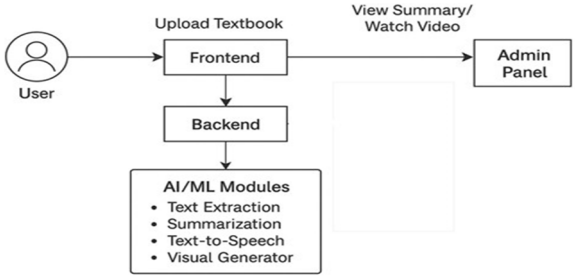


Fig. 1. System architecture of the EDUVID framework illustrating interaction between user interface, backend server, OCR module, NLP summarization engine, Text-to-Speech module, visual generation module, and video rendering component.

B. System Workflow

The end-to-end system workflow is described as follows.

1) The user opens the EDUVID application interface and uploads a textbook in PDF format.
2) The system displays all pages of the uploaded document for user selection.
3) The user selects the required page range for video generation.
4) The system performs OCR on the selected pages to extract textual content.
5) The extracted text is summarized using NLP-based sum- marization techniques.
6) The summarized content is converted into speech using Text-to-Speech.
7) AI-based visuals and slide templates are generated for explanation.
8) The narration and visuals are combined to generate the final educational video.
9) The generated video is displayed to the user with a download option.
10) The user submits feedback including name, email, rat- ing, and comments.

This workflow enables educational videos to be generated with minimal user effort while maintaining a fully automated processing pipeline.

C. Content Processing Techniques and Algorithms

The proposed system integrates multiple AI-based process- ing techniques to ensure reliable and accurate video genera- tion.

1) Text Extraction: The OCR module extracts textual content from uploaded PDF pages using optical character recognition techniques, enabling processing of both digital and scanned documents.
2) Text Summarization: The extracted text is processed using NLP-based summarization algorithms to condense lengthy content into concise and meaningful explanations suitable for video narration.
3) Text-to-Speech Conversion: The summarized script is converted into natural-sounding speech using Text-to-Speech engines, providing audio narration for the educational video.

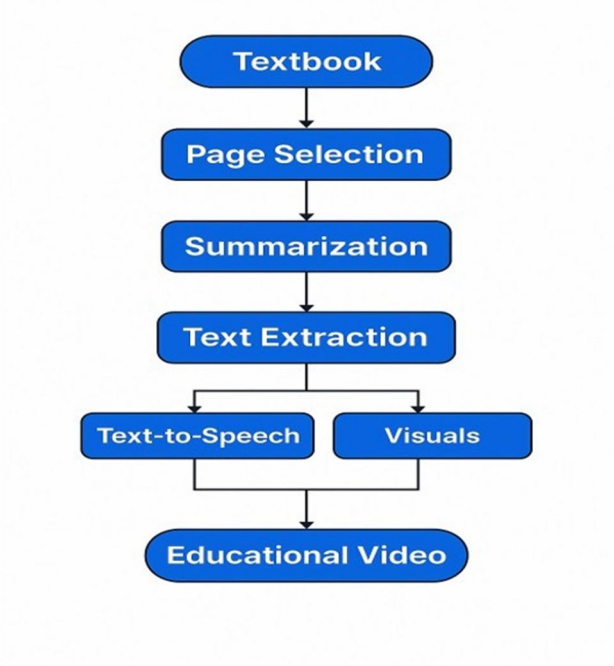


Fig. 2. System workflow of the EDUVID framework showing interaction between user interface, backend processing, OCR, NLP summarization, Text- to-Speech conversion, visual generation, and final video output.

4) Visual Generation: Simple AI-based slide visuals and templates are generated to represent key points and concepts extracted from the summarized content.
5) Video Composition: The generated visuals and audio nar- ration are merged using video rendering techniques to produce the final educational video output.

D. Development Approach

The project followed a structured Spiral development methodology consisting of requirement analysis, system de- sign, implementation, testing, and evaluation phases. The re- quirement phase identified essential features such as selective page extraction, automated summarization, AI-based video generation, chatbot-assisted guidance, and feedback collection. The system design phase included the development of architecture diagrams, data flow diagrams, use case diagrams, database design, and user interface layouts. The implementa- tion phase involved backend development using Python and Flask, integration of OCR, NLP, Text-to-Speech, and video generation modules, and frontend development using HTML, CSS, and JavaScript. The testing phase included unit testing, module testing, integration testing, and system-level testing under realistic usage conditions to validate system performance and reliability.

IV. IMPLEMENTATION

A. Software Implementation

The EDUVID application was developed as a web-based system using Python and the Flask framework for backend processing, along with HTML, CSS, and JavaScript for the user interface. The system utilizes OCR libraries for ex- tracting text from PDF documents, NLP-based summarization models for simplifying complex textbook content, Text-to- Speech (TTS) libraries for converting summarized text into audio narration, and video processing libraries for generating educational videos. The backend handles file uploads, page selection, text extraction, summarization, audio generation, and video rendering in an automated pipeline. User inputs and feedback details are validated and processed dynamically, ensuring a lightweight and maintainable system design.

The application is organized into multiple functional mod- ules:

1) Textbook Upload Module: Allows users to upload PDF textbooks and validates file format before processing.
2) Page Selection Module: Enables users to view and select specific pages from the uploaded textbook for focused video generation.
3) Text Extraction Module: Extracts readable text from selected pages using OCR and PDF parsing techniques.
4) Text Summarization Module: Processes extracted con- tent using NLP-based summarization to generate concise explanations.
5) Text-to-Speech Module: Converts summarized text into natural-sounding audio narration for the video.
6) Video Generation Module: Generates slide-based visu- als and combines them with audio narration to create the final educational video.
7) Feedback Module: Collects user feedback including name, email, rating, and comments for evaluation and improvement.

B. Hardware Implementation

The EDUVID system is implemented as a software-only solution and does not require any specialized hardware com- ponents. All processing, including text extraction, summa- rization, audio generation, and video rendering, is performed on standard computing devices such as laptops or desktop systems. This design choice reduces hardware dependency, cost, and maintenance requirements while ensuring ease of deployment across different environments. The system can run on commonly available hardware configurations with standard peripherals, making it suitable for use in educational institutions and by individual learners without the need for additional wearable or IoT devices.

C. Tools and Technologies

The development and implementation of the system utilized the following tools and technologies:

1) Python: Used for backend logic, OCR processing, NLP summarization, Text-to-Speech conversion, and video generation.
2) Flask Framework: Used to build the backend server, manage user requests, and coordinate AI processing mod- ules.
3) OCR Libraries (Tesseract, PDFPlumber): Used for extracting text from scanned and digital PDF files.
4) NLP Libraries: Used for summarization and text pro- cessing.
5) Text-to-Speech Libraries (gTTS / pyttsx3): Used to generate audio narration from summarized content.
6) Video Processing Libraries (OpenCV / MoviePy):
7) Used to generate and render educational videos.
8) Frontend Technologies (HTML, CSS, JavaScript): Used to design the user interface for uploading docu- ments, selecting pages, and viewing generated videos.

V. RESULTS AND DISCUSSION

A. Testing Strategy

The EDUVID system was validated through unit, module, integration, and system testing to ensure reliable behavior across individual components and complete workflow execu- tion. Unit testing focused on verifying individual functions such as PDF upload validation, page range selection, OCR text extraction, text summarization accuracy, Text-to-Speech audio generation, and video rendering.

Module testing verified that major functional components, including the document upload module, page selection mod- ule, text extraction module, summarization module, Text- to-Speech module, video generation module, and feedback module, operated correctly as integrated units.

Integration testing evaluated interactions between modules, including PDF upload to OCR processing, OCR output to sum- marization pipeline, summarization output to Text-to-Speech conversion, synchronization between audio generation and visual slide creation, and final video rendering. Error handling during invalid file uploads, incorrect page ranges, and partial content extraction was also tested.

System testing examined complete end-to-end operational scenarios such as full textbook upload to video generation workflow execution, system behavior during large PDF pro- cessing, performance under low computational resources, re- peated video generation requests, invalid feedback submis- sions, and system recovery from partial processing failures.

Across all defined test cases, the expected system behavior was observed. The system consistently generated educational videos from selected textbook pages while maintaining stable processing and user interaction. The working user interface of the EDUVID system is shown in Fig. 3.

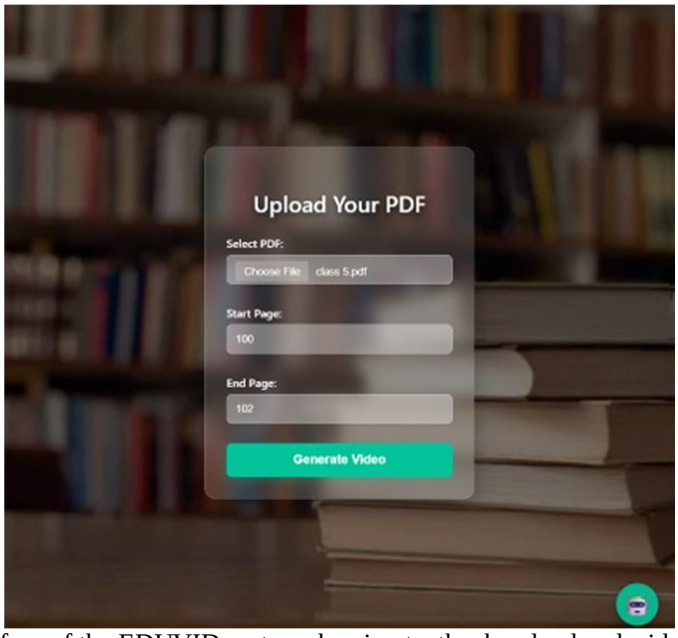


Fig. 3. Main interface of the EDUVID system showing textbook upload and video generation options.

B. Performance Analysis

To quantify system effectiveness, performance evaluation was conducted using metrics such as content extraction time, summarization accuracy, video generation time, system stabil- ity, and overall processing reliability.

Manual methods such as reading textbooks and searching for suitable videos typically require several minutes for each topic. In contrast, the proposed EDUVID system generated summarized educational videos within a few seconds to under a minute using automated processing techniques. The OCR module achieved high accuracy in extracting read- able text from digital and scanned documents under normal conditions. The summarization module effectively condensed long textbook content into concise explanations suitable for narration. The video generation pipeline maintained stable synchronization between audio narration and visual slides across multiple test runs. The system remained operational during repeated processing cycles and handled multiple user requests with minimal failure rate.

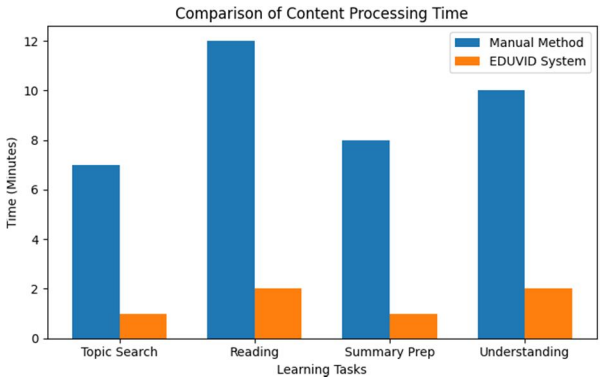


Fig. 4. Comparison of content processing time between manual methods and EDUVID system.

C. Discussion

The experimental evaluation confirms that EDUVID suc- cessfully achieves its primary objectives by enabling auto- mated conversion of textbook content into educational videos with minimal user effort. The system significantly reduces the time required to understand complex topics and provides a structured, visual learning experience aligned with selected textbook pages.

The integration of OCR, NLP-based summarization, Text- to-Speech narration, and automated visual generation offers a practical and cost-effective educational support system suitable for students and educators. The system improves learning efficiency while maintaining reliable processing and consistent output quality. However, certain improvement areas remain. Enhancing OCR performance for low-quality scanned documents could further improve extraction accuracy, and refining summarization models could produce more context-aware explanations. Optimizing video rendering performance would also improve user experience for larger documents. Despite these limitations, the proposed system demonstrates that an AI-driven educational video generation framework can significantly enhance personalized learning without requiring complex manual content creation.

VI. FUTURE WORK

Future enhancements to EDUVID will focus on improving system intelligence, accessibility, and learning support. Multilingual video generation can be added to support learners from different language backgrounds. Advanced AI-based summarization techniques may further improve the quality and accuracy of generated educational content.

TABLE I
PERFORMANCE COMPARISON BETWEEN MANUAL METHODS AND EDUVID

Metric	Manual	EDUVID
Processing Time	Several mins	Few seconds
Summarization	Not Available	Available
Video Generation	Manual	Automated
User Effort	High	Minimal
Reliability	Variable	High

The system can also be extended with features such as automated quiz generation, progress tracking, and interactive learning modules to increase student engagement. Mobile application support and cloud-based processing may improve accessibility, scalability, and processing speed for larger educational resources. Integration with learning management systems and real- time performance monitoring can further enhance usability in academic environments. These improvements can help EDUVID evolve into a more efficient and personalized digital learning platform. The proposed framework highlights the effectiveness of integrating AI-based text processing with automated multimedia generation to create intuitive and learner-friendly educational tools. By minimizing manual effort and supporting automated content transformation, EDUVID improves accessibility and real-world applicability of personalized digital learning systems. Future enhancements such as multilingual support, advanced AI-based summarization models, interactive learning features, mobile application support, and cloud-based scalability can further strengthen the system's functionality and intelligence. With these advancements, EDUVID has the potential to evolve from a basic automated video generator into a comprehensive intelligent learning ecosystem capable of adaptive content delivery, personalized learning support, and scalable educational resource creation.

VII. ACKNOWLEDGMENT

The authors would like to thank the Department of Artificial Intelligence and Machine Learning, Ballari Institute of Technology and Management, for their continuous support and guidance during the development of this project.

VIII. CONCLUSION

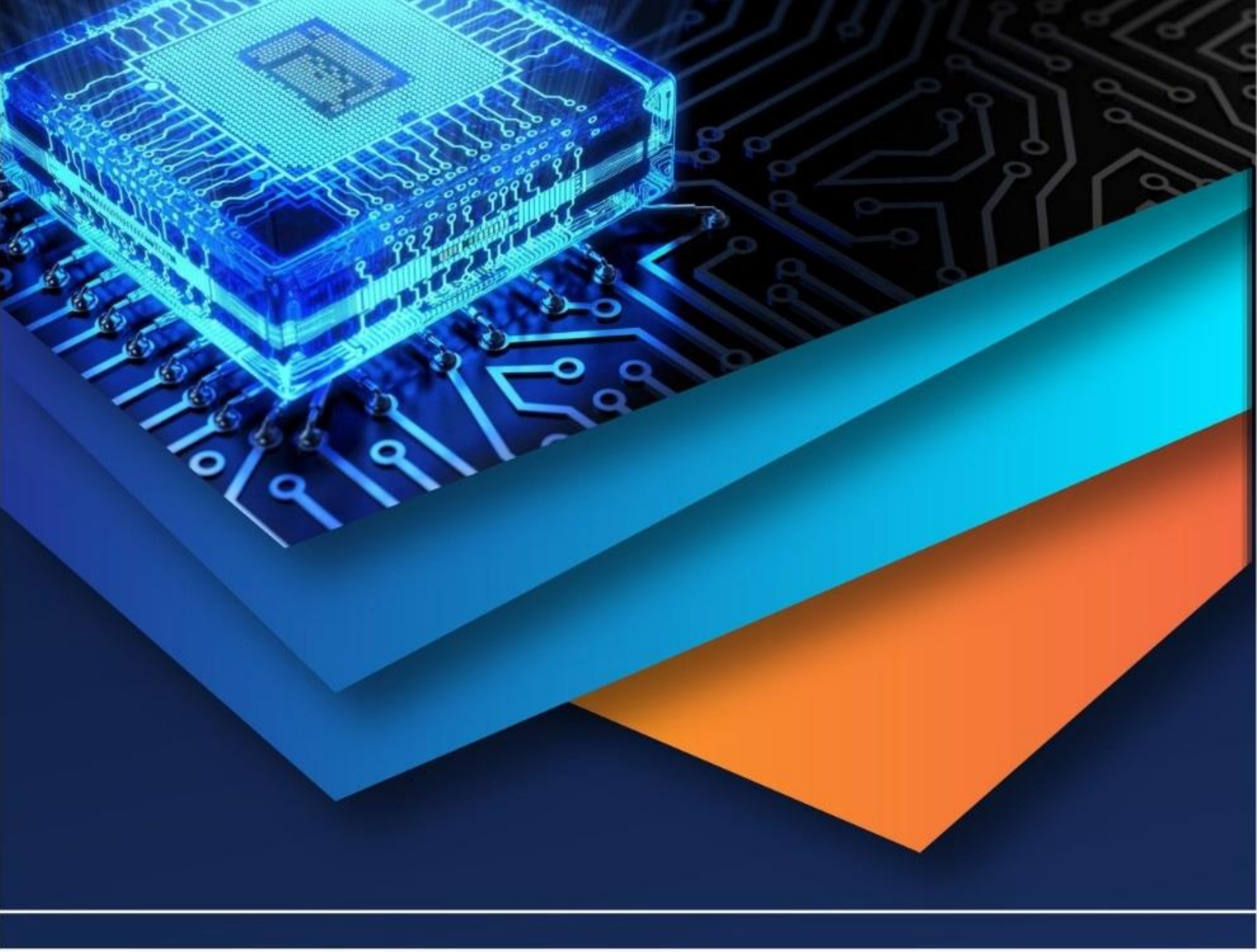
This paper presented *EDUVID*, an AI-powered educational video generation system developed to provide simple and effective learning support by converting textbook content into engaging educational videos. The proposed framework integrates document upload, selective page extraction, OCR- based text extraction, NLP-driven summarization, Text-to- Speech narration, and automated visual generation to create a reliable and accessible learning solution. By incorporating multiple processing stages—including text extraction, content simplification, audio narration, and visual slide creation—the system ensures that learners can understand complex topics through structured multimedia explanations. The framework further enhances usability by allowing users to generate videos from selected textbook pages without requiring advanced technical skills.

The EDUVID system plays a significant role in enabling personalized and focused learning. By allowing learners to choose specific pages from textbooks, the system generates targeted educational videos that match the exact syllabus or topics required. Through automated content processing and narration, users can access learning material without spending time searching for suitable external videos. The system oper- ates as a standalone educational support platform and does not depend on manual content creation, making it suitable for use in resource-constrained academic environments.

The design and development of EDUVID followed a structured engineering approach involving system architec- ture modeling, modular implementation, and comprehensive testing. Extensive unit, integration, and system-level testing demonstrated that the framework consistently performs as expected across various document sizes and usage scenarios. Performance analysis indicated that the system significantly reduces learning preparation time compared to manual meth- ods while maintaining stable processing and reliable video generation. Additionally, the lightweight software-based de- sign ensures low deployment cost and ease of use, making the framework practical for everyday academic use.

REFERENCES

[1] I. Sommerville, *Software Engineering*, 9th ed., Pearson Education, 2011.
[2] L. Nyame and S. Bengesi, “Generative Artificial Intelligence Trend on Video Generation,” Preprints.org, Sep. 2024, doi: 10.20944/preprints.202409.0195.v1.
[3] F. Amato *et al.*, “AI-Powered Learning: Personalizing Education for Each Student,” Unpublished Manuscript, University of Naples Federico II, Italy, 2024.
[4] O. Koraishi, “Teaching English in the Age of AI: Embracing ChatGPT to Optimize EFL Materials and Assessment,” *Lang. Educ. Technol.*, vol. 3, no. 1, pp. 55–72, May 2023.
[5] T. Williams and J. Carter, “Enhancing Learning Experiences Through Automated Video Creation Using NLP and TTS Technologies,” *IEEE Trans. Learn. Technol.*, vol. 17, no. 2, pp. 145–156, 2024.
[6] S. Das and P. Mehta, “Integrating Artificial Intelligence in Digital Education: A Review of Emerging Tools and Techniques,” *Educ. Inf. Technol.*, vol. 29, no. 1, pp. 112–130, Jan. 2024.
[7] A. Kumar and V. Rao, “Automated Educational Content Generation Using Machine Learning Techniques,” *Int. J. Adv. Res. Comput. Sci.*, vol. 15, no. 6, pp. 320–328, Jun. 2024.
[8] H. Tanaka and Y. Sato, “Advances in Text-to-Speech and Summarization Models for E-Learning Systems,” *J. Artif. Intell. Res.*, vol. 59, no. 3, pp. 450–468, 2023.
[9] Python Software Foundation, “Python Documentation,” [Online]. Avail- able: https://docs.python.org/3/.
[10] Google Developers, “Text-to-Speech Overview,” 2024. [Online]. Avail- able: https://cloud.google.com/text-to-speech/docs.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY